



Bibliography of Language Testing

Volume 3 2021–2024

This bibliography has been compiled by the Graduate Student Assembly of ILTA.

Entries and corrections may be submitted in December of each calendar year. Please watch listservs and social media for announcements about the call for entries.

Editorial Board

Volume	Years	Editor(s)	
1	1990–1999	Jayanti Banerjee Caroline Clapham	Phoebe Clapham Dianne Wall
	1999–2012	Annie Brown	
	2013–2018	Tineke Brunfaut	
2	ILTA GSA Bibliography Committee		
	2019	Olena Rossi	Wenjun (Elyse) Ding
	2020	Dylan Burton (Editor in Chief) Jorge Beltrán Hyunah Kim Wenjun (Elyse) Ding	Minh Ngo Haoshan (Sally) Ren Olena Rossi Yan (Niles) Zhao
	2021	Dylan Burton (Editor in Chief) Wenjun (Elyse) Ding Haeun Kim Hyunah Kim	Minh Ngo Liberato Silva dos Santos Monique Yoder Yan (Niles) Zhao
	2022	Dylan Burton (Editor-in-Chief) Shanshan He (Editor-in-Chief) Jorge Beltrán Andy Jiahao Liu Coral Yiwei Qin	Laura Yi Tan Chenyang Zhang Jing Zhang Yan (Niles) Zhao
	2023	Shanshan He (Editor in Chief) Deise Amaral Andy Jiahao Liu Coral Yiwei Qin	Laura Yi Tan Chia-Hsin Yin Chenyang Zhang Yan (Niles) Zhao
	2024	Shanshan He (Editor in Chief) Deise Amaral Xianzhi Chen I-Chun Vera Hsiao Coral Yiwei Qin	Monique Yoder Chengyuan Yu Chenyang Zhang Niles Zhao

Questions about the bibliography may be directed to Shanshan He: she296@uwo.ca

Table of Contents

Editorial Board.....	1
Bibliography by Alphabet.....	4
A.....	4
B.....	10
C.....	16
D.....	23
E.....	28
F.....	30
G.....	33
H.....	37
I.....	45
J.....	48
K.....	50
L.....	58
M.....	66
N.....	73
O.....	77
P.....	80
Q.....	84
R.....	86
S.....	91
T.....	101
U.....	105
V.....	106
W.....	108
X.....	113
Y.....	115
Z.....	119
Bibliography by Category.....	125
Assessing Specific Competences and Populations.....	125
Assessing Reading.....	125
Assessing Listening.....	131
Assessing Speaking.....	137
Assessing Writing.....	153
Assessing Integrated Skills.....	181
Assessing Sign Languages.....	185
Assessing Vocabulary and Grammar.....	186
Assessing Interactional Competence and Pragmatics.....	195
Assessing Mediation, Translation, and Interpretation.....	199
Assessing Young Learners.....	201
Assessing Learners with Specific Learning Differences.....	207

Assessing Digital and Multimodal Literacy.....	209
Language Test Construction and Development.....	212
Validity and Validation.....	212
Test Design and Development.....	230
Rating and Rating Scales.....	240
Specific Purpose Testing.....	249
Technology and Language Assessment.....	255
Artificial Intelligence and Language Assessment.....	271
Social Aspects of Language Assessment.....	274
Test Impact, Washback, and Consequences.....	274
Standards and Frameworks.....	282
Language Assessment Literacy.....	285
Approaches to Assessment.....	296
Fairness and Ethics.....	313
Language Testing Policy.....	318
Research in Language Assessment.....	323
Research Methodology.....	323
Research Synthesis.....	328
Open Science and Academic Publishing.....	331
Other.....	331
Reflections and Perspectives on the Field.....	331
Bibliography of Material and Datasets.....	333

Bibliography by Alphabet

A

- Abbott, M., Lee, K. & Ricioppo, S. (2021). Does portfolio-based language assessment align with learning-oriented assessment? Evidence from literacy learners and their Instructors. *Canadian Journal of Applied Linguistics / Revue Canadienne de Linguistique Appliquée*, 24(2), 260–285. <https://doi.org/10.37213/cjal.2021.31338>
- Abdellatif, M. S., Alshehri, M. A., Alshehri, H. A., Hafez, W. E., Gafar, M. G., & Lamouchi, A. (2024). I am all ears: Listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement. *Language Testing in Asia*, 14, Article 54. <https://doi.org/10.1186/s40468-024-00329-6>
- Abdi Tabari, M., & Wang, Y. (2022). Assessing linguistic complexity features in L2 writing: Understanding effects of topic familiarity and strategic planning within the realm of task readiness. *Assessing Writing*, 52, Article 100605. <https://doi.org/10.1016/j.asw.2022.100605>
- Abdolrezapour, P., & Ghanbari, N. (2021). Enhancing learning potential score in EFL listening comprehension and self-regulation through self-regulated dynamic assessment procedures. *Language Testing in Asia*, 11, Article 10. <https://doi.org/10.1186/s40468-021-00126-5>
- Abdulaal, M. A. A. D., Alenazi, M. H., Tajuddin, A. J. A., & Hamidi, B. (2022). Dynamic vs. diagnostic assessment: Impacts on EFL learners' speaking fluency and accuracy, learning anxiety, and cognitive load. *Language Testing in Asia*, 12, Article 32. <https://doi.org/10.1186/s40468-022-00179-0>
- Abdullaeva, B. S., Çakmak, F., & Abdullaev, D. (2024). Paper assessment or online assessment: Exploring the impact of assessment modes on EFL students' language learning outcomes and personal development. *Language Testing in Asia*, 14(, Article 35. <https://doi.org/10.1186/s40468-024-00309-w>
- Abrar-ul-Hassan, S., & Nassaji, H. (2024). Rescoping language assessment literacy: An expanded perspective. *System*, 120, Article 103195. <https://doi.org/10.1016/j.system.2023.103195>
- Abrar-ul-Hassan, S., Douglas, D., & Turner, J. (2021). Revisiting second language portfolio assessment in a new age. *System*, 103, Article 102652. <https://doi.org/10.1016/j.system.2021.102652>
- Acar, A. S. (2023). Genre pedagogy: A writing pedagogy to help L2 writing instructors enact their classroom writing assessment literacy and feedback literacy. *Assessing Writing*, 56, Article 100717. <https://doi.org/10.1016/j.asw.2023.100717>
- Agrawal, S., Oza, P., Kakkar, R., Tanwar, S., Jetani, V., Undhad, J., & Singh, A. (2024). Analysis and recommendation system-based on PRISMA checklist to write systematic review. *Assessing Writing*, 61, Article 100866. <https://doi.org/10.1016/j.asw.2024.100866>
- Ahmadi Safa, M., & Sheykhmololuki, H. (2023). An impact study of the Iranian National University Entrance Exam from students and parents' perspectives. *Language Testing in Asia*, 13, Article 40. <https://doi.org/10.1186/s40468-023-00254-0>
- Ahmadi, A. (2024). Patterns of interaction in a paired speaking test: Comparing L1 and L2 interactions. *Language Testing in Asia*, 14, Article 31.

- <https://doi.org/10.1186/s40468-024-00299-9>
- Ahmed Abdel-Al Ibrahim, K., Cuba Carbajal, N., Zuta, M. E. C., & Bayat, S. (2023). Collaborative learning, scaffolding-based instruction, and self-assessment: Impacts on intermediate EFL learners' reading comprehension, motivation, and anxiety. *Language Testing in Asia*, 13, Article 16. <https://doi.org/10.1186/s40468-023-00229-1>
- Ahmed Abdel-Al Ibrahim, K., Karimi, A. R., Abdelrasheed, N.S.G., & Shatalebi, V. (2023). Group dynamic assessment vs. computerized dynamic assessment: Impacts on L2 listening development, L2 students' perfectionism, foreign language anxiety, and intrinsic motivation. *Language Testing in Asia*, 13, Article 32 (2023). <https://doi.org/10.1186/s40468-023-00245-1>
- Akase, M. (2022). Longitudinal measurement of growth in vocabulary size using Rasch-based test equating. *Language Testing in Asia*, 12, Article 5. <https://doi.org/10.1186/s40468-022-00155-8>
- Akbarian, I. H., & Farrokhi, J. (2021). Lexical availability and writing ability of EFL learners. *International Journal of Applied Linguistics*, 31(1), 48–64. <https://doi.org/10.1111/ijal.12319>
- Akgün, H. & Ünalı, A. (2022). Investigating cognitive processes in different item formats in reading tests through eye-tracking and verbal protocols. *Studies in Language Assessment*, 11(2), 98–129. <https://doi.org/10.58379/UIXK2220>
- Al Lawati, Z. A. (2023). Investigating the characteristics of language test specifications and item writer guidelines, and their effect on item development: A mixed-method case study. *Language Testing in Asia*, 13, Article 21. <https://doi.org/10.1186/s40468-023-00233-5>
- Al Sharoufi, H. (2022). Towards a unified English technology-based writing curriculum in the Arabian Gulf countries: the case of Oman. *Language Testing in Asia*, 12, Article 33. <https://doi.org/10.1186/s40468-022-00178-1>
- Al-Ghezi, R., Voskoboinik, K., Getman, Y., Von Zansen, A., Kallio, H., Kurimo, M., Huhta, A., & Hildén, R. (2023). Automatic speaking assessment of spontaneous L2 Finnish and Swedish. *Language Assessment Quarterly*, 20(4-5), 421–444. <https://doi.org/10.1080/15434303.2023.2292265>
- Al-Hawamdeh, B. O. S., Hussien, N., & Abdelrasheed, N. S. G. (2023). Portfolio vs. summative assessment: Impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13, Article 12. <https://doi.org/10.1186/s40468-023-00225-5>
- Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2023). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition*, 46(2), 307–329. <https://doi.org/10.1017/S0272263123000487>
- Al-Rashidi, A. H., Asif, M., Vanani, M. G., & Aberash, A. (2022). Learner-oriented assessment (LOA) practice: The comparative study of self-assessment, peer assessment, and teacher assessment on EFL learners' writing complicity, accuracy, and fluency (CAF), speaking CAF, and attitude. *Language Testing in Asia*, 12, Article 59. <https://doi.org/10.1186/s40468-022-00209-x>
- AL-TAMIMI, Y.r A., QUEVEDO-CAMARGO, G., & MELO, R. S. R. (2022). Washback of a High-stakes Language Exam on Spanish Teachers' Practice in Brazil. *Journal of Linguistics and Language Teaching*, v. 13, p. 1-20.

- Aladini, A., Mahmud, R., & Ali, A. A. H. (2024). The importance of needs satisfaction, teacher support, and L2 learning experience in Intelligent Computer-Assisted Language Assessment (ICALA): A probe into the state of willingness to communicate as well as academic motivation in EFL settings. *Language Testing in Asia*, 14, Article 58. <https://doi.org/10.1186/s40468-024-00334-9>
- Alam, S., Vadivel, B., Banu, S., & Jamalyar, R. (2024). Reflecting the voices of EFL teachers in the world of Intelligent Computer-Assisted Language Assessment (ICALA): An insight into teacher immunity, reflective teaching, job satisfaction, and L2-teacher grit. *Language Testing in Asia*, 14, Article 43. <https://doi.org/10.1186/s40468-024-00314-z>
- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Alazemi, A. F. T., Jember, B., & Al-Rashidi, A. H. (2023). How to decrease test anxiety: A focus on academic emotion regulation, L2 grit, resilience, and self-assessment. *Language Testing in Asia*, 13, Article 28. <https://doi.org/10.1186/s40468-023-00241-5>
- Albrecht, S. (2022). What do teachers do in the age of Zoom? The Covid-19 crisis and the role of the language teacher (Research Notes Issue 82). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/654641-research-notes-82.pdf>
- Aldosari, M., Heydarnejad, T., Hashemifardnia, A., & Abdalgane, M. (2023). The interplay among self-assessment, using reflection for assessment, classroom enjoyment, and immunity: Into prospects of effective language learning. *Language Testing in Asia*, 13, Article 1. <https://doi.org/10.1186/s40468-022-00213-1>
- Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing*, 52, Article 100610. <https://doi.org/10.1016/j.asw.2022.100610>
- Ali, M. H., & Abdulrazzaq, D. M. (2024). Dynamic versus static assessment: Effects on Iraqi EFL learners' resilience, burnout, positive orientation, and academic success. *Language Testing in Asia*, 14, Article 37. <https://doi.org/10.1186/s40468-024-00308-x>
- Ali, M. M. E. (2023). The foreign-accentedness, comprehensibility, and intelligibility of L2 Arabic speech. *Language Teaching Research*, 0(0). <https://doi-org.ezproxy2.library.arizona.edu/10.1177/13621688231158787>
- Ali, M. M., & Hamid, M. O. (2023). Teacher agency and washback: Insights from classrooms in rural Bangladesh. *ELT Journal*, 77(1), 52-61. <https://doi.org/10.1093/elt/ccac005>
- Alicia Kim, A., Yumsek, M., Kemp, J. A., Chapman, M., & Gary Cook, H. (2023). Universal tools activation in English language proficiency assessments: A comparison of Grades 1–12 English learners with and without disabilities. *Language Testing*, 40(4), 877-903. <https://doi.org/10.1177/02655322221149009>
- Allahyari, R., Abbasabady, M. M., Akhter, S., & Alibakhshi, G. (2023). EFL teachers' cognition of social and psychological consequences of high-stake national language tests: Role of teacher training workshops. *Language Testing in Asia*, 13, Article 54. <https://doi.org/10.1186/s40468-023-00262-0>
- Allal, L. (2021). Involving primary school students in the co-construction of formative

- assessment in support of writing. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 584–601. <https://doi.org/10.1080/0969594X.2021.1951164>
- Allen, D., & Koizumi, R. (2024). The English Speaking Achievement Test for Japanese Junior High School Students (ESAT-J): An analysis of public criticisms using the socio-cognitive framework. *Language Assessment Quarterly*, 21(4–5), 404–426. <https://doi.org/10.1080/15434303.2024.2438982>
- Allen, D., & Nakamura, K. (2023). The distribution of cognates and their impact on response accuracy in the EIKEN tests. *Language Testing*, 40(3), 771–795. <https://doi.org/10.1177/02655322231158551>
- Allen, D., & Tahara, T. (2021). A review of washback research in Japan. *Japanese Language Testing Association Journal*, 24, 3–22. https://doi.org/10.20622/jltajournal.24.0_3
- Allen, D., Yamamura, K., Meguro, S., & Nakamura, T. (2024). Washback from Four-Skills English Exams in the Japanese secondary school context. *Language Assessment Quarterly*, 21(4–5), 473–505. <https://doi.org/10.1080/15434303.2024.2440895>
- Almossa, S. Y., & Alzahrani, S. M. (2022). Assessment approaches of English language teachers in the Saudi higher education context. *Language Testing in Asia*, 12, Article 10. <https://doi.org/10.1186/s40468-022-00160-x>
- Almukhaild, H., & King, J. (2024). Scaling the apex of engagement: Anxious language learners' positive flow experiences during speaking tasks. *System*, 124, 103373. <https://doi.org/10.1016/j.system.2024.103373>
- Alpizar, D., Li, T., Norris, J. M., & Gu, L. (2023). Psychometric approaches to analyzing C-tests. *Language Testing*, 40(1), 107–132. <https://doi.org/10.1177/02655322211062138>
- Altherr Flores, J. A. (2021). The interplay of text and image on the meaning-making processes of adult L2 learners with emerging literacy: Implications for test design and evaluation frameworks. *Language Assessment Quarterly*, 18(5), 508–529. <https://doi.org/10.1080/15434303.2021.1984491>
- Alves-Wold, A., Walgermo, B. R., & Foldnes, N. (2024). Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses? *Assessing Writing*, 60, Article 100833. <https://doi.org/10.1016/j.asw.2024.100833>
- Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, 42(2), 292–314. <https://doi.org/10.1093/applin/amaa020>
- Amjadi, A., & Talebi, S. H. (2024). Extending the efficacy of Collaborative Strategic Reading (CSR) in teaching reading strategies: A socio-cultural emotional perspective. *Language Teaching Research*, 28(2), 714–748. <https://doi.org/10.1177/13621688211011501>
- An, H., & Li, S. (2024). Task-specific writing anxiety and self-efficacy are separate from general L2 writing anxiety and self-efficacy and they have differential associations with the effects of written corrective feedback in pre-task and within-task planning. *System*, 126, Article 103480. <https://doi.org/10.1016/j.system.2024.103480>
- An, N., Wright, C., & Wang, J. (2024). How does background knowledge affect second language reading? An eye movement study. *International Journal of Applied Linguistics*, 34(4), 1767–1789. <https://doi.org/10.1111/ijal.12601>
- Anderson, R. C., Chaparro, E. A., Smolkowski, K., & Cameron, R. (2023). Visual thinking

- and argumentative writing: A social-cognitive pairing for student writing development. *Assessing Writing*, 55, Article 100694.
<https://doi.org/10.1016/j.asw.2023.100694>
- Angelo, D., & Hudson, C. (2023). Phantom young English language learners: The shadowy presence of second language proficiency in (English medium) early childhood assessment. *Studies in Language Assessment*, 12(2), 1–27.
<https://doi.org/10.58379/KCDT5699>
- Antia, B. (2021). Multilingual examinations: Towards a schema of politicization of language in end of high school examinations in sub-Saharan Africa. *International Journal of Bilingual Education and Bilingualism*, 24(1), 138–153.
<https://doi.org/10.1080/13670050.2018.1450354>
- Antoniou, K., & Milaki, E. (2021). Irony comprehension in bidialectal speakers. *The Modern Language Journal*, 105(3), 697–719. <https://doi.org/10.1111/modl.12724>
- Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, 110, Article 102899.
<https://doi.org/10.1016/j.system.2022.102899>
- Appel, R., Arias, A., Baker, B., & Loignon, G. (2023). Insights from lexical and syntactic analyses of a French for academic purposes assessment. *Assessing Writing*, 58, Article 100789. <https://doi.org/10.1016/j.asw.2023.100789>
- Arai, Y. (2023). Comparability between graded readers and the Common Test in Japan in terms of text difficulty perceived by learners. *Language Assessment Quarterly*, 20(3), 277–295. <https://doi.org/10.1080/15434303.2023.2237497>
- Arce, A. J., & Young, M. J. (2022). Test efficacy: Refocusing validation from college exams to candidates. *International Journal of Testing*, 22(1), 100–119.
<https://doi.org/10.1080/15305058.2021.2019752>
- Archila, P. A., Molina, J., Danies, G., Truscott de Mejía, A.-M., & Restrepo, S. (2022). Using formal formative assessment (FFA) to promote undergraduates' bilingual written scientific argumentation (BWSA). *International Journal of Bilingual Education and Bilingualism*, 25(9), 3137–3150. <https://doi.org/10.1080/13670050.2021.2021386>
- Ardasheva, Y., Crosson, A. C., Carbonneau, K. J., & French, B. F. (2021). Unpacking contributions of morphosyntactic awareness and vocabulary to science reading comprehension among linguistically diverse students. *TESOL Quarterly*, 55(3), 931–965. <https://doi.org/10.1002/tesq.3039>
- Arias-Hermoso, R., Agirre, A. I., & Larrañaga, E. G. (2024). A comparison between input modalities and languages in source-based multilingual argumentative writing. *Assessing Writing*, 60, 100813. <https://doi.org/10.1016/j.asw.2024.100813>
- Arias, A., Vessey, R., & Sheyholislami, J. (2022). Jumping through hoops in the Canadian immigration system: A critical view of the immigrant's journey to citizenship. *TESOL Quarterly*, 56(4), 1430–1444. <https://doi.org/10.1002/tesq.3184>
- Arques, A. C., & Ferrero, C. L. (2023). Peer-feedback of an occluded genre in the Spanish language classroom: A case study. *Assessing Writing*, 57, Article 100756.
<https://doi.org/10.1016/j.asw.2023.100756>
- Aryadoust, V. (2023). The vexing problem of validity and the future of second language assessment. *Language Testing*, 40(1), 8–14.
<https://doi.org/10.1177/02655322221125204>
- Aryadoust, V. (2024). Topic and accent coverage in a commercialized L2 listening test: Implications for test-takers' identity. *Applied Linguistics*, 45(5), 765–785.

- <https://doi.org/10.1093/applin/amad062>
- Aryadoust, V. (2025). Generative Artificial Intelligence in assessment of listening. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from <https://meteor.springer.com/container/contribute.jsf?id=263244>
- Aryadoust, V., & Luo, L. (2023). The typology of second language listening constructs: A systematic review. *Language Testing*, 40(2), 375–409. <https://doi.org/10.1177/02655322221126604>
- Aryadoust, V., & Yichen, J. (2025). ChatGPT in assessment of writing. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from <https://meteor.springer.com/container/contribute.jsf?id=263243>
- Aryadoust, V., & Zakaria, A. (2025). ChatGPT in assessment of speaking. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from <https://meteor.springer.com/container/contribute.jsf?id=263242>
- Aryadoust, V., Foo, S., & Ng, L. Y. (2022). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Language Testing*, 39(1), 56–89. <https://doi.org/10.1177/02655322211026876>
- Aryadoust, V., Min, S., & Chen, X. (2024). Investigating differential item functioning across interaction variables in listening comprehension assessment. *Studies in Educational Evaluation*, 80, 101322. <https://doi.org/10.1016/j.stueduc.2024.101322>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6–40. <https://doi.org/10.1177/0265532220927487>
- Aryadoust, V., Ng, L. Y., Foo, S., & Esposito, G. (2022). A neurocognitive investigation of test methods and gender effects in listening assessment. *Computer Assisted Language Learning*, 35(4), 743–763. <https://doi.org/10.1080/09588221.2020.1744667>
- Aryadoust, V., Zakaria, A., & Jia, Y. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence*. Advance online publication. <https://doi.org/10.1016/j.caeai.2024.100204>
- Atak, N., & Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. *Assessing Writing*, 47, Article 100506. <https://doi.org/10.1016/j.asw.2020.100506>
- Aull, L. (2021). Directed Self-Placement: Subconstructs and group differences at a US university. *Assessing Writing*, 48, Article 100522. <https://doi.org/10.1016/j.asw.2021.100522>
- Azizi, Z. (2022). Fairness in assessment practices in online education: Iranian University English teachers' perceptions. *Language Testing in Asia*, 12, Article 14. <https://doi.org/10.1186/s40468-022-00164-7>
- Azizi, Z., & Ismail, S. M. (2023). What ethical requirements should be considered in language classroom assessment? Insights from high school students. *Language Testing in Asia*, 13, Article 23. <https://doi.org/10.1186/s40468-023-00235-3>

B

- Badham, L., & Furlong, A. (2023). Summative assessments in a multilingual context: What comparative judgment reveals about comparability across different languages in Literature. *International Journal of Testing*, 23:2, 111-134.
<https://doi.org/10.1080/15305058.2022.2149536>
- Baker, B. (2024). Introduction. In B. Baker and L. Taylor (Eds.), *Language assessment literacy and Competence volume 2: Case studies from around the world*. Cambridge.
- Baker, B. & Gebril, A. (2024). Editorial: The assessment of writing in languages other than English. *Assessing Writing*, 60, 1-2. <https://doi.org/10.1016/j.asw.2024.100840>.
- Baker, B. & Taylor, L. (2024). *Language assessment literacy and competence volume 2: Case studies from around the world*. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/717471-silt-volume-56.pdf>
- Baker, B. & Taylor, L. (2024). *Language assessment literacy and competence volume 1: Research and reflections from the field*. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- Baker, B., & Gebril, A. (2024). The assessment of writing in languages other than English (LOTE). *Assessing Writing*, 60, 100840. <https://doi.org/10.1016/j.asw.2024.100840>
- Baker, B., Arias, A., Bibeau, L. D., Qin, Y., Norenberg, M., & St-John, J. (2023). Rethinking student placement to enhance efficiency and student agency. *Language Testing*, 41(1), 181–191. <https://doi.org/10.1177/02655322231179128>
- Baker, B., Arias, A., Bibeau, L.-D., Qin, Y. (Coral), Norenberg, M., & St-John, J. (2024). Rethinking student placement to enhance efficiency and student agency. *Language Testing*, 41(1), 181–191. <https://doi.org/10.1177/02655322231179128>
- Baker, B., Polikar, S., & Homayounzadeh, M. (2021). “I was not thoughtful enough before:” Exploiting the learning-oriented-assessment potential of a test taker-oriented rubric for summative assessment. In A. Gebril. (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 162–181). Routledge.
<https://doi.org/10.4324/9781003014102-12>
- Balbino, T. R. T. & Moraes, I. T. (2023). Letramento em avaliação em contexto de línguas para pais e responsáveis por crianças [Language assessment literacy for children’s parents and guardians]. *Revista X*, 18(1), 5-38.
<https://doi.org/10.5380/rvx.v18i1.88026>
- Ball, J. (2023). Encouraging EAP students to engage with the literature of their disciplines. In A. Burn, & S. O’Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 61–72). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Banerjee, J., Lestari, S. B., & Rossi, O. (2021). Choosing test formats and task types. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.78–89). Routledge.
<https://doi.org/10.4324/9781351034784>
- Banitz, B. (2022). Language assessment in Mexico: Exploring university language teachers’ backgrounds, practices, and opinions. *System*, 110, Article 102898.
<https://doi.org/10.1016/j.system.2022.102898>
- Bao, G. (2023). Factor structures of speed and breakdown fluency in EFL learners’ story retelling performances. *International Review of Applied Linguistics in Language*

- Teaching*, 61(2), 631-654. <https://doi.org/10.1515/iral-2022-0125>
- Bao, G. (2024). Differential effects of EFL proficiency and mode on noun phrase use. *System*, 121, 103244. <https://doi.org/10.1016/j.system.2024.103244>
- Barkaoui, K. (2021). *Evaluating tests of second language development*. Peter Lang. <https://doi.org/10.3726/b18059>
- Barkaoui, K. (2023). Exploring task effects on register variation in second language learners' writing. *TASK: Journal on Task-Based Language Teaching and Learning*, 3, 109-139. <https://doi.org/10.1075/task.22009.bar>
- Barkaoui, K. (2024). Examining performance on an integrated writing task from a Canadian English language proficiency test. *Canadian Modern Language Review*, 80(2), 77-115. <https://doi.org/10.3138/cmlr-2023-0022>
- Barkaoui, K. (2024). Exploring the effects of task difficulty and learner variables on performance on picture description writing tasks. *Assessing Writing*, 60, 100827. <https://doi.org/10.1016/j.asw.2024.100827>
- Barkaoui, K. (2024). Multi-faceted Rasch analysis for evaluating second language tests. In Kunnan, A. (Ed.), *The concise companion to language assessment* (pp. 602-618). Wiley-Blackwell.
- Barkaoui, K. (2024). The academic achievement of undergraduate students with different score profiles on the TOEFL iBT: A replication study. *Language Assessment Quarterly*. <https://doi.org/10.1080/15434303.2024.2438142>
- Barkaoui, K. & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26-63. 2._sila_12_1__barkaoui_and_woodworth.pdf (altaanz.org)
- Barkaoui, K., & Hadidi, A. (2021). *Assessing changes in second Language writing performance*. Routledge. <https://doi.org/10.4324/9781003092346>
- Barkaoui, K., & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26–63. <https://doi.org/10.58379/QCFS2805>
- Barkaoui, K., Shi, L., & Riazi, M. (2021). "Follow your interests because those will motivate you to excel": An interview with Alister Cumming. *Language Assessment Quarterly*, 18(1), 56–75. <https://doi.org/10.1080/15434303.2020.1846190>
- Barker, B. (2023). EAL/D or an additional need? A cycle of learning, teaching and assessment to discern if an EAL/D student has additional needs to learning English. *Studies in Language Assessment*, 12(2), 272-285. <https://doi.org/10.58379/XSLO8306>
- Barnau, A. (2023). Testing English for medical purposes: The effects of traditional and distance education on learning outcomes. *Language Learning in Higher Education*, 13(2), 451-460. <https://doi.org/10.1515/cercles-2023-2027>
- Barnawi, O.Z., Alharbi, M.S., & Alzahrani, A.A. (Eds.). (2022). *Transnational English language assessment practices in the age of metrics*. Routledge. <https://doi.org/10.4324/9781003252382>
- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing*, 47, Article 100510. <https://doi.org/10.1016/j.asw.2020.100510>

- Barrows, J., & Cox, T. (2021). The effect of prompt accent on elicited imitation assessments in English as a second language. *Papers in Language Testing and Assessment*, 10(2), 1–29.
http://www.altanz.org/uploads/5/9/0/8/5908292/2._plta_10_2__barrows_cox.pdf
- Bastola, M. N., & Hu, G. (2024). Engagement with supervisory feedback on master's theses: Do supervisors and students see eye to eye? *Assessing Writing*, 60, 100841. <https://doi.org/10.1016/j.asw.2024.100841>
- Batista, E. G. & Moraes, I. T. (2020). Avaliação orientada para a aprendizagem no ensino de línguas para crianças [Learning-oriented language assessment in the context of young learners]. *REVELLI*, 12, 1-19. <https://doi.org/10.51913/revelli.v12i0.10234>
- Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 38(4), 511–535. <https://doi.org/10.1177/0265532220951504>
- Batty, A. O. (2021). Shūjukudo to seibetsu ga bideo L2 chōkairyoku tesuto ni okeru shikaku no kōka ni ataeru eikyō [The impact of proficiency and gender on the effect of visuals in video L2 listening comprehension tests]. *Journal of Foreign Language Education*, 17, 1–18.
- Batty, A. O., Haug, T., Ebling, S., Tissi, K., & Sidler-Miserez, S. (2023). Challenges in rating signed production: A mixed-methods study of a Swiss German Sign Language form-recall vocabulary test. *Language Testing*, 40(2), 352-374.
<https://doi.org/10.1177/02655322221122774>
- Baxter, G., & Neumann, H. (2023). Source-based writing of the high- and low-proficiency adolescent writer in the high-school L2 classroom. *Journal of Second Language Writing*, 62, Article 101064. <https://doi.org/10.1016/j.jslw.2023.101064>
- Bazvand, A. D. (2023). Exploring student teachers' perceptions of assessment ethics across university-based teacher education programs in Iran. *Language Testing in Asia*, 13, Article 9. <https://doi.org/10.1186/s40468-022-00205-1>
- Belda-Medina, J., & Kokošková, V. (2024). Assessing Pre-Service Teachers' Knowledge and Attitudes Towards Language Testing for Citizenship Through LAMAS (Language Accreditation for Migration Attitudes Scale). *Language Assessment Quarterly*, 21(4–5), 427–451. <https://doi.org/10.1080/15434303.2024.2438995>
- Bell, D. E. (2024). EAP materials and EAP assessment. In D. E. Bell (Ed.), *English for academic purposes: Perspectives on the past, present and future* (pp. 105-123). Multilingual Matters. <https://doi.org/10.2307/jj.21995548.11>
- Bell, T. R., & Borden, R. S. (2022). Examining target language proficiency among in-service K-12 language teachers. *Foreign Language Annals*, 55 (3), 815–831.
<https://doi.org/10.1111/flan.12637>
- Beltrán-Palanques, V. (2024). Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP. *Assessing Writing*, 60, 100809.
<https://doi.org/10.1016/j.asw.2024.100809>
- Bennett, R. E., Zhang, M., Sinharay, S., Guo, H., & Deane, P. (2022). Are there distinctive profiles in examinee essay-writing processes? *Educational Measurement: Issues and Practice*, 41(2), 55–69. <https://doi.org/10.1111/emip.12469>
- Berberovic, V. (2022). *Implementing peer feedback for writing tasks* (Research Notes Issue 83). Cambridge University Press and Assessment.
<https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Berger, A. (2023). A Difficulty-Informed Approach to developing Language Assessment Literacy for classroom purposes. *Chinese Journal of Applied Linguistics*, 46(2),

- 289-307. <https://doi.org/10.1515/CJAL-2023-0209>
- Betts, J., Muntean, W., Kim, D., & Kao, S. (2022). Evaluating different scoring methods for multiple response items providing partial credit. *Educational and Psychological Measurement*, 82(1), 151–176. <https://doi.org/10.1177/0013164421994636>
- Bijani, H., Hashempour, B., Ibrahim, K. A. A., Orabah, S. S. B., & Heydarnejad, T. (2022). Investigating the effect of classroom-based feedback on speaking assessment: A multifaceted Rasch analysis. *Language Testing in Asia*, 12, Article 26. <https://doi.org/10.1186/s40468-022-00176-3>
- BijanKhan, M., ShayesteFar, P., & Mohebbi, H. (2023). Assessing the content typicality and construct of Persian language proficiency test (PLPT) for non-Persian speakers: A corpus-informed study. *Language Testing in Asia*, 13, Article 11. <https://doi.org/10.1186/s40468-023-00217-5>
- Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: an exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14, Article 45. <https://doi.org/10.1186/s40468-024-00322-z>
- Binti Hasrol, S., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), Article 100023. <https://doi.org/10.1016/j.rmal.2022.100023>
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing*, 40(1), 133-152. <https://doi.org/10.1177/02655322211064629>
- Boggs, J. A., & Manchón, R. M. (2023). Feedback literacy in writing research and teaching: Advancing L2 WCF research agendas. *Assessing Writing*, 58, Article 100786. <https://doi.org/10.1016/j.asw.2023.100786>
- Bøhn, H. & Tsagari, D. (2022) Language assessment literacy: Understanding the construct from Norwegian EFL teachers' perspective. *Studies in Language Assessment*, 11(1), 119–148.
- Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>
- Bolender, B., Foster, C., & Vispoel, S. (2023). The criticality of implementing principled design when using AI technologies in test development. *Language Assessment Quarterly*, 20(4-5), 512-519. <https://doi.org/10.1080/15434303.2023.2288266>
- Bondar, N. A., Day, S. X., Mountain, L., Turchi, L. B., & Hutchison, L. F. (2023). Composition Organization and Development Analysis (CODA) Scale: Equipping high school students to evaluate argumentative essays. *Assessing Writing*, 57, Article 100724. <https://doi.org/10.1016/j.asw.2023.100724>
- Bonvin, A., Brugger, L., & Berthele, R. (2023). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*, 61(2), 257-285. <https://doi.org/10.1515/iral-2020-0093>
- Bormanaki, H. B., & Ajideh, P. (2022). Item performance across native language groups on the Iranian National University Entrance English Exam: A nationwide study. *Language Testing in Asia*, 12, Article 29. <https://doi.org/10.1186/s40468-022-00185-2>
- Bourdeaud'Hui, H., Aesaert, K., & van Braak, J. (2021). Exploring the validity of a

- comprehensive listening test to identify differences in primary school students' listening skills. *Language Assessment Quarterly*, 18(3), 228–252.
<https://doi.org/10.1080/15434303.2020.1860059>
- Bouwer, R., Koster, M., & Van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*, 30(3-4), 302-319.
<https://doi.org/10.1080/0969594X.2023.2241656>
- Bowles, M. A. (2022). Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, 39(3), 355–376. <https://doi.org/10.1177/02655322221076033>
- Braund, H., DeLuca, C., Panadero, E., & Cheng, L. (2021). Exploring formative assessment and co-regulation in kindergarten through interviews and direct observation. *Frontiers in Education: Educational Psychology*, 6, Article 732373.
<https://doi.org/10.3389/educ.2021.732373>
- Britton, E. R. (2023). Developing teacher feedback literacy through self-study: Exploring written commentary in a critical language writing curriculum. *Assessing Writing*, 56, Article 100709. <https://doi.org/10.1016/j.asw.2023.100709>
- Britton, E. R., & Kraemer, A. Assessing Languages Across the Curriculum: Critical reflection and student learning perspectives. *Foreign Language Annals*, 57(4), 1051-1071.
<https://doi.org/10.1111/flan.12775>
- Brookhart, S. M. (2023). Assessment Literacy in a better assessment future. *Chinese Journal of Applied Linguistics*, 46(2), 162-179.
<https://doi.org/10.1515/CJAL-2023-0202>
- Brown, J. D., Panahi, A., & Mohebbi, H. (2023). James Dean Brown's 50 years of work in language testing: A systematic review. *Language Teaching Research Quarterly*, 37, 4–75. <https://doi.org/10.32038/ltrq.2023.37.02>
- Browne, E. (2023). Making room for learning and assessing together in bilingual, bicultural teaching teams at a remote Warlpiri school in Central Australia. *Studies in Language Assessment*, 12(2), 151-167. <https://doi.org/10.58379/JKSA5503>
- Bruce, E., & Clark, T. (2023). IELTS Online: Overview. *Research Reports*.
<https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/ielts-online-overview-bruce-et-al-2023.pdf>
- Bruce, E., Dunn, K., & Clark, T. (2023). Review of recommended two-year limit on IELTS validity: A summary of preliminary internal research. *Research Reports*.
<https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/review-of-recommended-2-year-limit-on-ielts-validity-bruce-et-al-2023.pdf>
- Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, 40(1), 15-23. <https://doi.org/10.1177/02655322221127896>
- Brunfaut, T., & Harding, L. (2024). Diagnostic approaches in teaching and assessing listening. In E. Wagner, A. O. Batty, & E. Galaczi (Eds.), *The Routledge Handbook of Second Language Acquisition and Listening* (pp. 255-267). (Routledge Handbooks in Second Language Acquisition). Routledge.
<https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Listening/Wagner-Batty-Galaczi/p/book/9781032113647>
- Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners' reading comprehension: Exploring the effects of working

- memory, grade level, and reading task. *Language Testing*, 38(3), 356–377.
<https://doi.org/10.1177/0265532221991480>
- Brunsting, N., Yu, Q., Smart, J., and Bingham, W.P. (2022). Investigating linkages between international students' English language proficiency, social-contextual outcomes, and wellbeing in U.S. universities. IELTS Research Reports Online Series, No. 4/22. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/-/media/research-reports/brunsting-yu-smart-bingham-2022.ashx>
- Buchanan, B. J., Pirone, J. S., Pudans-Smith, K. K., Leffler, B. W., Cohen, B. S., Simmons, P. K., Ballentine, P. M., Venable, T., & Soza, K. (2023). American Sign Language level 1 assessment: What is out there? *Foreign Language Annals*, 56(4), 912–926.
<https://doi.org/10.1111/flan.12717>
- Buffington, J., Demos, A., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, 43(3), 635–662.
<https://doi.org/10.1017/S0272263121000127>
- Bui, H. P. (2023). Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment. *Language Testing in Asia*, 13, Article 10 (2023). <https://doi.org/10.1186/s40468-023-00220-w>
- Bullock, N., & Westbrook, C. (2021). Testing in ESP: Approaches in aviation and maritime English. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world*. (67–77). Springer.
https://doi.org/10.1007/978-981-33-4232-3_7
- Bulut, H. C., Bulut, O., & Arian, S. (2023). Evaluating group differences in online reading comprehension: The impact of item properties. *International Journal of Testing*, 23(1), 10-33. <https://doi.org/10.1080/15305058.2022.2044821>
- Burns, A. (2023). Action research for sustained and powerful professional learning. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 5–14). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Burns, A., O'Keefe, S., & Savage, J. (2024). *Research notes 87: Findings of action research in ELICOS program 2023*. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/714674-research-notes-87.pdf>
- Burton, J. D. (2021). The face of communication breakdown: Multimodal repair in L2 oral proficiency interviews. *Papers in Language Testing and Assessment*, 10(2), 30–61.
http://www.altanz.org/uploads/5/9/0/8/5908292/3._plta_10_2__burton.pdf
- Burton, J. D. (2023). Gazing into cognition: Eye behavior in online L2 speaking tests. *Language Assessment Quarterly*, 20(2), 190-214.
<https://doi.org/10.1080/15434303.2022.2143680>
- Burton, J. D. (2023). Reflections on the past and future of language testing and assessment: An emerging scholar's perspective. *Language Testing*, 40(1), 24-30.
<https://doi.org/10.1177/02655322221126607>
- Burton, J. D. (2024). Evaluating the impact of nonverbal behavior on language ability ratings. *Language Testing*, 41(4), 729–758.
<https://doi.org/10.1177/02655322241255709>
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3),

- 429–455. <https://doi.org/10.1177/0265532221992274>
- Buzick, H. M., Wolf, M. K., & Ballard, L. (2024). Achievement and growth on English language proficiency and content assessments for English learners in elementary grades. *Educational Measurement: Issues and Practice*, 43(1), 83-95. <http://doi.org/10.1111/emip.12588>

C

- Cai, H., & Yan, X. (2024). Examining the direct and indirect impacts of verbatim source use on linguistic complexity in integrated argumentative writing assessment. *Assessing Writing*, 61, 100868. <https://doi.org/10.1016/j.asw.2024.100868>
- Cai, H., & Yan, X. (2024). Triangulating natural language processing (NLP)-based analysis of rater comments and many-facet Rasch measurement (MFRM): An innovative approach to investigating raters' application of rating scales in writing assessment. *Language Testing*, 41(2), 384–411. <https://doi.org/10.1177/02655322231210231>
- Cai, J., Zhang, X., & Wang, Y. (2024). Effects of processing tasks on children's learning of second language words. *System*, 125, 103435. <https://doi.org/10.1016/j.system.2024.103435>
- Cai, Y., & Chen, H. (2022). The fluctuating effect of thinking on language performance: New evidence for the Island Ridge Curve. *Language Assessment Quarterly*, 19(5), 465–479. <https://doi.org/10.1080/15434303.2022.2080553>
- Çakmak, F., Ismail, S. M., & Karami, S. (2023). Advancing learning-oriented assessment (LOA): Mapping the role of self-assessment, academic resilience, academic motivation in students' test-taking skills, and test anxiety management in Telegram-assisted-language learning. *Language Testing in Asia*, 13, Article 20. <https://doi.org/10.1186/s40468-023-00230-8>
- Camara, W. J., & Mattern, K. (2022). Inflection point: The role of testing in admissions decisions in a postpandemic environment. *Educational Measurement: Issues and Practice*, 41(1), 10–15. <https://doi.org/10.1111/emip.12493>
- Campbell, M. & Rosenhan, C. (2024). Assessment and creativity through a translingual lens: transdisciplinary insights. *Applied Linguistics Review*, 15(5), 1839-1859. <https://doi.org/10.1515/applirev-2023-0090>
- Canals, E. (2023). Modified output and metalanguage during conversational interaction: A qualitative look at interactional feedback. *Studies in Second Language Learning and Teaching*, 13(3), 601-626. <https://doi.org/10.14746/ssllt.31128>
- Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, Article 102464. <https://doi.org/10.1016/j.system.2021.102464>
- Canning, D., McLean, S., & Vitta, J. (2024). Relative complexity in a model of word difficulty: The role of loanwords in vocabulary size tests. *Studies in Second Language Learning and Teaching*, 14(4), 631-659. <https://doi.org/10.14746/ssllt.38492>
- Cardwell, R. L., Nydick, S. W., Lockwood, J. R., & von Davier, A. A. (2024). Practical considerations when building concordances between English tests. *Language Testing*, 41(1), 192–202. <https://doi.org/10.1177/02655322231195027>
- Carey, M. D., & Szocs, S. (2024). Revisiting raters' accent familiarity in speaking tests : Evidence that presentation mode interacts with accent familiarity to variably affect comprehensibility ratings. *Language Testing*, 41(2), 290–315.

- <https://doi.org/10.1177/02655322231200808>
- Carlsen, C. H., & Rocca, L. (2021). Language test misuse. *Language Assessment Quarterly*, 18(5), 477–491. <https://doi.org/10.1080/15434303.2021.1947288>
- Carlsen, C. H., & Rocca, L. (2022). Language test activism. *Language Policy*, 21(4), 597–616. <https://doi.org/10.1007/s10993-022-09614-7>
- Carney, N. (2021). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*, 55(2), 536–567. <https://doi.org/10.1002/tesq.3000>
- Carney, N. (2024). Referencing word and phrase lists in an online vocabulary notebook. *Language Teaching Research*, 28(1), 253–272. <https://doi.org/10.1177/1362168821997613>
- Carr, N. (2023). Feedback on writing through the lens of activity theory: An exploration of changes to peer-to-peer interactions. *Assessing Writing*, 56, Article 100720. <https://doi.org/10.1016/j.asw.2023.100720>
- Carrió-Pastor, M. L. (2021). The assessment of metadiscourse devices in English as a foreign language. *Assessing Writing*, 50, Article 100560. <https://doi.org/10.1016/j.asw.2021.100560>
- Castillo, P. A. P., Soto, C., Asún, R. A., & Gutiérrez, F. (2023). Profiling support in literacy development: Use of natural language processing to identify learning needs in higher education. *Assessing Writing*, 58, Article 100787. <https://doi.org/10.1016/j.asw.2023.100787>
- Cen, Y., & Zheng, Y. (2024). The motivational aspect of feedback: A meta-analysis on the effect of different feedback practices on L2 learners' writing motivation. *Assessing Writing*, 59, 100802. <https://doi.org/10.1016/j.asw.2023.100802>
- Chalhoub-Deville, M., & O'Sullivan, B. (2020). *Validity: Theoretical development and integrated arguments*. University of Toronto Press.
- Chambers, L., Vitello, S., & Vidal Rodeiro, C. (2024). Moderation of non-exam assessments: a novel approach using comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 31(1), 32–55. <https://doi.org/10.1080/0969594X.2024.2313237>
- Chan, J. Y. H. (2021). Bridging the gap between ELF and L2 learners' use of communication strategies: Rethinking current L2 assessment and teaching practices. *System*, 101, Article 102609. <https://doi.org/10.1016/j.system.2021.102609>
- Chan, J. Y. H. (2022). The evolution of assessment in English pronunciation: The case of Hong Kong (1978–2018). *Language Assessment Quarterly*, 19(1), 1–26. <https://doi.org/10.1080/15434303.2021.1935960>
- Chan, K. K. Y., Bond, T., & Yan, Z. (2023). Application of an Automated Essay Scoring engine to English writing assessment using Many-Facet Rasch Measurement. *Language Testing*, 40(1), 61–85. <https://doi.org/10.1177/02655322221076025>
- Chan, S., & May, L. (2023). Towards more valid scoring criteria for integrated reading-writing and listening-writing summary tasks. *Language Testing*, 40(2), 410–439. <https://doi.org/10.1177/02655322221135025>
- Chan, S., & Yamashita, J. (2022). Integrated writing and its correlates: A meta-analysis. *Assessing Writing*, 54, Article 100662. <https://doi.org/10.1016/j.asw.2022.100662>
- Chang, H. (2022). Evaluation of language and teaching skill domains for international teaching assistants: An approach based on invariant measurement. *Language Assessment Quarterly*, 19(3), 264–288.

- <https://doi.org/10.1080/15434303.2021.2013487>
- Chang, T. (2023). A move analysis of Chinese L2 student essays from the sociocognitive perspective: Genres, languages, and writing quality. *Assessing Writing*, 57, Article 100750. <https://doi.org/10.1016/j.asw.2023.100750>
- Chang, Y.-T., Choe, A. T., Holden, D., & Isbell, D. R. (2024). Making each point count: Revising a local adaptation of the Jacobs et al.'s (1981) ESL COMPOSITION PROFILE rubric. *Language Testing*, 41(2), 443–455. <https://doi.org/10.1177/02655322231217979>
- Chapelle, C. A. & Lee, H. (2022). Validation of second language spoken assessments. In T. Haug, W. Mann & U. Knoch (Eds.), *The handbook of language assessment across modalities* (pp. 273–284). Oxford University Press. <https://doi.org/10.1093/oso/9780190885052.003.0023>
- Chapelle, C. A., & Lee, H. (2022). Conceptions of validity. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 17–31). Routledge. <https://doi.org/10.4324/9781003220756-3>
- Chapelle, C. A., & Lee, H.-W. (2021). Understanding argument-based validity in language testing. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 19–44). Cambridge University Press. <https://doi.org/10.1017/9781108669849>
- Chapelle, C. A., & Ockey, G. J. (2024). Open Science in language assessment research contexts: A reply to Winke. *Language Testing*, 41(4), 882–885. <https://doi.org/10.1177/02655322241239377>
- Chapelle, C. A., & Voss, E. (Eds.). (2021). *Validity argument in language testing: Case studies of argument-based validation research*. Cambridge University Press. <https://doi.org/10.1017/9781108669849>
- Chapelle, C. A., Hauser, P. C., Lee, H., Rathmann, C., & Schönström, K. (2022). Discussion of validation issues in signed and spoken assessments for adult L2 learners. In T. Haug, W. Mann & U. Knoch (Eds.), *The handbook of language assessment across modalities* (pp. 295–298). Oxford University Press. <https://doi.org/10.1093/oso/9780190885052.003.0025>
- Chapelle, C. A., Voss, E., & Kim, H. (2024). Designing evaluations for validation of language assessments. In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment*. (pp. 67–79). Wiley-Blackwell.
- Chapman, M., Kemp, J. A., Kim, A. A., MacGregor, D., & MacMillan, F. Testing young multilingual learners remotely: Prioritizing equity, fairness, and social justice. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 153–170). Routledge. <https://doi.org/10.4324/9781003292395-13>
- Chappuis, S., Brookhart, S., & Chappuis, J. (2021). *Ten assessment literacy goals for school leaders*. SAGE.
- Chedid, G., & Stephan, M. (2024). Adaptation and norm determination of the Boston Naming Test for healthy Lebanese adults aged between 50 and 88 years. *Language Testing in Asia*, 14, Article 25. <https://doi.org/10.1186/s40468-024-00294-0>
- Chen Hsieh, J. (2021). Digital storytelling outcomes and emotional experience among middle school EFL Learners: Robot-assisted versus PowerPoint-assisted mode. *TESOL Quarterly*, 55(3), 994–1010. <https://doi.org/10.1002/tesq.3043>
- Chen, H., Cai, Y., & de la Torre, J. (2023). Investigating second language (L2) reading

- subskill associations: A cognitive diagnosis approach. *Language Assessment Quarterly*, 20(2), 166-189. <https://doi.org/10.1080/15434303.2022.2140050>
- Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*, 57, Article 100915. <https://doi.org/10.1016/j.jslw.2022.100915>
- Chen, T.-H. (2024). The effects of task complexity on L2 English rapport-building language use and its relationship with paired speaking test task performance. *Applied Linguistics Review*, 15(2), 737–769. <https://doi.org/10.1515/applirev-2021-0199>
- Chen, X., Aryadoust, V., & Zhang, W. (2024). A systematic review of differential item functioning in second language assessment. *Language Testing*, 0(0). <https://doi.org/10.1177/02655322241290188>
- Cheng, L. (2024). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–1146). Chichester, UK: John Wiley & Sons. doi:10.1002/9781118411360.wbcla071. <https://onlinelibrary.wiley.com/doi/10.1002/9781118411360.wbcla071>
- Cheng, L. & Sultana, N. (2022). Washback: Looking backward and forward. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp 136–152). <https://doi.org/10.4324/9781003220756-12>
- Cheng, L., & Im, G.-H. (2021). Translation of surveys for international research: Beyond equivalence. In G. Harbaugh & U. Luhanga (Eds.). *Basic elements of survey research in education: Addressing the problems your advisor never told you about* (pp. 219–236). Information Age Publishing.
- Cheng, L., & Zheng, Y. (2021). Measuring anxiety. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 177–186). <https://doi.org/10.4324/9781351034784-20>
- Cheng, L., Im, G.-H., & Jabeen, R. (2021). Whose English(es). are we assessing and by whom? In B. Lantaigne, C. Coombe., & J.D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp. 169–179). Springer. https://doi.org/10.1007/978-981-33-4232-3_12
- Cheng, Y. (2023). Concerto Software for Computerized Adaptive Testing - Free Version. *Measurement: Interdisciplinary Research and Perspectives*, 21(3), 194–202. <https://doi.org/10.1080/15366367.2023.2187274>
- Cheong, C. M., Liu, Y., & Mu, R. (2024). Characteristics of students' task representation and its association with argumentative integrated writing performance. *Assessing Writing*, 60, 100845. <https://doi.org/10.1016/j.asw.2024.100845>
- Cheong, C. M., Yao, Y., & Zhang, J. (2023). Growth mindset and emotions in tandem: Their effects on L2 writing performance based on writers' proficiency levels. *Assessing Writing*, 58, Article 100785. <https://doi.org/10.1016/j.asw.2023.100785>
- Chernova, S., & Abbasi, M. (2022). Using ipsative assessment in teaching and learning English grammar through e-portfolios (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Chinpakdee, M., & Gu, P. Y. (2024). The impact of explicit strategy instruction on EFL

- secondary school learners' reading. *Language Teaching Research*, 28(1), 296–319. <https://doi.org/10.1177/1362168821994157>
- Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487–2512. <https://doi.org/10.1177/13621688211046740>
- Choi, I. & Zu, J. (2022). The impact of using synthetically generated listening stimuli on test-taker performance: A case study with multiple-choice, single-selection items. *ETS Research Report Series*, 2022(1), 1–14. <https://doi.org/10.1002/ets2.12347>
- Choi, I., & Deane, P. (2021). Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly*, 18(2), 107–132. <https://doi.org/10.1080/15434303.2020.1804913>
- Choi, I., Hao, J., Deane, P., & Zhang, M. (2021). Benchmark keystroke biometrics accuracy from high-stakes writing tasks. *ETS Research Report Series*, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12326>
- Choi, J. S., & Loewen, S. (2022). Exploring young learners' strategic behaviors in a speaking test. *TESOL Quarterly*, 56(4), 1384–1396. <https://doi.org/10.1002/tesq.313>
- Choi, S., & Kang, O. (2023). The roles of suprasegmental features in assessing paired speaking tasks in high-stakes language assessment. *System*, 119, Article 103183. <https://doi.org/10.1016/j.system.2023.103183>
- Choi, Y. (2021). What interpretations can we make from scores on graphic-prompt writing (GPW) tasks? An argument-based approach to test validation. *Assessing Writing*, 48, Article 100523. <https://doi.org/10.1016/j.asw.2021.100523>
- Choi, Y. D. (2022). Validity of score interpretations on an online English placement writing test. *Language Testing in Asia*, 12, Article 42. <https://doi.org/10.1186/s40468-022-00187-0>
- Choi, Y., & Chun, J. (2022). Test review: French examination of the College Scholastic Ability Test in Korea. *Language Testing in Asia*, 12, Article 49. <https://doi.org/10.1186/s40468-022-00199-w>
- Choi, Y., Vo, S., & Ockey, G. J. (2024). Investigation into the factor structure of a local English placement test of oral communication. *International Journal of Applied Linguistics*, 34(1), 63–85. <https://doi.org/10.1111/ijal.12478>
- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, Article 102408. <https://doi.org/10.1016/j.system.2020.102408>
- Chong, J. J. Q., & Aryadoust, V. (2023). Investigating the effect of multimodality and sentiments on speaking assessments: A facial emotional analysis. *Education and Information Technologies*, 28, 7413–7436. <https://doi.org/10.1007/s10639-022-11478-7>
- Chong, S. W., & Isaacs, T. (2023). An ecological perspective on classroom-based assessment. *TESOL Quarterly*, 57(4), 1558–1570. <http://doi.org/10.1002/tesq.3201>
- Chong, S. W., & Reinders, H. (Eds.). (2023). *Innovation in learning-oriented language assessment*. Springer.
- Chong, S. W., Isaacs, T., & McKinley, J. (2023). Research timeline: Ecological systems theory and L2 research. *Language Teaching*, 56(3), 333–348. <http://doi.org/10.1017/S0261444822000283>
- Chou, M. H. (2021). An integrated approach to developing and assessing EFL students'

- speaking ability and strategy use. *Language Education & Assessment*, 4(1), 19–37. <https://doi.org/10.29140/lea.v4n1.428>
- Chovancová, B., Bilová, Š., & Hradilová, A. (2023). Turning the tables on online exam cheating via language mediation tasks. *Language Learning in Higher Education*, 13(2), 517–533. <https://doi.org/10.1515/cercles-2023-2033>
- Christensen, L. L., Shyyan, V. V., & MacMillan, F. (2023). Toward a systematic accessibility review process for English language proficiency tests for young learners. *Language Testing*, 40(4), 856–876. <https://doi.org/10.1177/02655322231168386>
- Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, 60, Article 101003. <https://doi.org/10.1016/j.jslw.2023.101003>
- Chuang, P., & Yan, X. (2022). An investigation of the relationship between argument structure and essay quality in assessed writing. *Journal of Second Language Writing*, 56, Article 100892. <https://doi.org/10.1016/j.jslw.2022.100892>
- Chukharev-Hudilainen, E., & Ockey, G. J. (2021). The development and evaluation of interactional competence elicitor for oral language assessments. *ETS Research Report Series*, 2021(1), 1–20. <https://doi.org/10.1002/ets2.12319>
- Clark, T., & Bruce, E. (2024). Open Science should be welcomed by test providers but grounded in pragmatic caution: A response to Winke. *Language Testing*, 41(4), 872–876. <https://doi.org/10.1177/02655322231223105>
- Clark, T., & Yu, G. (2021). Beyond the IELTS test: Chinese and Japanese postgraduate UK experiences. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1512–1530. <https://doi.org/10.1080/13670050.2020.1829538>
- Clark, T., & Yu, G. (2024). The pedagogical remit of test preparation: the case of writing acquisition on an IELTS course. *Applied Linguistics Review*, 15(3), 1009–1045. <https://doi.org/10.1515/applirev-2021-0164>
- Clark, T., & Endres, H. (2021). Online diagnostic grammar testing for high school students—An international trial. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 602–632. <https://doi.org/10.1080/0969594X.2021.1970513>
- Clark, T., Holland, M., & Spiby, R. (2023). Seeking empirical evidence to support online test validation: Building on the IELTS indicator assessment model. In K. Sadeghi (Ed), *Technology-assisted language assessment in diverse contexts* (pp. 16–33). Routledge.
- Clark, T., Spiby, R., & Tasviri, R. (2021). Crisis, collaboration, recovery: IELTS and COVID-19. *Language Assessment Quarterly*, 18(1), 17–25. <https://doi.org/10.1080/15434303.2020.1866575>
- Clews, P. (2022). *Towards a peer feedback scaffold* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Cliff, A. (2021). The use of mediation and feedback in a standardised test of academic literacy: Theoretical and design considerations. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 75–92). Multilingual Matters.
- Cobb, T., & Laufer, B. (2021). The nuclear word family list: A list of the most frequent family members, including base and affixed words. *Language Learning*, 71(3), 834–871. <https://doi.org/10.1111/lang.12452>

- Cohen, A. D., Rahmati, T., & Sadeghi, K. (2023). Test-taking strategies in technology-assisted language assessment. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 235-254). Routledge.
<https://doi.org/10.4324/9781003292395-19>
- Colman, A., Segers, W., & Verplaetse, H. (2023). Extending Preselected Items Evaluation (PIE) with the ATA (American Translators Association) error categories for objective and targeted translation evaluation. *Journal of Applied Linguistics and Professional Practice*, 17(3), 274-293. <https://doi.org/10.1558/jalpp.21681>
- Coniam, D., Lee, T., Milanovic, M., Pike, N., & Zhao, W. (2022). Role of expert judgement in language test validation. *Language Education and Assessment*, 5(1), 18–33.
<https://doi.org/10.29140/lea.v5n1.769>
- Consolini, C. H., & Kyle, K. (2024). The relationship between L2 Spanish proficiency and features of written lexical and lexicogrammatical use. *Applied Linguistics*, 45(6), 992–1009. <https://doi.org/10.1093/applin/amae032>
- Cook, B., Luke, J., Valoe, A., & Barkaoui, K. (2021). Institutional language policy and ESL teachers' conceptions of L2 writing and classroom assessment practices. *Canadian Modern Language Review*, 77(2), 93–109. <https://doi.org/10.3138/cmlr-2020-0013>
- Coombe, T., Clark, T., & Mohebbi, H. (2023). Volume Introduction: Special Issue in Honour of James Dean Brown's Five-Decade Contribution to Language Testing and Assessment. *Language Teaching Research Quarterly*, 37, 1-3.
- Cox, T. L., Brown, A. V., & Thompson, G. L. (2023). Temporal fluency and floor/ceiling scoring of intermediate and advanced speech on the ACTFL Spanish Oral Proficiency Interview–computer. *Language Testing*, 40(2), 325-351.
<https://doi.org/10.1177/02655322221114614>
- Crosby, S. (2023). A fair go: Translanguaging and assessment practices in a New Zealand junior college. *Studies in Language Assessment*, 12(2), 117-150.
- Crossley, S. A., Baffour, P., Tian, Y., Franklin, A., Benner, M., & Boser, U. (2023). A large-scale corpus for assessing written argumentation: PERSUADE 2.0. *Assessing Writing*, 61, 100865. <https://doi.org/10.1016/j.asw.2024.100865>
- Crowther, D., Isbell, D. R., & Nishizawa, H. (2023). Second language speech comprehensibility and acceptability in academic settings: Listener perceptions and speech stream influences. *Applied Psycholinguistics*, 44(5), 858–888.
<https://doi.org/10.1017/S0142716423000346>
- Crusan, D. (2023). Shaping teachers' writing assessment practices: The effect of teacher beliefs. In R. Kohls & C. P. Casanave (Eds.), *Perspectives on good writing in Applied Linguistics and TESOL* (pp. w60-74). University of Michigan Press.
<http://doi.org/10.3998/mpub.11930285>
- Crusan, D. (2024). Ungrading: Revolution or evolution. *Journal of Second Language Writing*, 66, 1-7. <https://doi.org/10.1016/j.jslw.2024.101149>
- Crusan, D., Masjedlou, A. P., & Mohebbi, H. (2023). Writing teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.) *Insights into teaching learning writing: A practical guide for teachers* (pp. 45-55). Castledown.
<http://doi.org/10.29140/9781914291159>
- Crusan, D., Panahi, A., & Mohebbi, H. (2023). Writing teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.), *Insights into teaching and*

- learning writing: A practical guide for early-career teachers* (pp. 45-55). Castledown. <https://castledown.online/reference/9781914291159/>
- Cui, Y., Jin, H., & Gao, Y. (2023). Developing EFL teachers' feedback literacy for research and publication purposes through intra- and inter-disciplinary collaborations: A multiple-case study. *Assessing Writing*, 57, Article 100751. <https://doi.org/10.1016/j.asw.2023.100751>
- Cui, Y., Liu, Y., Yu, H., & Gao, Y. (2022). Developing English teachers' language assessment literacy in an EAP reform context through test design: A case study. *System*, 109, 102866. <https://doi.org/10.1016/j.system.2022.102866>
- Curtis, A. & Cheng, L. (2024). The Complexities of Diversity: A Personal Journey Across Time and Space. *The Language Teacher*, 48(5), 14-17 <https://doi.org/10.37546/JALTTLT48.5-3>
- Cushing, S. T., Ren, H., & Tan, Y. (2024). The use of TOEFL iBT® in admissions decisions: Stakeholder perceptions of policies and practices. ETS Research Report Series, 2024(1), 1–16. <https://doi.org/10.1002/ets2.12375>

D

- Dai, D. W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test*. Peter Lang. <https://doi.org/10.3726/b21295>
- Dalal, D. K., Randall, J. G., Cheung, H. K., Gorman, B. C., Roch, S. G., & Williams, K. J. (2022). Is there bias in alternatives to standardized tests? An investigation into letters of recommendation. *International Journal of Testing*, 22(1), 21-42. <https://doi.org/10.1080/15305058.2021.2019751>
- Daller, M., Müller, A., & Wang-Taylor, Y. (2021). The C-test as predictor of the academic success of international students. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1502–1511. <https://doi.org/10.1080/13670050.2020.1747975>
- DAMACENA, M., & QUEVEDO-CAMARGO, G. (2021) Avaliação e formação de professores de línguas [Assessment and language teacher education]. *Revista Olhares e Trilhas*, v. 23, p. 1054-1073. DOI: <https://doi.org/10.14393/OT2021v23.n.3.62193>
- Dang, C. N., & Dang, T. N. Y. (2023). The predictive validity of the IELTS test and contribution of IELTS preparation courses to international students' subsequent academic study: Insights from Vietnamese international students in the UK. *RELJ Journal*, 54(1), 84-98. <https://doi.org/10.1177/0033688220985533>
- Dao, P., Nguyen, M. X. N. C., & Nguyen, N. B. C. (2021). Effect of pronunciation instruction on L2 learners' listening comprehension. *Journal of Second Language Pronunciation*, 7(1), 10–37. <https://doi.org/10.1075/jslp.19012.dao>
- Dao, P., Nguyen, M.X.N.C, & Nguyen, H.V. (2024). 'IELTS juniors' in Vietnam: Perceptions of learners, parents and IELTS preparation course providers. IELTS Research Reports Online Series, No. 3/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at <https://ielts.org/researchers/our-research/research-reports>
- Darabad, M. A., Abbasian, G., Mowlaie, B., & Abusaeedi, A. R. (2023). On the construct and perceptual validity measures of L1-based vs. L2-based elicitation as a measure of L2 classroom performance assessment. *Language Testing in Asia*, 13, Article 13 (2023). <https://doi.org/10.1186/s40468-023-00218-4>

- Davidson, S. (2022). The domain expert perspective: A qualitative study into the views expressed in a standard-setting exercise on a language for specific purposes (LSP) test for health professionals. *Language Testing*, 39(1), 117–141.
<https://doi.org/10.1177/02655322211010737>
- Davis, L., & Norris, J. (2021). Developing an innovative elicited imitation task for efficient English proficiency assessment. *ETS Research Report Series*, 2021(1), 1–30.
<https://doi.org/10.1002/ets2.12338>
- Davis, L., & Papageorgiou, S. (2021). Complementary strengths? Evaluation of a hybrid human-machine scoring approach for a test of oral academic English. *Assessment in Education: Principles, Policy & Practice*, 28(4), 437–455.
<https://doi.org/10.1080/0969594X.2021.1979466>
- Davison, C. (2023). Assessment Literacy: Changing Cultures, Enculturing Change in Hong Kong. *Chinese Journal of Applied Linguistics*, 46(2), 180–197.
<https://doi.org/10.1515/CJAL-2023-0203>
- De John, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9(1), 541–560.
<https://doi.org/10.1146/annurev-linguistics-030521-052114>
- DE JONG, J. H. A. L. (2023). Reading: A precondition for Writing. In R. Horowitz (Ed.), *The Routledge international handbook of research on writing* (2nd. ed., pp. 556–562). Routledge.
- De Jong, J. H.A.L. (2022) The Action-oriented approach and language testing: A critical view. In D. Little & N. Figueras (Eds.), *Reflecting on the Common European Framework of reference for languages and its companion volume* (pp. 53–63). Multilingual Matters. <https://doi.org/10.21832/9781800410206-009>
- De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476.
<https://doi.org/10.1080/0969594X.2021.1951162>
- De Malsche, F., Tobback, E., & Vandenbroucke, M. (2024). Small talk as relational practice during assessment. *Journal of Applied Linguistics and Professional Practice*, (0), 1–24.
- De Silva, R. (2021). Assessment literacy and assessment practices of teachers of English in a South-Asian context: Issues and possible washback. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world* (pp. 433–446). Springer. https://doi.org/10.1007/978-981-33-4232-3_31
- Deane, P., Yan, D., Castellano, K., Attali, Y., Lamar, M., Zhang, M., Blood, I., Bruno, J. V., Li, C., Cui, W., Ruan, C., Appel, C., James, K., Long, R., & Qureshi, F. (2024). Modeling writing traits in a formative essay corpus. *ETS Research Report Series*, 2024(1), 1–62. <https://doi.org/10.1002/ets2.12377>
- deBoer, M., Leontjev, D., & Friederich, L. (2023). From language to function: developing self-and peer-assessment tools. *ELT Journal*, 77(1), 94–104.
<https://doi.org/10.1093/elt/ccac014>
- Dee, T. S., & Domingue, B. W. (2021). Assessing the impact of a test question: Evidence from the “Underground Railroad” controversy. *Educational Measurement: Issues and Practice*, 40(2), 81–88. <https://doi.org/10.1111/emip.12411>
- DeKeyser, R., & Li, S. (2021). Epilogue: Evidence for the validity of implicit aptitude and the need for construct validation. *Studies in Second Language Acquisition*, 43(3),

- 692–697. <https://doi.org/10.1017/S0272263121000140>
- del Mar Sánchez-Pérez, M. (2023). The impact of EMI on student English writing proficiency in a Spanish undergraduate engineering context. *Studies in Second Language Learning and Teaching*, 13(2), 373–397. <https://doi.org/10.14746/ssl.t.38279>
- Delgado-Osorio, X., Koval, V., Hartig, J., & Harsch, C. (2023). Strategic processing of source text in reading-into-writing tasks: A comparison between summary and argumentative tasks. *Journal of English for Academic Purposes*, 62, Article 101227. <https://doi.org/10.1016/j.jeap.2023.101227>
- Deng, Z., Uccelli, P., & Snow, C. (2022). Diversity of Advanced Sentence Structures (DASS) in writing predicts argumentative writing quality and receptive academic language skills of fifth-to-eighth grade students. *Assessing Writing*, 53, Article 100649. <https://doi.org/10.1016/j.asw.2022.100649>
- Derakhshan, A., & Fathi, J. (2024). Growth mindset, self-efficacy, and self-regulation: A symphony of success in L2 speaking. *System*, 123, 103320. <https://doi.org/10.1016/j.system.2024.103320>
- Derakhshan, A., & Ghasvand, F. (2022). Demystifying Iranian EFL teachers' perceptions and practices of learning-oriented assessment (LOA): challenges and prospects in focus. *Language Testing in Asia*, 12, Article 55. <https://doi.org/10.1186/s40468-022-00204-2>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE.
- Devine, A., Lopes, S., & Harrison, G. (2021). Young learners' positive attitudes to learning in South America. *Research Notes*, 80, 21–28. Cambridge Assessment English.
- DeVore, S., & Kyle, K. (2023). Assessing syntactic and lexicogrammatical use in second language Mandarin writing samples. *Journal of Second Language Writing*, 60, Article 101014. <https://doi.org/10.1016/j.jslw.2023.101014>
- Devos, N., Nizonkiza, D., & Lynch, S. (2024). Developing and validating a post-admission screening-diagnostic assessment procedure to offer language support in college diploma programs. *Canadian Journal of Applied Linguistics*, 27(1), 51–77. <https://doi.org/10.37213/cjal.2024.33259>
- Deygers, B. (2021). The CEFR companion volume: Between research-based policy and policy-based research. *Applied Linguistics*, 42(1), 186–191. <https://doi.org/10.1093/applin/amz024>
- Deygers, B., & Vanbuel, M. (2022). Gauging the impact of literacy and educational background on receptive vocabulary test scores. *Language Testing*, 39(2), 191–211. <https://doi.org/10.1177/02655322211049097>
- Deygers, B., Bigelow, M., Lo Bianco, J., Nadarajan, D., & Tani, M. (2021). Low print literacy and its representation in research and policy. *Language Assessment Quarterly*, 18(5), 463–476. <https://doi.org/10.1080/15434303.2021.1903471>
- Di Zhang, E., & Yu, S. (2023). Investigating the relationship between linguistic changes in L2 writers' paraphrasing, paraphrasing performance and L2 proficiency. *Applied Linguistics Review*, 14(5), 1451–1473. <https://doi.org/10.1515/applirev-2021-0014>
- Dimitrov, D. M., & Atanasov, D. V. (2021). An approach to test equating under the latent D-scoring method. *Measurement: Interdisciplinary Research and Perspectives*, 19(3), 153–162. <https://doi.org/10.1080/15366367.2020.1843107>
- Dimitrov, D. M., & Atanasov, D. V. (2022). Testing for differential item functioning under the

- D-Scoring method. *Educational and Psychological Measurement*, 82(1), 107–121.
<https://doi.org/10.1177/00131644211001524>
- Ding, W., & Yu, G. (2023). Using causal explanation speaking tasks to assess young EFL learners' speaking ability: The effects of age, cognitive, and L2 linguistic development. *Language Assessment Quarterly*, 20(3), 251-276.
<https://doi.org/10.1080/15434303.2023.2206811>
- Dissake, E. M. K. (2022). Assessing litigant's language proficiency: the case of the Bafoussam Court of First Instance. *Language Policy*, 21(2), 217-234.
<https://doi.org/10.1007/s10993-021-09603->
- Dobrić, N. (2024). Effects of errors on ratings of writing performances—Evidence from a high-stakes exam. *Assessing Writing*, 59, 100806.
<https://doi.org/10.1016/j.asw.2023.100806>
- Dobrić, N., Cesnik, H., Harsch, C. (2023). *Festschrift in honour of Günther Sigott: Advanced methods in language testing*. Peter Lang. <https://doi.org/10.3726/b20407>
- Dobrić, N., Sigott, G., Ilc, G., Lazović, V., Cesnik, H., & Stopar, A. (2021). Errors as indicators of writing task difficulty at the Slovene general matura in English. *International Journal of Applied Linguistics*, 31(3), 475–491.
<https://doi.org/10.1111/ijal.12345>
- Doeniyas, C., Gül, Z. T., & Alci, B. (2023). A non-Western adaptation of the Situated Academic Writing Self-Efficacy Scale (SAWSES). *Assessing Writing*, 57, Article 100763. <https://doi.org/10.1016/j.asw.2023.100763>
- Dong, J., Zhao, Y., & Buckingham, L. (2024). Thirty years of writing assessment: A bibliometric analysis of research trends and future directions. *Assessing Writing*, 61, 100862. <https://doi.org/10.1016/j.asw.2024.100862>
- Dong, M. (2022). Do school status and classroom English learning practices influence Chinese learners' perceptions of a high-stakes English test?. *Language Testing in Asia*, 12, Article 28.
<https://doi-org.proxy.bib.uottawa.ca/10.1186/s40468-022-00182-5>
- Douglas, D. (2023). Specificity, authenticity, and inseparability: Assessing integrated oral skills for pilots and air traffic controllers. *Applied Linguistics Papers*, 27(1), 4–13.
<https://doi.org/10.32612/uw.25449354>
- Drackert, A., Zimmermann, S., Marks, D. & Korflür, D. (2023). New test format – new research agenda: An overview of the technology-related research at g.a.s.t. In: *Conference Proceedings of the ALTE 8th International Conference: Language Assessment Fit for the Future* (pp. 34–38).
- Drennan, L. (2022). Transitioning from undergraduate to postgraduate studies : A multistage evaluation of a discipline-specific writing intervention. *Journal for Language Teaching*, 56(2), 1-24. <https://doi.org/10.56285/jltVol56iss2a5293>
- Drennan, L., Joubert, M., Weideman, A. & Posthumus, R. (2021). Combined measures of identifying language risk for first-year students. *Journal for Language Teaching*, 55(2), 93-116. <https://doi.org/10.4314/jlt.v55i2.4>
- Drown, L., Giovannone, N., Pisoni, D., & Theodore, R. (2023). Validation of two measures for assessing English vocabulary knowledge on web-based testing platforms: brief assessments. *Linguistics Vanguard*, 9(1), 99-111.
<https://doi.org/10.1515/lingvan-2022-0116>
- Du Plessis, C. (2021): Basic education and academic literacy: conflicting constructs in the South African National Senior Certificate (NSC) Language Examination. In A.

- Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 95-116). Multilingual Matters.
- Du Plessis, T. (2021). Institutional language policy and academic literacy in South African higher education. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 3-21). Multilingual Matters.
- Du, G., & Man, D. (2022). Person factors and strategic processing in L2 listening comprehension: Examining the role of vocabulary size, metacognitive knowledge, self-efficacy, and strategy use. *System*, 107, 102801. <https://doi.org/10.1016/j.system.2022.102801>
- Duan, S., & Shi, Z. (2024). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Language Teaching Research*, 28(2), 497-530. <https://doi.org/10.1177/13621688211002942>
- Ducasse, A.M. (2022) Oral reflection tasks: Advanced Spanish L2 learner insights on emergency remote teaching assessment practices in a higher education context. *Languages*, 7(1), Article 26. <https://doi.org/10.3390/languages7010026>
- Dudley, A., & Marsden, E. (2024). The lexical content of high-stakes national exams in French, German, and Spanish in England. *Foreign Language Annals*, 57(2), 311-338.
- Dudley, A., Marsden, E., & Bovolenta, G. (2024). A Context-Aligned Two Thousand Test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England. *Language Testing*, 41(4), 759–791. <https://doi.org/10.1177/02655322241261415>
- Duke, G. (2023). Training L2 learners to be autonomous through instructional scaffolding. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 83–104). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Dunlea, J., O'Sullivan, B., Dunn, K., Felice, M., Lightfoot, A., Patel, M., Spiby, R., & Westbrook, C. (2024). Assessing English as an international language. In: Galloway, N. & Fuad Selvi, A. (Eds.). *Routledge Handbook of English as an International Language* (pp.184-198), Routledge. DOI: 10.4324/9781003203261-16
- Dunn, K. J. (2024). Random-item Rasch models and explanatory extensions: A worked example using L2 vocabulary test item responses. *Research Methods in Applied Linguistics*, 3(3), Article 100143. <https://doi.org/10.1016/j.rmal.2024.100143>
- Duong, P. T., Perez, M. M., Nguyen, L. Q., Desmet, P., & Peters, E. (2023). The impact of input, input repetition, and task repetition on L2 lexical use and fluency in speaking. *Studies in second language learning and teaching*, 13(1), 101-124. <https://doi.org/10.14746/ssl.t.29727>
- Durrant, P., & Brenchley, M. (2021). The development of academic collocations in children's writing. In P. Szudarski & S. Barclay (Eds.), *Vocabulary theory, patterning and teaching* (pp. 99–120). Multilingual Matters.
- Durrant, P., & Durrant, A. (2022). Appropriateness as an aspect of lexical richness: What do quantitative measures tell us about children's writing? *Assessing Writing*, 51, Article 100596. <https://doi.org/10.1016/j.asw.2021.100596>

E

- East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–273. <https://doi.org/10.1017/s0261444823000034>
- Ebadi, S., Karimi, E., & Vakili, S. (2023). An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment. *Language Testing in Asia*, 13, Article 5 (2023). <https://doi.org/10.1186/s40468-023-00221-9>
- Eberharter, K., Kormos, J., Guggenbichler, E., Ebner, V. S., Suzuki, S., Moser-Frötscher, D., Konrad, E., & Kremmel, B. (2023). Investigating the impact of self-pacing on the L2 listening performance of young learner candidates with differing L1 literacy skills. *Language Testing*, 40(4), 960–983. <https://doi.org/10.1177/02655322221149642>
- Ebrahimi, B., & Ebadi, S. (2024). Exploring the impact of screencast-based ipsative assessment on EFL students' writing development. *Language Testing in Asia*, 14, Article 41. <https://doi.org/10.1186/s40468-024-00315-y>
- Eckes, T., & Jin, K. Y. (2021). Examining severity and centrality effects in TestDaF writing and speaking assessments: An extended Bayesian many-facet Rasch analysis. *International Journal of Testing*, 21(3–4), 131–153. <https://doi.org/10.1080/15305058.2021.1963260>
- Edmonds, A., Clenton, J., & Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System*, 108, 102855. <https://doi.org/10.1016/j.system.2022.102855>
- Eerdemutu, L., Dewaele, J.-M., & Wang, J. (2024). Developing a short language classroom engagement scale (Lces) and linking it with needs satisfaction and achievement. *System*, 120, 103189. <https://doi.org/10.1016/j.system.2023.103189>
- Effatpanah, F., Baghaei, P., & Karimi, M. N. (2024). A mixed Rasch model analysis of multiple profiles in L2 writing. *Assessing Writing*, 59, 100803. <https://doi.org/10.1016/j.asw.2023.100803>
- Eguchi, M., Suzuki, S., & Suzuki, Y. (2022). Lexical competence underlying second language word association tasks: Examining the construct validity of response type and response time measures. *Studies in Second Language Acquisition*, 44(1), 112–142. <https://doi.org/10.1017/S0272263121000164>
- Eibenstein, L. (2023). L3 acquisition of aspect: the influence of structural similarity, analytic L2 and general L3 proficiency. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1827–1858. <https://doi.org/10.1515/iral-2021-0220>
- Ekiert, M., & di Gennaro, K. (2021). Focused written corrective feedback and linguistic target mastery: Conceptual replication of Bitchener and Knoch (2010). *Language Teaching*, 54(1), 71–89. <http://doi.org/10.1017/S0261444819000120>
- El Ebyary, K., Shabara, R., & Boraie, D. (2024). The differential role of AI-operated WCF in L2 students' noticing of errors and its impact on writing scores. *Language Testing in Asia*, 14, Article 59. <https://doi.org/10.1186/s40468-024-00312-1>
- el Majidi, A., de Graaff, R., & Janssen, D. (2024). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*, 28(6), 2431–2452. <https://doi.org/10.1177/13621688211050619>
- Elder, C. (2021). Towards 'policy responsible' language assessment: Framing the Language Testing Research Centre's contribution in languages other than English. *Papers in Language Testing and Assessment*, 10(1), 4–29.

- http://www.altaan.org/uploads/5/9/0/8/5908292/2._elder.pdf
- Ellis, R., Roever, C., Shintani, N., & Zhu, Y. (2024). *Measuring second language pragmatic competence: A psycholinguistic perspective*. Multilingual Matters.
<https://doi.org/10.2307/jj.22679670>
- Elosua, P. (2022). Validity evidences for scoring procedures of a writing assessment task. A case study on consistency, reliability, unidimensionality and prediction accuracy. *Assessing Writing*, 54, Article 100669. <https://doi.org/10.1016/j.asw.2022.100669>
- Elosua, P. (2024). A Three-Step DIF Analysis of a Reading Comprehension Test Across Regional Dialects to Improve Test Score Validity. *Language Assessment Quarterly*, 21(2), 141–158. <https://doi.org/10.1080/15434303.2024.2307621>
- Elsayed, A. M., Kholikov, A., Abdullayeva, I., Al-Farouni, M., & Wodajo, M. R. (2024). Teacher support in AI-assisted exams: An experimental study to inspect the effects on demotivation, anxiety management in exams, L2 learning experience, and academic success. *Language Testing in Asia*, 14, Article 53.
<https://doi.org/10.1186/s40468-024-00328-7>
- Emmanouil, K., Zoe, G., Christos, A., & Markos, A. (2024). Linguistic factors affecting L1 language evaluation in argumentative essays of students aged 16 to 18 attending secondary education in Greece. *Assessing Writing*, 60, 100844.
<https://doi.org/10.1016/j.asw.2024.100844>
- Engelhard, G., & Wang, J. (2021). *Rasch models for solving measurement problems: Invariant measurement in the social sciences*. SAGE.
- Erguvan, I. D. (2021). The rise of contract cheating during the COVID-19 pandemic: A qualitative study through the eyes of academics in Kuwait. *Language Testing in Asia*, 11, Article 34. <https://doi.org/10.1186/s40468-021-00149-y>
- Erguvan, I. D. (2022). University students' understanding of contract cheating: A qualitative case study in Kuwait. *Language Testing in Asia*, 12, Article 56.
<https://doi.org/10.1186/s40468-022-00208-y>
- Erickson, G., & Tholin, J. (2022). Overall, a good test, but...—Swedish lower secondary teachers' perceptions and use of national test results of English. *Languages*, 7(1), Article 64. <https://doi.org/10.3390/languages7010064>
- Erickson, G., Borger, L., & Olsson, E. (2022). National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language Testing*, 39(3), 474–493. <https://doi.org/10.1177/02655322221075067>
- Erlam, R., & Rau, W. T. (2023). Fairer for all: Teachers identifying and solving problems with the assessment of English language learners in the mainstream classroom. *Studies in Language Assessment*, 12(2), 94–116.
- Erlam, R., & Wei, L. (2024). The importance of increased processing demands in the design of Elicited Imitation tests. *Language Teaching Research*, 28(4), 1472–1500.
<https://doi.org/10.1177/13621688211026032>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14, Article 51.
<https://doi.org/10.1186/s40468-024-00326-9>
- Eskin, D. (2023). Writing task performance and first language background on an ESL placement exam: A many-facets rasch analysis of facet main effects and differential facet functioning. *Studies in Applied Linguistics & TESOL*, 23(1), 37–65.
<https://doi.org/10.52214/salt.v23i1.11805>

- Estaji, M. (2024). Unveiling the impact of language assessment teacher education on EFL teachers' assessment agency and professional development trajectory: Perceptions, opportunities, and challenges. *Language Testing in Asia*, 14, Article 20.
<https://doi.org/10.1186/s40468-024-00292-2>
- Estaji, M., & Banitalebi, Z. (2023). A study of Test-Taking strategies of Iranian IELTS repeaters: Any change in the strategy use? *International Journal of Testing*, 23(3), 205-230. <https://doi.org/10.1080/15305058.2023.2195662>
- Estaji, M., & Hashemi, M. (2022). Phraseological competence in IELTS academic writing task 2: Phraseological units and test-takers' perceptions and use. *Language Testing in Asia*, 12, Article 34.
<https://doi-org.proxy.bib.uottawa.ca/10.1186/s40468-022-00180-7>
- Estaji, M., & Kogani, M. (2024). Assessment polarity or praxis parity? Uncovering the learning-oriented assessment (LOA) landscape of novice and experienced Iranian EFL teachers. *Language Testing in Asia*, 14, Article 34.
<https://doi.org/10.1186/s40468-024-00310-3>
- Estaji, M., & Safari, F. (2023) Learning-oriented assessment and its effects on the perceptions and argumentative writing performance of impulsive vs. reflective learners. *Language Testing in Asia*, 13, Article 31 (2023).
<https://doi.org/10.1186/s40468-023-00248-y>
- Estrada-Araoz, E. G., Sayed, B. T., Niyazova, G. G., & Lami, D. (2023). Comparing the effects of computerized formative assessment vs. computerized dynamic assessment on developing EFL learners' reading motivation, reading self-concept, autonomy, and self-regulation. *Language Testing in Asia*, 13, Article 39 (2023).
<https://doi.org/10.1186/s40468-023-00253-1>

F

- Falk, C. F., & Feuerstahler, L. M. (2022). On the performance of semi- and nonparametric item response functions in computer adaptive tests. *Educational and Psychological Measurement*, 82(1), 57–75. <https://doi.org/10.1177/00131644211014261>
- Fan, J. (2022). Ethics and fairness. In K. Luo (Ed.), *Major Issues in Language Assessment* (pp. 207-220). Foreign Language Testing and Research Press.
- Fan, J., & Zhang, X. (2024). Applying the Rasch model in language assessment. Foreign Language Teaching and Research Press.
- Fan, J., Frost, K., & Jin, Y. (2022). Local English testing in China's tertiary education: Contexts, policies, and practices. *Language Testing*, 39(3), 453–473.
<https://doi.org/10.1177/02655322211070839>
- Fan, J., Jin, T., & Chen, I. (2022). *Applying the Lexile Framework to evaluate text complexity in the Chinese context: A corpus-based study* (Report 1/2022).
https://www.britishcouncil.org/sites/default/files/rirg_fan_jin_chen_layout_1.pdf
- Fan, J., Knoch, U., & Chen, I. (2021). *Linking the GEPT writing subtest (Part 1) to the Common European Framework of Reference (CEFR)*. Language Testing Research Centre.
https://www.lttc.ntu.edu.tw/lttc-grants/doc/108report/GEPT_linking_study_Final_report.pdf
- Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices. *Language Testing in Asia*, 11, Article

2. <https://doi.org/10.1186/s40468-020-00117-y>
- Farhady, H., & Tavassoli, K. (2021). EFL teachers' perceptions and practices of their language assessment knowledge. *Language Testing in Asia*, 11, Article 17, 1-19. <https://doi.org/10.1186/s40468-021-00135-4>.
- Farhady, H., & Tavassoli, K. (2021). EFL teachers' perceptions and practices of their language assessment knowledge. *Language Testing in Asia*, 11, Article 17. <https://doi.org/10.1186/s40468-021-00135-4>
- Fazel, I., & Ali, A. M. (2022). EAP teachers' knowledge and use of learning-oriented assessment: A cross-contextual study. *System*, 104, Article 102685. <https://doi.org/10.1016/j.system.2021.102685>.
- Fechter, T. M., & Yoon, H. (2024). Evaluating methodological enhancements to the Yes/No Angoff standard-setting method in language proficiency assessment. *Language Testing*, 41(3), 530–556. <https://doi.org/10.1177/02655322231222600>
- Feng, H.-H., & Chukharev-Hudilainen, E. (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology*, 26(2), 58–77. <https://doi.org/10.1257/73479>
- Feng, K.-Z. (2024). Exploration of vocational high school students experiencing difficulty in cloze test performances: A mixed-methods study in Taiwan. *Language Testing in Asia*, 14, Article 16. <https://doi.org/10.1186/s40468-024-00274-4>
- Fernández Álvarez, M., & Gordenstein Montes, A. L. (Eds.). (2023). *Global perspectives on effective assessment in English language teaching*. IGI Global. <https://doi.org/10.4018/978-1-6684-8213-1>
- Finch, M., & Willis, J. (2021). Learning to be expert writers: Feedback in secondary English. *Assessment in Education: Principles, Policy & Practice*, 28(2), 101–117. <https://doi.org/10.1080/0969594X.2021.1908226>
- Finkelstein, S., & Netz, H. (2023). Challenging folk-linguistics: Grammatical and spelling variation in students' writing in Hebrew on WhatsApp and in essays. *Applied Linguistics*, 44(3), 555–575. <https://doi.org/10.1093/applin/amac072>
- Finkenstaedt-Quinn, S. A., Watts, F. M., & Shultz, G. V. (2024). Reading, receiving, revising: A case study on the relationship between peer review and revision in writing-to-learn. *Assessing Writing*, 59, 100808. <https://doi.org/10.1016/j.asw.2024.100808>
- Firoozi, T., Mohammadi, H., & Gierl, M. J. (2023). Using active learning methods to strategically select essays for automated scoring. *Educational Measurement, Issues and Practice*, 42(1), 34–43. <https://doi.org/10.1111/emip.12537>
- Fjørtoft, H., Lai, M. K., & Li, M. (2024). How curriculum and assessment policies affect the role of reading in an assessment culture: a New Zealand case study. *Assessment in Education: Principles, Policy & Practice*, 31(2), 1-31. <https://doi.org/10.1080/0969594X.2024.2365645>
- Fladung, I. T., Gruhn, S., Österbauer, V., & Jost, J. (2023). Assessing writing in fourth grade: Rhetorical specification effects on text quality. *Assessing Writing*, 57, Article 100764. <https://doi.org/10.1016/j.asw.2023.100764>
- Flores, J. A. A. (2021). The semiotics of writing: How adult L2 learners with emergent literacy make meaning in assessment texts through writing. *Journal of Second Language Writing*, 51, Article 100793. <https://doi.org/10.1016/j.jslw.2021.100793>
- Fong, C. J., Schallert, D. L., Williamson, Z. H., Lin, S., Williams, K. M., & Kim, Y. W. (2023). Are self-compassionate writers more feedback literate? Exploring undergraduates'

- perceptions of feedback constructiveness. *Assessing Writing*, 57, Article 100761. <https://doi.org/10.1016/j.asw.2023.100761>
- Fouché, I. (2023). Measuring the impact of academic literacy interventions: Refining an evaluation design through self-reflection and feedback. *Applied Linguistics*, 44(4), 718–746. <https://doi.org/10.1093/applin/amad025>
- Fox, J. (2021). *A mathematical primer for social statistics* (2nd ed.). SAGE.
- Fox, J., & Artemeva, N. (2022). *Reconsidering context in language assessment: Transdisciplinary perspectives, social theories, and validity*. Routledge. <https://doi.org/10.4324/9781351184571>
- Fox, J., Abdulhamid, N. & Turner, C.E. (2022). Classroom based assessment. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing*, 2nd ed. (pp. 119-135). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781003220756-6>
- François, J., & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. *Language Learning & Technology*, 25(2), 26–41. <http://hdl.handle.net/10125/73429>
- Franken, M. (2023). The new national literacy tests for post-primary students in Aotearoa New Zealand: How process and design issues undermine principles of a strong and fair qualification. *Studies in Language Assessment*, 12(2), 59-93.
- Frisén, L. B. (2024). Eyes on the ball: Teachers' rating processes when assessing a national L2 speaking test through the lens of their scoring document. *Studies in Language Assessment (SILA)*, 13(1), 1-36. <https://doi.org/10.58379/DAGH1483>
- Frost, K. (2021). Negotiating the boundaries of responsibility: Rethinking test takers and the ethics of testing. *Papers in Language Testing and Assessment*, 10(1), 70–83. http://www.altanz.org/uploads/5/9/0/8/5908292/5._frost.pdf
- Frost, K., Wigglesworth, G., & Clothier, J. (2021). Relationships between comprehension, strategic behaviours and content-related aspects of test performances in integrated speaking tasks. *Language Assessment Quarterly*, 18(2), 133–153. <https://doi.org/10.1080/15434303.2020.1835918>
- Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. *Studies in Second Language Acquisition*, 44(1), 2-34. <https://doi.org/10.1017/S0272263120000388>
- Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(1-2), 179-221. <https://doi.org/10.1080/09588221.2022.2033787>
- Fukuda, S. T., Lander, B. W., & Pope, C. J. (2022). Formative assessment for learning how to learn: Exploring university student learning experiences. *RELC Journal*, 53(1), 118-133. <https://doi.org/10.1177/0033688220925927>
- Fukunaga, T. (2023). L2 writing development through two types of writing task repetition. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1109-1138. <https://doi.org/10.1515/iral-2021-0144>
- Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56. <https://doi.org/10.32038/ltrq.2022.29.03>
- Furr, R. M. (2021). *Psychometrics: An introduction* (4th ed.). SAGE.
- Furtoso, V. B., Egido, A. A., & Moraes, I. (2022). Social justice in language assessment:

Intercultural communicative competence in the Celpe-Bras exam. In E. Meletiadiou (Ed.), *Handbook of research on fostering social justice through intercultural and multilingual communication* (pp. 162-186). IGI Global.
<https://doi.org/10.4018/978-1-6684-5083-3.ch009>

G

- Gaillat, T., Lafontaine, A., & Knefati, A. (2023). Visualizing linguistic complexity and proficiency in learner English writings. *CALICO Journal*, 40(2), 178-197.
<https://doi.org/10.1558/cj.19487>
- Galaczi, E. D., & Taylor, L. (2021). Measuring interactional competence. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 338–348). Routledge.
<https://doi.org/10.4324/9781351034784-36>
- Gallardo del Puerto, F., & Basterrechea, M. (2024). The role of proficiency and pair formation method in language-related episodes: A study of young CLIL learners' interaction. *Language Teaching Research*, 28(3), 1087-1105.
<https://doi.org/10.1177/13621688211024363>
- Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503-525.
<https://doi.org/10.1080/15434303.2022.2128802>
- Gan, L., & Lam, R. (2023). A tale of two teachers: Construction of novice university English teachers' assessment identity. *Chinese Journal of Applied Linguistics*, 46(2), 219-236. <https://doi.org/10.1515/CJAL-2023-0205>
- Gao, J., & Sun, P. P. (2024). How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics*, 34(1), 276-291.
<https://doi.org/10.1111/ijal.12493>
- Gao, S., & Pan, M. (2024). Implementing dynamic assessment in writing in Chinese as a second language: Towards an adjusted cumulative group dynamic assessment. *System*, 123, 103327. <https://doi.org/10.1016/j.system.2024.103327>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- García Laborda, J., & Fernández Álvarez, M. (2021). Multilevel Language tests: Walking into the land of the unexplored. *Language Learning & Technology*, 25(2), 1–25.
<http://hdl.handle.net/10125/73428>
- Garcia-Ponce, E. E., & Tavakoli, P. (2022). Effects of task type and language proficiency on dialogic performance and task engagement. *System*, 105, Article 102734.
<https://doi.org/10.1016/j.system.2022.102734>
- Gebril, A. (2021). Learning-oriented assessment: Main issues and an overview. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 1–10). Routledge. <https://doi.org/10.4324/9781003014102-1>
- Gebril, A. (2021). *Learning-oriented language assessment: Putting theory into practice*. Routledge. <https://doi.org/10.4324/9781003014102>
- Gebril, A. (2021). Nation-wide assessment reform in Egypt: Challenges and potentials of learning-oriented assessment in an examinations-based instructional setting. In A.

- Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 109–122). Routledge. <https://doi.org/10.4324/9781003014102-9>
- Gebril, A., & Bali, M. (2024). A Global South perspective on Open Science in language assessment: A response to Winke. *Language Testing*, 41(4), 886–891. <https://doi.org/10.1177/02655322241260121>
- George, A. C., & Robitzsch, A. (2021). Validating theoretical assumptions about reading with cognitive diagnosis models. *International Journal of Testing*, 21(2), 105–129. <https://doi.org/10.1080/15305058.2021.1931238>
- Geramipour, M. (2021). Rasch testlet model and bifactor analysis: How do they assess the dimensionality of large-scale Iranian EFL reading comprehension tests? *Language Testing in Asia*, 11, Article 3. <https://doi.org/10.1186/s40468-021-00118-5>
- Geranpayeh, a. (2023). Leveraging computational psychometrics for language testing. *Language Teaching Research Quarterly*, 37, 204–212, <https://doi.org/10.32038/ltrq.2023.37.11>
- Gesuato, S., Pagliarini, E., & Sanfelici, E. (2023). YOUNG ADULTS' INFERENTIAL READING SKILLS - RECOGNISING IMPLICIT AND UNSTATED CONTENT IN ITALIAN L1 AND ENGLISH L2. *Italiano LinguaDue*, 15(2), 20–20. <https://doi.org/10.54103/2037-3597/22011>
- Getman, E. P. (2024). Speaking test tasks for young language learners: Development and validation. In L. Davis & J. M. Norris (Eds.), *Challenges and Innovations in Speaking Assessment*. Routledge. <https://doi.org/10.4324/9781315201504-7>
- Getman, E. P. (2024). Four task characteristics and their influence on speaking performance and engagement. *English Teaching & Learning*, 48(2), 265–290. <https://doi.org/10.1007/s42321-024-00176-x>
- Ghaemi, H. (2022). Application of nonparametric item response theory in determining the one-dimensionality and adaptability of TOEFL iBT listening test. *Language Testing in Asia*, 12, Article 22. <https://doi.org/10.1186/s40468-022-00171-8>
- Ghaffarvand-Mokari, P. (2024). Effects of listeners' dialectal variation on comprehensibility and accentedness judgements of second language speech. *Journal of Second Language Pronunciation*, 10(1), 35–58. <https://doi.org/10.1075/jslp.23018.gha>
- Ghamoushi, M., Zenouzagh, Z. M., & Hashamdar, M. (2022). Development and validation of a potential assessment inventory for assessing EFL teachers' ecological agency. *Language Testing in Asia*, 12, Article 37. <https://doi.org/10.1186/s40468-022-00190-5>
- Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, 11, Article 27. <https://doi.org/10.1186/s40468-021-00143-4>
- Ghaneiarani, S., Alavi, S. M., & Kaivanpanah, S. (2024). Enhancing writing ability of Iranian EFL learners through learning-oriented assessment: Peer and teacher feedback implementation. *Language Testing in Asia*, 14, Article 39. <https://doi.org/10.1186/s40468-024-00298-w>
- Ghiasvand, F., Jahanbakhsh, A. A. & Sharifpour, P. (2023). Designing and validating an assessment agency questionnaire for EFL teachers: an ecological perspective. *Language Testing in Asia*, 13, Article 41. <https://doi.org/10.1186/s40468-023-00255-z>
- Gilbert, S., & May, L. (2023). Video-call mode speaking tests: an examiner verbal protocol study. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp.

- 71–86). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247.
<https://doi.org/10.1111/ijal.12317>
- Ginther, A. (2023). Administration, labor, and love. *Language Testing*, 40(1), 31–39.
<https://doi.org/10.1177/02655322221127365>
- Glasson, N. (2022). Is the devil you know better? Testwiseness and eliciting evidence of interactional competence in familiar versus unfamiliar triadic speaking tasks. *Studies in Language Assessment*, 11(2), 58–97.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0004/4396261/SiLA_11.2_Glasson.pdf
- Glasson, N., & Devine, A. (2023). The eye of the stakeholder—perceptions of remote speaking. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 53–70). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Glasson, N., & Halley, K. (2024). Less talk, more action? Exploring proficiency scores and embodied resources in online L2 interactions. *Classroom Discourse*, 1–22.
<https://doi.org/10.1080/19463014.2024.2408421>
- Goh, R., & Kiat Tan, K. (2023). Teachers' qualitatively different ways of experiencing assessment feedback: Implications for teacher assessment literacy. *Chinese Journal of Applied Linguistics*, 46(2), 253–269. <https://doi.org/10.1515/CJAL-2023-0207>
- Gokturk, N., & Chukharev-Hudilainen, E. (2023). Strategy use in a spoken dialog system–delivered paired discussion task: A stimulated recall study. *Language Testing*, 40(3), 630–657. <https://doi.org/10.1177/02655322231152620>
- Gokturk, N., & Chukharev, E. (2024). Exploring the Potential of a Spoken Dialog System-Delivered Paired Discussion Task for Assessing Interactional Competence. *Language Assessment Quarterly*, 21(1), 60–99.
<https://doi.org/10.1080/15434303.2023.2289173>
- Golparvar, S. E., & Abolhasani, H. (2022). Unpacking the contribution of linguistic features to graph writing quality: An analytic scoring approach. *Assessing Writing*, 53, Article 100644. <https://doi.org/10.1016/j.asw.2022.100644>
- Golparvar, S. E., & Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assessing Writing*, 47, Article 100504.
<https://doi.org/10.1016/j.asw.2020.100504>
- Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple-text source-based writing. *System*, 99, Article 102524. <https://doi.org/10.1016/j.system.2021.102524>
- Gomez Laich, M., & Taguchi, N. (2023). Effects of task complexity on the learning of genre specific rhetorical moves and linguistic forms: the case of contrast and argumentative essays. *International Review of Applied Linguistics in Language Teaching*, 61(2), 571–599. <https://doi.org/10.1515/iral-2021-0111>
- Gong, T., Zhang, M., & Li, C. (2022). Association of keyboarding fluency and writing performance in online-delivered assessment. *Assessing Writing*, 51, Article 100575. <https://doi.org/10.1016/j.asw.2021.100575>
- Goodwin, R., & Kirkpatrick, R. (2023). Using rubrics to improve writing skills: a study in Kuwait. *Language Testing in Asia*, 13, Article 17 (2023).

- <https://doi.org/10.1186/s40468-023-00224-6>
- Goriot, C., Hout, R., Broersma, M., Lobo, V., McQueen, J., & Unsworth, S. (2021). Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, 24(4), 546–568. <https://doi.org/10.1080/13670050.2018.1494131>
- Gottlieb, M. (2021). *Assessment in multiple languages: A handbook for school and district leaders*. SAGE.
- Gottlieb, M. (2021). *Classroom assessment in multiple languages: A handbook for teachers*. Corwin.
- Gràcia, M., Adam-Alcocer, A., & Castillo Mardones, P. (2023). Exploring the impact of a teacher development programme using a digital application on linguistic interactions in the classroom: a multiple case study. *Applied Linguistics Review*, 14(5), 1305–1343. <https://doi.org/10.1515/applirev-2020-0132>
- Graham, K. M. (2024). Examining the criterion validity of curriculum-based measurement of writing (CBM-W) with young EFL learners. *Research Methods in Applied Linguistics*, 3(2), Article 100108. <https://doi.org/10.1016/j.rmal.2024.100108>
- Graham, S., Harbaugh-Schattenkirk, A. G., Aitken, A., Harris, K. R., Ng, C., Ray, A., Wilson, J. M., & Wdowin, J. (2022). Writing motivation questionnaire: Validation and application as a formative assessment. *Assessment in Education: Principles, Policy & Practice*, 29(2), 238–261. <https://doi.org/10.1080/0969594X.2022.2080178>
- Grant, R., Miller, G. A., Design, Y. C., & Cook, G. (2023). A culturally sustaining and valid Alaska Native Language (Yugtun) assessment for school children. *Studies in Language Assessment*, 12(2), 205–234. <https://doi.org/10.58379/LRBE1549>
- Grapin, S. E. (2023). Assessment of English learners and their peers in the content areas: Expanding what “Counts” as evidence of content learning. *Language assessment quarterly*, 20(2), 215–234. <https://doi.org/10.1080/15434303.2022.2147072>
- Grapin, S. E., & Llosa, L. (2022). Dynamic assessment of English learners in the content areas: An exploratory study in fifth-grade science. *TESOL Quarterly*, 56(1), 201–229. <https://doi.org/10.1002/tesq.3059>
- Graves, B. C. (2023). Specifications grading to promote student engagement, motivation and learning: Possibilities and cautions. *Assessing Writing*, 57, Article 100754. <https://doi.org/10.1016/j.asw.2023.100754>
- Green, A. and Chan, S. (2022). Report to the Nursing and Midwifery Council on language test policy. *Nursing and Midwifery Council*, 1–27. <https://www.nmc.org.uk/globalassets/sitedocuments/english-language-consultation/crella-report-to-the-nmc-june-22.pdf>
- Green, B. A., & Lung, Y. S. M. (2021). English language placement testing at BYU-Hawaii in the time of COVID-19. *Language Assessment Quarterly*, 18(1), 6–11. <https://doi.org/10.1080/15434303.2020.1863966>
- Green, J., Weideman, A., Davis, C., Judith, K., & Harmes, M. (2024). Using a five-phase applied linguistics design to develop a contextualized academic literacy placement test for pre-university pathway students. *Literacy Research and Instruction*, 1–27. <https://doi.org/10.1080/19388071.2024.2340031>
- Greenblatt, P., & McDonald, P. (2024). Using Multimodal Public Service Announcement Video Projects as Communicative Language Assessments. *RELC Journal*, 55(2), 559–566. <https://doi.org/10.1177/00336882221102234>
- Greenier, V., Liu, X. & Xiao, Y. (2024). Creative translanguaging in formative assessment:

- Chinese teachers' perceptions and practices in the primary EFL classroom. *Applied Linguistics Review*, 15(5), 1861–1885. <https://doi.org/10.1515/applirev-2023-0085>
- Grocott, C., Moe, E., & Verhelst, N. (2024). Investigating level descriptors and their impact on validity. *Studies in Language Assessment (SILA)*, 13(1), 199–236. <https://doi.org/10.58379/LIHU1555>
- Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58–76. <https://doi.org/10.1080/0969594X.2020.1735995>
- Guapacha Chamorro, M. E. (2022). Cognitive validity evidence of computer- and paper-based writing tests and differences in the impact on EFL test-takers in classroom assessment. *Assessing Writing*, 51, Article 100594. <https://doi.org/10.1016/j.asw.2021.100594>
- Gui, M., Cao, X., & Zhao, J. (2023). EAP reading ability development during the process of generating research questions: A CDST perspective. *English for Academic Purposes*, 62, Article 101249. <https://doi.org/10.1016/j.jeap.2023.101249>
- Gujord, A.-K. H. (2023). Who succeeds and who fails? Exploring the role of background variables in explaining the outcomes of L2 language tests. *Language Testing*, 40(2), 227–248. <https://doi.org/10.1177/02655322221100115>
- Guo, J. & Cheng, L. (2024). Chinese Parent Involvement in Young Learners' English Tests: A Systematic Review of the Literature. 語言測試和評估 (Language Testing and Assessment), 4, 1-20.
- Guo, J., & East, M. (2024). The Role of Vocabulary in Speaking: Voices from Learners and Raters. *Language Assessment Quarterly*, 21(4–5), 452–472. <https://doi.org/10.1080/15434303.2024.2440887>
- Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing*, 61, 100864. <https://doi.org/10.1016/j.asw.2024.100864>
- Guo, Y., Xu, J., & Chen, C. (2023). Measurement of engagement in the foreign language classroom and its effect on language achievement: the case of Chinese college EFL students. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1225–1270. <https://doi.org/10.1515/iral-2021-0118>
- Guzman-Orth, D., Steinberg, J., & Albee, T. (2023). English learners who are blind or visually impaired: A participatory design approach to enhancing fairness and validity for language testing accommodations. *Language Testing*, 40(4), 933–959. <https://doi.org/10.1177/02655322231159143>
- Gyllstad, H., McLean, S., & Stewart, J. (2021). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing*, 38(4), 558–579. <https://doi.org/10.1177/0265532220979562>

H

- Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11, Article 16. <https://doi.org/10.1186/s40468-021-00132-7>

- Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11, Article 20. <https://doi.org/10.1186/s40468-021-00131-8>
- Ha, H. T., & Nguyen, D. T. B. (2023). Using structural equation modeling to examine the relationship between receptive vocabulary knowledge and receptive language proficiency: A new light on an old issue. *International Journal of Applied Linguistics*, 33(2), 382–398. <https://doi.org/10.1111/ijal.12475>
- Ha, H. T., Nguyen, D. T. B., & Stoeckel, T. (2024). What is the best predictor of word difficulty? A case of data mining using random forest. *Language Testing*, 41(4), 828–844. <https://doi.org/10.1177/02655322241263628>
- Haddadian, G., Radmanesh, S., & Haddadian, N. (2024). Construction and validation of a Computerized Formative Assessment Literacy (CFAL) questionnaire for language teachers: An exploratory sequential mixed-methods investigation. *Language Testing in Asia*, 14, Article 33. <https://doi.org/10.1186/s40468-024-00303-2>
- Hadidi, A. (2023). Comparing summative and dynamic assessments of L2 written argumentative discourse: Microgenetic validity evidence. *Assessing Writing*, 55, Article 100691. <https://doi.org/10.1016/j.asw.2022.100691>
- Haghi, I., & Tuersley, I. (2022). An investigation into how various textual features of academic journals can interact with students' engagement with these texts: A case of novice and non-traditional learners in STEM subjects (Report 3/2022). British Council and MetaMetrics. https://www.britishcouncil.org/sites/default/files/rirg_tuersley_and_haghi_layout_22.pdf
- Haghi, I., Clark, T. (2024). Assessment in the EAP Classroom. In E. Zehir Topkaya & H. Çelik (Eds.), *Teaching English for Academic Purposes*. Palgrave Macmillan/Cham. https://doi.org/10.1007/978-3-031-72545-6_13
- Hair Jr, J.F., Hult, G.T., Ringle, C.M., Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. (3rd ed.). SAGE.
- Hamada, A., & Takaki, S. (2021). Approximate replication of Matsuda and Gobel (2004) for psychometric validation of the Foreign Language Reading Anxiety Scale. *Language Teaching*, 54(4), 535–551. <http://doi.org/10.1017/S0261444819000296>
- Hamada, Y., & Suzuki, S. (2021). Listening to Global Englishes: Script-assisted shadowing. *International Journal of Applied Linguistics*, 31(1), 31–47. <https://doi.org/10.1111/ijal.12318>
- Hammad Al-Rashidi, A., Vadivel, B., Ramadan Khalil, N., & Basim, N. (2023). The comparative impacts of portfolio-based assessment, self-assessment, and scaffolded peer assessment on reading comprehension, vocabulary learning, and grammatical accuracy: insights from working memory capacity. *Language Testing in Asia*, 13, Article 24 (2023). <https://doi.org/10.1186/s40468-023-00237-1>
- Han, C. (2021). Analytic rubric scoring versus comparative judgment: A comparison of two approaches to assessing spoken-language interpreting. *Meta: Translators' Journal*, 66(2), 337–361. <https://doi.org/10.7202/1083182ar>
- Han, C. (2021). Assessing spoken-language interpretation: The method of comparative judgment. *Interpreting*, 24(1), 59–83. <https://doi.org/10.1075/intp.00068.han>
- Han, C. (2022). Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, 39(1), 30–55. <https://doi.org/10.1177/02655322211036100>
- Han, C. (2023). Interrogating the predictive validity of aptitude testing for interpreting: A

- systematic methodological review. *The Interpreter and Translator Trainer*, 17 (1), 7–28. <https://doi.org/10.1080/1750399X.2023.2170049>
- Han, C., & Lu, X-L. (2021). Interpreting quality assessment re-imagined: The synergy between human and machine scoring. *Interpreting and Society*, 1(1), 70-90. <https://doi.org/10.1177/27523810211033670>
- Han, C., & Lu, X-L. (2023). Can automated machine translation evaluation metrics be used to assess students' interpretation in the language learning classroom? *Computer Assisted Language Learning*, 36(5-6), 1064–1087. <https://doi.org/10.1080/09588221.2021.1968915>
- Han, C., & Xiao, X. (2022). A comparative judgment approach to assessing Chinese Sign Language interpreting. *Language Testing*, 39(2), 289–312. <https://doi.org/10.1177/02655322211038977>
- Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment and Evaluation in Higher Education*, 46(8), 1299–1313. <https://doi.org/10.1080/02602938.2020.1855624>
- Han, C., Hu, J., & Deng, Y. (2023). Effects of language background and directionality on raters' assessments of spoken-language interpretation. *Revista Española de Lingüística Aplicada*, 36(2), 556–584. <https://doi.org/10.1075/resla.21009.han>
- Han, C., Xiao, R., & Su, W. (2021). Assessing the fidelity of consecutive interpretation: The effects of using source versus target text as the reference material. *Interpreting*, 23(2), 245-268. <https://doi.org/10.1075/intp.00058.han>
- Han, C., Zheng, B-H., Xie, M-Q., & Chen, S-R. (2024). Raters' scoring process in assessment of interpreting: an empirical study based on eye tracking and retrospective verbalisation. *Interpreter and Translator Trainer*, 18(3), 400-422. <https://doi.org/10.1080/1750399X.2024.2326400>
- Han, D., & Qian, D. D. (2024). Evaluating the roles of breadth and depth of aural vocabulary knowledge in listening comprehension of EFL learners: An investigation applying auditory measures. *System*, 120, 103207. <https://doi.org/10.1016/j.system.2023.103207>
- Handley, Z. L., & Wang, H. (2023). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency? *Language Assessment Quarterly*, 21(1), 3–32. <https://doi.org/10.1080/15434303.2023.2283839>
- Handley, Z. L., & Wang, H. (2024). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency?. *Language Assessment Quarterly*, 21(1), 3-32.
- Hanjani, A. M. (2021). Insights into the cognitive processes of trained vs untrained EFL peer reviewers on writing: An exploratory study. *Assessing Writing*, 49, Article 100530. <https://doi.org/10.1016/j.asw.2021.100530>
- Hanjani, A. M. (2024). Comparing trained EFL peer reviewers' feedback: From claim to reality. *Assessing Writing*, 60, 100836. <https://doi.org/10.1016/j.asw.2024.100836>
- Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity arguments for automated essay scoring of young students' writing traits. *Language Assessment Quarterly*, 20(4-5), 399–420. <https://doi.org/10.1080/15434303.2023.2288253>
- Hannah, L., Kim, H., & Jang, E. E. (2022). Investigating the effects of task type and linguistic background on accuracy in automated speech recognition systems: Implications for use in language assessment of young learners. *Language Assessment Quarterly*, 19(3), 289–313.

- <https://doi.org/10.1080/15434303.2022.2038172>
- Hanzawa, K. (2024). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, 28(3), 816-838. <https://doi.org/10.1177/13621688211008693>
- Hao, Y., Wang, X., Bin, S., Yang, Q., & Liu, H. (2024). How syntactic complexity indices predict Chinese L2 writing quality: An analysis of unified dependency syntactically-annotated corpus. *Assessing Writing*, 61, 100847. <https://doi.org/10.1016/j.asw.2024.100847>
- Harding, L., Macqueen, S., & Pill, J. (2023). Assessing communicative competence. In M. Solon & M. Kanwit (Eds.), *Communicative competence in a second language: Theory, method, and applications* (pp. 187–207). Routledge. <http://doi.org/10.4324/9781003160779-14>
- Harrasi, K. T. S. AL. (2023). Reexamining feedback in the context of different rhetorical patterns of writing. *Language Testing in Asia*, 13, Article 4 (2023). <https://doi.org/10.1186/s40468-023-00216-6>
- Harsch, C., Koval, V., Delgado-Osorio, X. & Hartig, J. (2024). Usability of CEFR Companion Volume scales for the development of an analytic rating scale for academic integrated writing assessment. In *CEFR Journal - Research and Practice*. Vol. 6, 155-177, <https://doi.org/10.37546/JALTSIG.CEFR6-9>
- Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing*, 62, 100894. <https://doi.org/10.1016/j.asw.2024.100894>
- Harsch, C., Seyferth, S., & Villa Larenas, S. (2021). Evaluating a collaborative and responsive project to develop language assessment literacy. *Language Learning in Higher Education*, 11(2), 311–342. <https://doi.org/10.1515/cercles-2021-2020>
- Hartwell, K., & Aull, L. (2024). Navigating innovation and equity in writing assessment. *Assessing Writing*, 61, 100873. <https://doi.org/10.1016/j.asw.2024.100873>
- Harvey, R. (2022). *Implementing an integrated skills test in a Direct Entry project* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Hashimoto, B. J. (2021). Is frequency enough?: The frequency model in vocabulary size testing. *Language Assessment Quarterly*, 18(2), 171–187. <https://doi.org/10.1080/15434303.2020.1860058>
- Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), 100023.
- Haug, T., Boers-Visker, E., & Van den Bogaerde, B. (2021). Testing sign language learners. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 432–442). Routledge. <https://doi.org/10.4324/9781351034784-46>
- Haug, T., Holzknecht, F., & Mann, W. (2024). The impact of COVID-19 on sign language assessment. *Language Education & Assessment*, 7(1), 1316–1316. <https://doi.org/10.29140/lea.v7n1.1316>
- Haug, T., Mann, W., & Knoch, U. (2022). *The handbook of language assessment across modalities*. Oxford University Press.
- Hayashi, Y., Kondo, Y., & Ishii, Y. (2023). Automated speech scoring of dialogue response by Japanese learners of English as a foreign language. *Innovation in Language*

- Learning and Teaching*, 18(1), 32–46.
<https://doi.org/10.1080/17501229.2023.2217181>
- He, L., Jiang, Z., & Min, S. (2021). Diagnosing writing ability using China's Standards of English Language Ability: Application of cognitive diagnosis models. *Assessing Writing*, 50, Article 100565. <https://doi.org/10.1016/j.asw.2021.100565>
- He, L., Xiong, L., & Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: Performance stability and variability across modalities and performance levels. *System*, 106, Article 102787. <https://doi.org/10.1016/j.system.2022.102787>
- Heeren, J., Speelman, D., & Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458–1473. <https://doi.org/10.1080/13670050.2019.1709411>
- Heidari Vincheh, M., Mirzaei, A., & Roohani, A. (2024). A cognitive diagnostic approach to IELTS speaking test: Unveiling the subskills and test-takers' perceptions. *Language Testing in Asia*, 14, Article 42. <https://doi.org/10.1186/s40468-024-00311-2>
- Heidari, N., Ghanbari, N., & Abbasi, A. (2022). Raters' perceptions of rating scales criteria and its effect on the process and outcome of their rating. *Language Testing in Asia*, 12, Article 20. <https://doi.org/10.1186/s40468-022-00168-3>
- Henry, N., Mandel, H., & VanderHeijden, V. (2024). Dialogue journals and traditional writing assignments: A comparison of writing in L2 German classrooms. *Foreign Language Annals*, 57(2), 339–361. <https://doi.org/10.1111/flan.12752>
- Heritage, M. (2021). *Formative assessment: Making it happen in the classroom* (2nd ed.). SAGE.
- Heydarnejad, T., Ibrahim, K. A. A. A., Abdelrasheed, N. S. G., & Rezvani, E. (2022). The effect of academic emotion regulation on EFL learners' core of self-assessment and academic buoyancy: a structural equation modeling. *Language Testing in Asia*, 12, Article 57. <https://doi.org/10.1186/s40468-022-00207-z>
- Heydarnejad, T., Ismail, S. M., Shakibaei, G., & Saeedian, A. (2022). Modeling the impact of L2 grit on EFL learners' core of self-assessment and foreign language anxiety. *Language Testing in Asia*, 12, Article 50. <https://doi.org/10.1186/s40468-022-00200-6>
- Heydarnejad, T., Saberi Dehkordi, E., & Parmadi, B. (2022). Learning-oriented assessment in the classroom: the contribution of self-assessment and critical thinking to EFL learners' academic engagement and self-esteem. *Language Testing in Asia*, 12, Article 60. <https://doi.org/10.1186/s40468-022-00210-4>
- Heydarnejad, T., Tagavipour, F., Patra, I., & Farid Khafaga, A. (2022). The impacts of performance-based assessment on reading comprehension achievement, academic motivation, foreign language anxiety, and students' self-efficacy. *Language Testing in Asia*, 12, Article 51. <https://doi.org/10.1186/s40468-022-00202-4>
- Hicks, N. S. (2024). Exploring systematic orthographic crosslinguistic similarities to enhance foreign language vocabulary learning. *Language Teaching Research*, 28(6), 2385–2407. <https://doi.org/10.1177/13621688211047353>
- Hidri, S. (2021). Linking the International English Language Competency Assessment suite of examinations to the Common European Framework of Reference. *Language Testing in Asia*, 11, Article 9. <https://doi.org/10.1186/s40468-021-00123-8>

- Hilden, R., Pulkkinen, J., & Rautopuro, J. (2022). The meaningfulness of two curriculum-based national tests of English as a foreign language. *Studies in Language Assessment*, 11(2), 130–163.
- Hill, K. & Ducasse, A. M. (2022). Contextual variables in written assessment feedback in a university-level Spanish program. *Studies in Language Assessment*, 11(1), 16–32. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4282558/SiLA-11.1-Hill_and_Ducasse.pdf
- Hirai, A., Oka, H., Kato, T., & Maeda, H. (2022). Development and validation of an English test measuring EFL learners' critical thinking skills. *Language Testing in Asia*, 12, Article 45. <https://doi.org/10.1186/s40468-022-00193-2>
- Hirsch, R. R. (2021). Serendipitous: Lessons and insights into language assessment from Catherine Elder. *Language Assessment Quarterly*, 18(3), 316–328. <https://doi.org/10.1080/15434303.2020.1863967>
- Hirsu, L., Nitecka Barche, M. & Futro, D. (2024). Assessment and creativity through a translingual lens: introduction. *Applied Linguistics Review*, 15(5), 1831-1838. <https://doi.org/10.1515/applirev-2023-0115>
- Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101, Article 102569. <https://doi.org/10.1016/j.system.2021.102569>
- Ho, P. J. (2022). Assessing the range of cognitive processes in the Hong Kong Diploma of Secondary Education Examination (HKDSE)'s English language reading literacy test. *Language Testing in Asia*, 12, Article 18. <https://doi.org/10.1186/s40468-022-00167-4>
- Ho, S.K., & Gan, Z. (2023). Instructional practices and students' reading performance: a comparative study of 10 top performing regions in PISA 2018. *Language Testing in Asia*, 13, Article 48 (2023). <https://doi.org/10.1186/s40468-023-00261-1>
- Hodges, T. S., Wright, K. L., & McTigue, E. M. (2021). The Preservice Teacher Self-Efficacy for Writing Inventory. (PTSWI): A tool for measuring beliefs about writing. *Assessing Writing*, 49, Article 100545. <https://doi.org/10.1016/j.asw.2021.100545>
- Hoeve, K. B. (2022). A validity framework for accountability: educational measurement and language testing. *Language Testing in Asia*, 12, Article 3. <https://doi.org/10.1186/s40468-021-00153-2>
- Holmström, I., Schönström, K., & Ryttervik, M. (2023). Development of a Sign Repetition Task for Novice L2 Signers. *Language Assessment Quarterly*, 21(1), 33–59. <https://doi.org/10.1080/15434303.2023.2256320>
- Holzknicht, F., & Harding, L. (2024). Repeating the listening text: effects on listener performance, metacognitive strategy use, and anxiety. *TESOL Quarterly*, 58(1), 451–478. <https://doi.org/10.1002/tesq.3249>
- Holzknicht, F., Guggenbichler, E., Zehentner, M., Yoder, M., Konrad, E., & Kremmel, B. (2022). Comparing EMI university reading materials with students' reading proficiency utilizing Lexile® measures, Aptis test results, student questionnaire responses, and interviews with lecturers (Report 2/2022). British Council and MetaMetrics. https://www.britishcouncil.org/sites/default/files/rirg_holzknicht_et_al_layout.pdf
- Holzknicht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiby, R., & Dunlea, J. (2021). The effect of response order on candidate viewing behaviour and

- item difficulty in a multiple-choice listening test. *Language Testing*, 38(1), 41–61. <https://doi.org/10.1177/0265532220917316>
- Holzknicht, F., Tornay, S., Battisti, A., Batty, A. O., Tissi, K., Haug, T., & Ebling, S. (2024). Automated Sign Language Vocabulary Assessment: Comparing Human and Machine Ratings and Studying Learner Perceptions. *Language Assessment Quarterly*, 21(3), 245–265. <https://doi.org/10.1080/15434303.2024.2364877>
- Homayouni, M. (2022). Peer assessment in group-oriented classroom contexts: on the effectiveness of peer assessment coupled with scaffolding and group work on speaking skills and vocabulary learning. *Language Testing in Asia*, 12, Article 61. <https://doi.org/10.1186/s40468-022-00211-3>
- Hong, J. (2023). Content teachers' and lecturers' corrective feedback in EMI classes in high school and university settings. *Studies in Second Language Learning and Teaching*, 13(2), 451–469. <https://doi.org/10.14746/ssllt.38282>
- Honko, M., Neittaanmäki, R., Jarvis, S., & Huhta, A. (2023). Beyond literacy and competency – The effects of raters' perceived uncertainty on assessment of writing. *Assessing Writing*, 57, Article 100768. <https://doi.org/10.1016/j.asw.2023.100768>
- Hooft, H., Schiepers, M., & Vandommele, G. (2021). Developing and validating a multilingual literacy test for asylum seekers. *Language Assessment Quarterly*, 18(5), 530–546. <https://doi.org/10.1080/15434303.2021.1931230>
- Hoomanfar, M. (2024). L2 master's and doctoral students' preferences for supervisor written feedback on their theses/dissertations. *Assessing Writing*, 62, 100891. <https://doi.org/10.1016/j.asw.2024.100891>
- Hopp, H., & Thoma, D. (2021). Effects of plurilingual teaching on grammatical development in early foreign-language learning. *The Modern Language Journal*, 105(2), 464–483. <https://doi.org/10.1111/modl.12709>
- Horák, T., & Gandini, E. A. M. (2021). Going off-grid: Multiple benefits of rethinking the marking criteria format. *Research Notes*, 80, 7–20, Cambridge Assessment English. <https://www.cambridgeenglish.org/Images/630216-research-notes-80.pdf>
- Hosseinnia, M., & Kafi, Z. (2024). Constructing and validating a Code of Ethics in Testing Inventory: Investigating EFL instructors' perspectives. *Language Testing in Asia*, 14, Article 56. <https://doi.org/10.1186/s40468-024-00331-y>
- Hosseinpour, R. M., & Kazemi, Z. (2022). Composing strategies employed by high-and low-performing Iranian EFL students in essay writing classes. *Assessing Writing*, 51, Article 100601. <https://doi.org/10.1016/j.asw.2021.100601>
- Hougham, D., Clenton, J., & Uchihara, T. (2024). Disentangling the contributions of shorter vs. Longer lexical bundles to L2 oral fluency. *System*, 121, 103243. <https://doi.org/10.1016/j.system.2024.103243>
- Hsieh, C. (2024). The role of task types and reading proficiency on Young English as a foreign language learners' writing performances. *TESOL Quarterly*, 58(2), 978–990. <https://doi.org/10.1002/tesq.3286>
- Hsieh, C.-N. (2023). *Evaluating the use and interpretations of the TOEIC Listening and Reading test score report: Perspectives of test takers in Japan*. ETS Research Report Series (RR-23-02). <https://doi.org/10.1002/ets2.12364>
- Hsieh, C.-N. (2023). *Interpretation and use of a workplace English language proficiency test score report: Perspectives of test takers and score users in Taiwan*. ETS Research Report Series (RR-23-10). <https://doi.org/10.1002/ets2.12373>
- Hsieh, C.-N. (2024). Building a validity argument for the TOEFL Junior® tests. ETS

- Research Report Series, 2024(1), 1–22. <https://doi.org/10.1002/ets2.12379>
- Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486–1501. <https://doi.org/10.1080/13670050.2019.1691498>
- Hu, T & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486-1501, DOI: 10.1080/13670050.2019.1691498
- Hu, X., Zhang, X., & McGeown, S. (2024). Foreign language anxiety and achievement: A study of primary school students learning English in China. *Language Teaching Research*, 28(4), 1594-1615. <https://doi.org/10.1177/13621688211032332>
- Huang, B. H., Bailey, A. L., Sass, D. A., & Shawn Chang, Y. (2021). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401–428. <https://doi.org/10.1177/0265532220925731>
- Huang, D., & Watzinger-Tharp, J. (2024). The use of cohesive devices as proficiency level discriminators in Chinese DLI learners' writing. *Foreign Language Annals*, 57(2), 472-498. <https://doi.org/10.1111/flan.12722>
- Huang, H. T. D., Hung, S. T. A., Chao, H. Y., Chen, J. H., Lin, T. P., & Shih, C. L. (2022). Developing and validating a computerized adaptive testing system for measuring the English proficiency of Taiwanese EFL university students. *Language Assessment Quarterly*, 19(2), 162-188. <https://doi.org/10.1080/15434303.2021.1984490>
- Huang, J., & Chen, G. (2022). Individualized feedback to raters in language assessment: Impacts on rater effects. *Assessing Writing*, 52, Article 100623. <https://doi.org/10.1016/j.asw.2022.100623>
- Huang, J., Shu, T., Dong, Y., & Zhu, D. (2023). Constructing and validating a self-assessment scale for Chinese college English-major students' feedback knowledge repertoire in EFL academic writing: Item response theory and factor analysis approaches. *Assessing Writing*, 56, Article 100716. <https://doi.org/10.1016/j.asw.2023.100716>
- Huang, J., Zhu, D., Xie, D., & Shu, T. (2023). Examining the reliability of an international Chinese proficiency standardized writing assessment: Implications for assessment policy makers. *Assessing Writing*, 55, Article 100693. <https://doi.org/10.1016/j.asw.2023.100693>
- Huang, T., Steinkrauss, R., & Verspoor, M. (2021). Variability as predictor in L2 writing proficiency. *Journal of Second Language Writing*, 52, Article 100787. <https://doi.org/10.1016/j.jslw.2020.100787>
- Huang, Y., Wu, J., Li, M., & Yang, Y. (2022). The trajectory of syntactic complexity development in L1 Chinese narrative writings of primary school children: A systematic 5-year longitudinal study. *Assessing Writing*, 52, Article 100622. <https://doi.org/10.1016/j.asw.2022.100622>
- Hudson, C., Angelo, D., & Creagh, S. (2023). Instantiating justice, fairness and inclusiveness in English as an Additional Language/Dialect assessment frameworks: Unpacking the evidence base for the Bandscales State Schools (Queensland). *Studies in Language Assessment*, 12(2), 235–271.
- Huhta, A., Harsch, C., Leontjev, D., & Niemegen, L. (2024). *Diagnosing Writing*. London: Routledge. DOI: 10.4324/9781315511979-5

- Huhta, A., Harsch, C., Leontjev, D., & Nieminen, L. (2024). The diagnosis of writing in a second or foreign language. Routledge.
- Hui, B., Wong, S. S. Y., & Au, R. K. (2022). Reading aloud listening test items to young learners: Attention, item understanding, and test performance. *System*, 108, Article 102831. <https://doi.org/10.1016/j.system.2022.102831>
- Hunte, M. R., McCormick, S., Shah, M., Lau, C., & Jang, E. E. (2021). Investigating the potential of NLP-driven linguistic and acoustic features for predicting human scores of children's oral language proficiency. *Assessment in Education: Principles, Policy & Practice*, 28(4), 477–505. <https://doi.org/10.1080/0969594X.2021.1999209>
- Hwang, H., & Kim, H. (2023). Automatic analysis of constructional diversity as a predictor of EFL students' writing proficiency. *Applied Linguistics*, 44(1), 127–147. <https://doi.org/10.1093/applin/amac046>
- Hwang, H., & Kim, H. (2024). Korean Syntactic Complexity Analyzer (KOSCA): An NLP application for the analysis of syntactic complexity in second language production. *Language Testing*, 41(3), 506–529. <https://doi.org/10.1177/02655322231222596>
- Hyung-Jo, Y., & Mahmoud, A.T. (2023). Authorial voice in source-based and opinion-based argumentative writing: Patterns of voice across task types and proficiency levels. *English for Academic Purposes*, 62, Article 101228. <https://doi.org/10.1016/j.jeap.2023.101228>

I

- Ibrahim, K. A. A.-A., Kassem, M. A. M., & Lami, D. (2024). Intelligent Computer-Assisted Language Assessment (ICALA) in philosophy-based language instruction: Unraveling the effects on critical thinking, self-evaluation, academic resilience, and speaking development. *Language Testing in Asia*, 14, Article 46. <https://doi.org/10.1186/s40468-024-00320-1>
- Ignacio, A., Jang, E. E., Heywood, A., Hunte, M. R., & Lai, H. (2024). Fostering young students' writing skill development and self-regulation through Dynamic Assessment feedback during pandemic disruptions. *Studies in Language Assessment (SILA)*, 13(2), 147–186. <https://doi.org/10.58379/asqh2857>
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276–299. <https://doi.org/10.1177/02655322221112364>
- Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, 55, Article 100676. <https://doi.org/10.1016/j.asw.2022.100676>
- Imperial, J. M., Forey, G., & Tayyar Madabushi, H. (2024). Standardize: Aligning Language Models with Expert-Defined Standards for Content Generation. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1573–1594). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.94>
- In'nami, Y., & Koizumi, R. (2021). *Factor structure and four-skill profiles of the Aptis test*. (ARAGs research reports, AR-G/2021/2). British Council. <https://www.britishcouncil.org/factor-structure-and-four-skill-profiles-aptis-test>
- In'nami, Y., & Koizumi, R. (2022). The relationship between L2 listening and metacognitive

- awareness across listening tests and learner samples. *International Journal of Listening*, 36(2), 100-117. <http://dx.doi.org/10.1080/10904018.2021.1955683>
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2023). Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach. *Language Learning*, 73(3), 759–798. <https://doi.org/10.1111/lang.12548>
- In'nami, Y., Hijikata, Y., & Koizumi, R. (2022). Working memory capacity and L2 reading: A meta-analysis. *Studies in Second Language Acquisition*, 44(2), 381-406. <https://doi.org/10.1017/S0272263121000267>
- In'nami, Y., Koizumi, R., Jeon, E. H., & Arai, Y. (2022). L2 listening and its correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 235-283). John Benjamins. <https://doi.org/10.1075/bpa.13>
- In'nami, Y., Mizumoto, A., Plonsky, L., & Koizumi, R. (2022). Promoting computationally reproducible research in applied linguistics: Recommended practices and considerations. *Research Methods in Applied Linguistics*, 1(3), Article 100030. <https://doi.org/10.1016/j.rmal.2022.100030>
- Inbar-Lourie, O. (2022). EMI programs and formative assessment: Implications for the assessment literacy of content lecturers. *Journal of English Medium Instruction* 1(2), 204-231. <https://doi.org/10.1075/jemi.21014.inb>
- Inbar-Lourie, O., & Shohamy, E. (2021). The Conflict and Consequences of Two Assessment Measures in Israel: Global PISA vs. the National MEITZAV. In B. Lantaigne (Ed.), *Challenges in Language Testing around the World* (pp. 191-200). Springer International Publishing
- Inoue, C., & Lam, D. M. K. (2021). *The effects of extended planning time on candidates' performance, processes, and strategy use in the lecture listening-into-speaking tasks of the TOEFL iBT Test*. (TOEFL Research Report, TOEFL-RR-93, ETS Research Report No. RR-21-09). <https://doi.org/10.1002/ets2.12322>
- Inoue, C., & Nakatsuhara, F. (2022). Validation of a large-scale task-based test: Functional progression in dialogic speaking performance. In N. P. Sudharshana & L. Mukhopadhyay (eds.), *Task-based language teaching and assessment: Contemporary reflections from across the world* (pp. 217 - 247). Springer Nature.
- Inoue, C., Khabbazzashi, N., Lam, D. M. K., & Nakatsuhara, F. (2021). Towards new avenues for the IELTS Speaking Test: Insights from examiners' voices. *IELTS Research Reports Online Series*, 2. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/online-series-2021-2.ashx>
- Isaacs, T. (2025). Assessing pronunciation. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 216-229). Wiley.
- Isaacs, T. & Thomson, R. I. (2022). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *The evolution of pronunciation teaching and research: 25 years of intelligibility, comprehensibility, and accentedness* (pp. 125-151). John Benjamins.
- Isaacs, T., & Rose, H. (2022). Redressing the balance in the native speaker debate: Assessment standards, standard language, and exposing double standards. *TESOL Quarterly*, 56(1), 401-412. <https://doi.org/10.1002/tesq.3041>

- Isaacs, T., & Winke, P. M. (2024). Purposeful turns for more equitable and transparent publishing in language testing and assessment. *Language Testing*, 41(1), 3-8. doi:10.1177/02655322231203234
- Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, 40(3), 748–770. <https://doi.org/10.1177/02655322231158550>
- Isbell, D. R. (2021). Can the test support student learning? Validating the use of a second language pronunciation diagnostic. *Language Assessment Quarterly*, 18(4), 331–356. <https://doi.org/10.1080/15434303.2021.1874382>
- Isbell, D. R., & Kim, J. (2023). Developer involvement and COI disclosure in high-stakes English proficiency test validation research: A systematic review. *Research Methods in Applied Linguistics*, 2(3), Article 100060. <https://doi.org/10.1016/j.rmal.2023.100060>
- Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 305–315). Routledge. <https://doi.org/10.4324/9781351034784-33>
- Isbell, D. R., & Son, Y. A. (2022). Measurement properties of a standardized elicited imitation test: An integrative data analysis. *Studies in Second Language Acquisition*, 44(3), 859-885. <https://doi.org/10.1017/S0272263121000383>
- Isbell, D. R., & Son, Y.-A. (2023). Explanatory item response models for instrument validation: A tutorial based on an elicited imitation test. *Research Methods in Applied Linguistics*, 2(3), Article 100080. <https://doi.org/10.1016/j.rmal.2023.100080>
- Isbell, D. R., Crowther, D., & Nishizawa, H. (2024). Speaking performances, stakeholder perceptions, and test scores: Extrapolating from the Duolingo English test to the university. *Language Testing*, 41(2), 233–262. <https://doi.org/10.1177/02655322231165984>
- Isbell, D. R., Delaney, M., & Park, L. (2024). Developing measures of L2 Korean vocabulary knowledge. *Vocabulary Learning and Instruction*, 13(1), 1-10. <https://doi.org/10.29140/vli.v13n1.1227>
- Isbell, D. R., Kim, K. M., & Chen, X. (2023). Exploring the potential of automated speech recognition for scoring the Korean Elicited Imitation Test. *Research Methods in Applied Linguistics*, 2(3), Article 100076. <https://doi.org/10.1016/j.rmal.2023.100076>
- Isbell, D. R., Kremmel, B., & Kim, J. (2023). Remote Proctoring in Language Testing: Implications for Fairness and Justice. *Language Assessment Quarterly*, 20(4-5), 469–487. <https://doi.org/10.1080/15434303.2023.2288251>
- Isbell, D. R., Lee, J., & Jang, J. (2024). Linguistic influences on comprehensibility and accentedness in second language Korean speech. *International Journal of Applied Linguistics*, 34(4), 1575-1591. <https://doi.org/10.1111/ijal.12580>
- Isbell, D., & Sakai, M. (2022). Pronunciation assessment in classroom contexts. In T. Derwing, J. Levis, & S. Sonsaat (Eds.), *Pronunciation in Second Language Learning and Teaching: Innovations and Developments in Research and Teaching* (pp. 194-214). Routledge.
- Isbell, D. R., & Lee, J. (2022). Self-Assessment of comprehensibility and accentedness in second language Korean. *Language Learning*, 72(3), 806-852. <https://doi.org/10.1111/lang.12497>
- Islam, M. S., & Stapa, M. B. (2021). Students' low proficiency in spoken English in private

- universities in Bangladesh: Reasons and remedies. *Language Testing in Asia*, 11, Article 22. <https://doi.org/10.1186/s40468-021-00139-0>
- Islam, M. S., Hasan, M. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, 11, Article 1. <https://doi.org/10.1186/s40468-020-00116-z>
- Ismail, S. M., Nikpoo, I., & Prasad, K. D. V. (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. *Language Testing in Asia*, 13, Article 27 (2023). <https://doi.org/10.1186/s40468-023-00239-z>
- Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, 12, Article 40. <https://doi.org/10.1186/s40468-022-00191-4>
- Ismail, S.M., & Heydarnejad, T. (2023). Probing into the influence of EFL learners' self-assessment and evaluation apprehension in predicting their personal best goals and self-efficacy skills: A structural equation modeling. *Language Testing in Asia*, 13, Article 8 (2023). <https://doi.org/10.1186/s40468-023-00219-3>
- Issa, B. I., Koronkiewicz, B., & Faretta-Stutenberg, M. (2022). Second-language writing in university-level basic language programs: A survey of student and instructor beliefs. *Foreign Language Annals*, 55(2), 383–407. <https://doi.org/10.1111/flan.12609>
- Ito, S. (2021). The effects of retelling on reading comprehension: Focusing on different levels of comprehension and non-textual information in retelling protocols. *Japanese Language Testing Association Journal*, 24, 23-42. https://doi.org/10.20622/jltajournal.24.0_23
- Iwashita, N., Sasaki, M., Stell, S., & Yucel, M. (2021). Japanese stakeholders' perceptions of IELTS writing and speaking tests and their impact on communication and achievement. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/teaching-and-research/research-reports>

J

- Jahara, S. F., Hussain, M., Kumar, T., Goodarzi, A., & Assefa, Y. (2022). The core of self-assessment and academic stress among EFL learners: The mediating role of coping styles. *Language Testing in Asia*, 12, Article 21. <https://doi.org/10.1186/s40468-022-00170-9>
- Jalali, H., & Rahimi, M. (2022). Incongruence between learning style and written corrective feedback type: Mediating effect of implicit theory of learning style. *Canadian Journal of Applied Linguistics*, 25(2), 1–22. <https://doi.org/10.37213/cjal.2022.32478>
- Jalilzadeh, K., & Coombe, C. (2023). Constraints in employing learning-oriented assessment in EFL classrooms: teachers' perceptions. *Language Testing in Asia*, 13, Article 7 (2023). <https://doi.org/10.1186/s40468-023-00222-8>
- Jalilzadeh, K., Rashtchi, M., & Mirzapour, F. (2024). Cheating in online assessment: A qualitative study on reasons and coping strategies focusing on EFL teachers' perceptions. *Language Testing in Asia*, 14, Article 29. <https://doi.org/10.1186/s40468-024-00304-1>
- Jamalzadeh, M., Lotfi, A. R., & Rostami, M. (2021). Assessing the validity of an IAU

- General English Achievement Test through hybridizing differential item functioning and differential distractor functioning. *Language Testing in Asia*, 11, Article 8. <https://doi.org/10.1186/s40468-021-00124-7>
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, Article 102521. <https://doi.org/10.1016/j.system.2021.102521>
- Janssen, G. (2022). Local placement test retrofit and building language assessment literacy with teacher stakeholders: A case study from Colombia. *Language Testing*, 39(3), 377–400. <https://doi.org/10.1177/02655322221076153>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). Discussion, limitations, and future research. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 369-385). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). L2 speaking and its external correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 339-367). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Ji, J. (2021). College-level placement for heritage language learners. *Foreign Language Annals*, 54(3), 690–713. <https://doi.org/10.1111/flan.12523>
- Jia, L., Cai, J., & Wang, J. (2023). Promoting learning potential among students of L2 Chinese through dynamic assessment. *Language Assessment Quarterly*, 20(1), 66–87. <https://doi.org/10.1080/15434303.2022.2156870>
- Jia, W., & Zhang, P. (2023). Rater cognitive processes in integrated writing tasks: from the perspective of problem-solving. *Language Testing in Asia*, 13, Article 50 (2023). <https://doi.org/10.1186/s40468-023-00265-x>
- Jiang, J., Bi, P., Xie, N., & Liu, H. (2023). Phraseological complexity and low- and intermediate-level L2 learners' writing quality. *International Review of Applied Linguistics in Language Teaching*, 61(3), 765–790. <https://doi.org/10.1515/iral-2019-0147>
- Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing*, 57, Article 100869. <https://doi.org/10.1016/j.jslw.2022.100869>
- Jin, K.-Y., & Eckes, T. (2022). Detecting Differential Rater Functioning in Severity and Centrality: The dual DRF facets model. *Educational and Psychological Measurement*, 82(4), 757–781. <https://doi.org/10.1177/00131644211043207>
- Jin, Y. (2022). Consequential research of accountability testing: the case of the CET. *Language Testing in Asia*, 12, Article 15. <https://doi.org/10.1186/s40468-022-00165-6>
- Jin, Y. (2023). Decision making in large-scale language testing: Intersections of policy, practice and research. *Studies in Language Assessment*, 12(1), 64–92. <https://doi.org/10.58379/HKKX5020>
- Jin, Y., & Fan, J. (2023). Test-taker engagement in AI technology-mediated language assessment. *Language Assessment Quarterly*, 20(4-5), 488–500. <https://doi.org/10.1080/15434303.2023.2291731>
- Jin, Y., & Fan, J. (2024). Open Science for language assessment research and practice in China: A response to Winke. *Language Testing*, 41(4), 877–881.

- <https://doi.org/10.1177/02655322231223100>
- Johnson, C., Cardoso, W., Zuercher, B., Brannen, K., & Springer, S. (2024). Assessing pronunciation using dictation tools: The use of Google Voice Typing to score a pronunciation placement test. *Journal of Second Language Pronunciation*, 10(1), 10–34. <https://doi.org/10.1075/jslp.23033.joh>
- Jones, J., & Isaacs, T. (2022). Assessing second language pronunciation. In H. Moehebbi, & C. Coombe (Eds.), *Research questions in language education: A reference guide for teachers* (pp. 305–310). Springer.
- Jönsson, A., Balan, A., & Hartell, E. (2021). Analytic or holistic? A study about how to increase the agreement in teachers' grading. *Assessment in Education: Principles, Policy & Practice*, 28(3), 212–227. <https://doi.org/10.1080/0969594X.2021.1884041>
- Joo, S. H., Seong, Y., Suh, J., Jung, J.-Y., & Purpura, J. E. (2023). Assessing Korean writing ability through a scenario-based assessment approach. *Assessing Writing*, 58, Article 100766. <https://doi.org/10.1016/j.asw.2023.100766>
- Jung Youn, S. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing*, 40(1), 54–60. <https://doi.org/10.1177/02655322221126606>
- Jung, D., & Randazzo, K. (2022). *The upward cycle: learner progress through critical reflection and strategic response* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Jung, J., & Fu, X. (2023). The impact of pragmalinguistic support on video-conferenced collaborative suggestion-giving task. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1083–1107. <https://doi.org/10.1515/iral-2021-0138>

K

- Kahn-Horwitz, J., & Goldstein, Z. (2024). English foreign language reading and spelling diagnostic assessments informing teaching and learning of young learners. *Language Testing*, 41(1), 60–88. <https://doi.org/10.1177/02655322231162838>
- Kahng, J., & Otonya, M. (2021). What elicited imitation can show us: Quantitative and qualitative analyses of longitudinal changes. *TESOL Quarterly*, 55(1), 284–295. <https://doi.org/10.1002/tesq.3020>
- Kaltsa, M., Prentza, A., Prela, L., & Tsimpli, I. M. (2024). The interplay of lexical and grammatical development in Greek-Albanian bilingual children: evidence from the majority and the heritage language. *International Journal of Bilingual Education and Bilingualism*, 27(10), 1390–1406. <https://doi.org/10.1080/13670050.2024.2359408>
- Kamiya, N. (2024). Is the Common Test for University Admissions in Japan enough to measure students' general English proficiency? The case of the TOEIC Bridge. *Language Testing in Asia*, 14, Article 2. <https://doi.org/10.1186/s40468-024-00272-6>
- Kang, H.-A., Han, S., Kim, D., & Kao, S.-C. (2022). Polytomous testlet response models for technology-enhanced innovative items: implications on model fit and trait inference. *Educational and Psychological Measurement*, 82(4), 811–838. <https://doi.org/10.1177/00131644211032261>
- Kang, M., & Lam, R. (2024). Understanding university English instructors' assessment literacy: A formative assessment perspective. *Language Testing in Asia*, 14, Article 52. <https://doi.org/10.1186/s40468-024-00323-y>

- Kang, O., & Johnson, D. (2021). *Linguistic features and automatic scoring of Aptis speaking performance*. (AR-G/2021/3). British Council.
https://www.britishcouncil.org/sites/default/files/kang_and_johnson_layout.pdf
- Kang, O., Ahn, H., Yaw, K., & Chung, S-Y. (2021). Investigation of relationships between learner background, linguistic progression, and score gain on IELTS. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia.
https://www.ielts.org/-/media/research-reports/ielts-rr_2021-1_kang-et-al.ashx
- Kang, O., Egbert, J., & Miao, Y. (2023). *Speaking and writing features: Distinguishing IELTS proficiency levels and progression over time* (IELTS Research Report Series No. 1/23). British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262788072871836>
- Kang, O., Hirschi, K., Miao, Y., Ahn, H., & Won, Y. (2022). Test-takers' IELTS preparations and their attitudes towards IELTS practices and the use of technologies in the global pandemic. *IELTS Research Reports Online Series, No. 2/22*. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/-/media/research-reports/kang-et-al-2022.ashx>
- Kang, O., Kermad, A., & Taguchi, N. (2021). The interplay of proficiency and study abroad experience on the prosody of L2 speech acts. *Journal of Second Language Pronunciation, 7*(3), 343–369. <https://doi.org/10.1075/jslp.20024.kan>
- Kang, O., Yan, X., Kostromitina, M., Thomson, R., & Isaacs, T. (2024). Fairness of using different English accents: The effect of shared L1s in listening tasks of the Duolingo English test. *Language Testing, 41*(2), 263–289.
<https://doi.org/10.1177/02655322231179134>
- Kang, S., & Kim, Y. (2024). Examining the quality of mobile-assisted, video-making task outcomes: The role of proficiency, narrative ability, digital literacy, and motivation. *Language Teaching Research, 28*(6), 2326–2353.
<https://doi.org/10.1177/13621688211047984>
- Kao, Y-T. (2022). Effects of group dynamic assessment on L2 Chinese learners' literacy development: learners' responsiveness to interactive mediation. *Applied Linguistics Review, 13*(5), 843–871. <https://doi.org/10.1515/applirev-2019-0077>
- Kara, Y., Kamata, A., Qiao, X., Potgieter, C. J., & Nese, J. F. (2024). Equating oral reading fluency scores: A model-based approach. *Educational and Psychological Measurement, 84*(1), 190–209. <https://doi.org/10.1177/00131644221148122>
- Karaca, M., & Uysal, H. H. (2021). The development and validation of an inventory on English writing teacher beliefs. *Assessing Writing, 47*, Article 100507.
<https://doi.org/10.1016/j.asw.2020.100507>
- Karaca, M., & Uysal, H. H. An investigation into L2 writing teacher beliefs and their possible sources. *Assessing Writing, 56*, Article 100710.
<https://doi.org/10.1016/j.asw.2023.100710>
- Karakolidis, A., O'Leary, M., & Scully, D. (2021). Animated videos in assessment: comparing validity evidence from and test-takers' reactions to an animated and a text-based situational judgment test. *International Journal of Testing, 21*(2), 57–79.
<https://doi.org/10.1080/15305058.2021.1916505>
- Karatay, Y. (2023). Using spoken dialog systems to assess L2 learners' oral skills in a local language testing context. In Y. Xun, D. Slobodanka & J. G. April (Eds.), *Local language testing: Practice across contexts* (pp, 231–252). Springer.

- https://link.springer.com/chapter/10.1007/978-3-031-33541-9_12
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, 103332.
<https://doi.org/10.1016/j.system.2024.103332>
- Karatay, Y., & Karatay, L. (2024). The impact of automated writing evaluation on language learning and teaching: A research synthesis. *System*, 123.
<https://doi.org/10.1016/j.system.2024.103332>
- Karimi, M. N., & Mousavi-Moghadam, T. (2022). Integrated writing across conflicting L2 texts: the effects of prior beliefs on perspectivism and propositional content. *Assessing Writing*, 53, Article 100648. <https://doi.org/10.1016/j.asw.2022.100648>
- Karyolemou, M. (2022). Teaching endangered languages of oral tradition: how and what to assess?. *Applied Linguistics*, 43(2), 389-411.
<https://doi.org/10.1093/applin/amab045>
- Kasahara, K., & Kanayama, K. (2021). When to conduct a vocabulary quiz, before the review or after the review? *System*, 103, Article 102641.
<https://doi.org/10.1016/j.system.2021.102641>
- Katagiri, W. N. (2023). How sociocultural factors mediate washback effect: a study of the Hong Kong Diploma of Secondary Education Examination English Language Paper 4. *Language Testing in Asia*, 13, Article 38 (2023).
<https://doi.org/10.1186/s40468-023-00250-4>
- Kaya, E., O'Grady, S., & Kalender, İ. (2022). IRT-based classification analysis of an English language reading proficiency subtest. *Language Testing*, 39(4), 541-566.
<https://doi.org/10.1177/02655322211068847>
- Kazemian, M. & Khonamri, F. (2022). Past, present, and future of language assessment: An interview with Dr. Hossein Farhady, *MEXTESOL Journal*, 46(4), 1-8.
https://www.mextesol.net/journal/index.php?page=journal&id_article=46395
- Kazemian, M. & Khonamri, F. (2024). The online assessment in the Middle East. A conversation with Dr. Christine Coombe, *MEXTESOL Journal*, 48(1), 1-6.
- Kazemian, M., Khodareza, M.R., Khonamri, F., & Rahimy, R. (2022) Problematizing language testing and assessment syllabi through intercultural competence assessment perspective in an EFL context. *The Journal of Asia TEFL*, 19(1), 309-316. <http://dx.doi.org/10.18823/asiatefl.2022.19.1.23.309>
- Kazemian, M., Khodareza, M.R., Khonamri, F., & Rahimy, R. (2023). ELT scholars' attitudes towards inclusion of intercultural competence assessment in language proficiency tests. *TESL-EJ Journal*, 26(4), 1-21.
<https://tesl-ej.org/wordpress/issues/volume26/ej104/ej104a6/>
- Kennedy, E., & Shiel, G. (2022). Writing assessment for communities of writers: rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy & Practice*, 29(2), 127-149.
<https://doi.org/10.1080/0969594X.2022.2047608>
- Kessler, M. (2024). Assessment. In M. Kessler (Ed.), *Digital multimodal composing: Connecting theory, research and practice in second language acquisition* (pp. 109-120). Multilingual Matters. <https://doi.org/10.2307/jj.22679723.11>
- Kessler, M., Ma, W., & Solheim, I. (2022). The effects of topic familiarity on text quality, complexity, accuracy, and fluency: a conceptual replication. *TESOL Quarterly*, 56(4), 1163-1190. <https://doi.org/10.1002/tesq.3096>
- Khabbazzbashi, N. (2021). *On topic validity of speaking test*. Cambridge University Press.

- Khabbazzbashi, N., Chan, S. and Clark, T. (2023). Towards the new construct of academic English in the digital age, *ELT Journal*, 77(2), 207–216
<https://doi.org/10.1093/elt/ccac010>
- Khabbazzbashi, N., Nakatsuhara, F., Inoue, C., Kaplan, G., & Green, A. (2022). The design and validation of an online speaking test for young learners in Uruguay: challenges and innovations. *International Journal of TESOL Studies*, 4(1), 141-168.
<https://doi.org/10.46451/ijts.2022.01.10>
- Khabbazzbashi, N., Xu, J., & Galaczi, E. (2021). Opening the black box: Exploring automated speaking assessment. In B. Lantaigne, C. Coombe & J.D. Brown (Eds.), *Issues in language testing around the world* (pp.333–344). Springer.
https://doi.org/10.1007/978-981-33-4232-3_25
- Khan, K. (2022). The securitisation of language borders and the (re) production of inequalities. *TESOL Quarterly*, 56(4), 1458-1470. <https://doi.org/10.1002/tesq.3186>
- Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, 14, Article 48. <https://doi.org/10.1186/s40468-024-00324-x>
- Khau, A. H., & Huynh, V. T. M. (2022). An investigation into oral fluency perceived by teachers and students—in a Vietnamese context of English education. *Language Testing in Asia*, 12(1), Article 27. <https://doi.org/10.1186/s40468-022-00174-5>
- Kho, S. Q. E., Aryadoust, V., & Foo, S. (2023). An eye-tracking investigation of the keyword-matching strategy in listening assessment. *Education and Information Technologies*, 28, 3739–3763.
<https://link.springer.com/article/10.1007/s10639-022-11322-y>
- Khodi, A. (2021). The affectability of writing assessment scores: A G-theory analysis of rater, task, and scoring method contribution. *Language Testing in Asia*, 11, Article 30. <https://doi.org/10.1186/s40468-021-00134-5>
- Khodi, A., Khezerlou, H., & Sahraei, H. (2024). Dependability and utility of using e-portfolios in assessing EFL learners' speaking proficiency. *Computer Assisted Language Learning*, 37(7), 1579-1601. <https://doi.org/10.1080/09588221.2022.2093379>
- Khodi, A., Ponniah, L. S., Farrokhi, A. H., & Sadeghi, F. (2024). Test review of Iranian English language proficiency test: MSRT test. *Language Testing in Asia*, 14, Article 4. <https://doi.org/10.1186/s40468-023-00270-0>
- Khosravani, M., Rostamian, M., Ashraf, H., & Motallebzade, K. (2022). A structural equation modeling of English tests' social and educational consequences: exploring target, leverage, risk, and critical variables. *Language Testing in Asia*, 12, Article 46.
<https://doi.org/10.1186/s40468-022-00177-2>
- Khushik, G. A. (2024). Is the variation in syntactic complexity features observed in argumentative essays produced by B1 level EFL learners in Finland and Pakistan attributable exclusively to their L1? *Assessing Writing*, 60, 100839.
<https://doi.org/10.1016/j.asw.2024.100839>
- Kim, A. A., Tywoniw, R. L., & Chapman, M. (2022). Technology-enhanced items in Grades 1–12 English language proficiency assessments. *Language Assessment Quarterly*, 19(4), 343–367 . <https://doi.org/10.1080/15434303.2022.2039659>
- Kim, H. S., Lee, J. H., & Lee, H. (2024). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, 28(1), 7-28.
<https://doi.org/10.1177/1362168820981394>

- Kim, H., & Hwang, H. (2022). Assessing verb-construction integration in young learners of English as a foreign language: analyses of written and spoken production. *Language Learning*, 72(2), 497–533. <https://doi.org/10.1111/lang.12480>
- Kim, H., & Ro, E. (2023). Assessment of sentence sophistication in L2 spoken production: Expansion of verbs and argument structure constructions. *System*, 119, Article 103175. <https://doi.org/10.1016/j.system.2023.103175>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73–95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Kim, H., Song, C., Kim, J., Jeong, H., & Park, J. (2024). Validation of an elicited Imitation test as a measure of Korean language proficiency. *Language Testing in Asia*, 14, Article 17. <https://doi.org/10.1186/s40468-024-00287-z>
- Kim, J. (2023). Memorizing and fabricating? Uncovering high-Stakes writing test preparation. *Language Education & Assessment*, 6(1), 19–37. <https://doi.org/10.29140/lea.v6n1.1175>
- Kim, K. M., Chen, X., & Liu, X. (2024). Accuracy scoring of elicited imitation: A tutorial of automating speech data with commercial NLP support. *Research Methods in Applied Linguistics*, 3(3), Article 100127. <https://doi.org/10.1016/j.rmal.2024.100127>
- Kim, M. K., Kim, J., & Heidari, A. (2024). Exploring the multi-dimensional human mind: Model-based and text-based approaches. *Assessing Writing*, 61, 100878. <https://doi.org/10.1016/j.asw.2024.100878>
- Kim, M., Crossley, S. A., & Kim, B.-K. (2022). Second language reading and writing in relation to first language, vocabulary knowledge, and learning backgrounds. *International Journal of Bilingual Education and Bilingualism*, 25(6), 1992–2005. <https://doi.org/10.1080/13670050.2020.1838434>
- Kim, M., Nam, Y., & Crossley, S. A. (2022). Roles of working memory, syllogistic inferencing ability, and linguistic knowledge on second language listening comprehension for passages of different lengths. *Language Testing*, 39(4), 593–617. <https://doi.org/10.1177/02655322211060076>
- Kim, M., Qiu, X., & Wang Y. (2024). Interrater agreement in genre analysis: A methodological review and a comparison of three measures, *Research Methods in Applied Linguistics*, 3(1), Article 100097. <https://doi.org/10.1016/j.rmal.2024.100097>
- Kim, P., Beltrán, J., & Joo, S. H. (2022). An interview with APPLE Lecture speaker Dr. Paul D. Deane. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University*, 22(1), 47–55. <https://doi.org/10.52214/salt.v22i1.9901>
- Kim, S. (2021). Generalizability of CEFR criterial grammatical features in a Korean EFL corpus across A1, A2, B1, and B2 levels. *Language Assessment Quarterly*, 18(3), 273–295. <https://doi.org/10.1080/15434303.2020.1855647>
- Kim, S. (2021). Prepping for the TOEFL iBT Writing test, Gangnam style. *Assessing Writing*, 49, Article 100544. <https://doi.org/10.1016/j.asw.2021.100544>
- Kim, S., & Kessler, M. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, Article

100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Kim, Y. J., & Hammill, M. J. (2024). Influence of prior educational contexts on directed self-placement of L2 writers. *Assessing Writing*, 61, 100870. <https://doi.org/10.1016/j.asw.2024.100870>
- Kim, Y., & Emeliyanova, L. (2021). The effects of written corrective feedback on the accuracy of L2 writing: Comparing collaborative and individual revision behavior. *Language Teaching Research*, 25(2), 234–255. <https://doi.org/10.1177/1362168819831406>
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247. <https://doi.org/10.1016/j.jeap.2023.101247>
- Kirkpatrick, R., Vafadar, H., & Mohebbi, H. (2024). A scoping review of willingness to communicate in language education: premises to doubt, lessons to learn, and future research questions to examine. *Language Testing in Asia*, 14, 9. <https://doi.org/10.1186/s40468-024-00284-2>
- Kirkwood, M. (2023). Using ipsative feedback for personal bests in grammatical accuracy in L2 writing. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 105–128). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Kılıçkaya, F. (2021). Reconsidering language assessment courses in teacher training programs during and after the COVID-19 pandemic. *Lublin Studies in Modern Languages and Literature*, 45(3), 45–55. <https://doi.org/10.17951/lsmll.2021.45.3.45-55>
- Kılıçkaya, F. (2022). Pre-service EFL teacher candidates' concerns over digital language assessment and test anxiety. In W. M. El-Henawy & M. del Mar Suárez (Eds.), *English as a foreign language in a new-found post-pandemic world* (pp. 164–181). IGI Global. <https://doi.org/10.4018/978-1-6684-4205-0>
- Klieger, D., Bochenek, J., Ezzo, C., Holtzman, S., Cline, F., & Olivera-Aguilar, M. (2022). Using third-party evaluations to assess socioemotional skills in graduate and professional school admissions. *International Journal of Testing*, 22(1), 72–99. <https://doi.org/10.1080/15305058.2021.2019748>
- Knoch, U., & Fan, J. (2024). Assessing clinician language skills. In Ortega, P., Martínez, G., Lor, M., & Ramírez, A. S. *The Handbook of Language in Public Health and Healthcare*. Wiley. <https://doi.org/10.1002/9781119853855.ch12>
- Knoch, U., & Fan, J. (2024). Test score comparison tables: How well are they serving test users? *Language Testing*, 41(3), 681–693. <https://doi.org/10.1177/02655322241239348>
- Knoch, U., Deygers, B., & Khamboonruang, A. (2021). Revisiting rating scale development for rater-mediated language performance assessments: Modelling construct and contextual choices made by scale developers. *Language Testing*, 38(4), 602–626. <https://doi.org/10.1177/0265532221994052>
- Knoch, U., Fairbairn, J., & Jin, Y. (2021). *Scoring second language spoken and written performance: Issues, options and directions*. University of Toronto Press.
- Knoch, U., Fan, J., Elder, C., Zhao, K., & Pitman, A. (2021). *Investigating the discourse produced at score levels B2.2 to C2 on the Aptis advanced writing test*

- (AR-G/2021/5). British Council.
https://www.britishcouncil.org/sites/default/files/knoch_et_al_report_layout.pdf
- Knoch, Ute. (2021). The challenges of providing expert advice in policy contexts. *Papers in Language Testing and Assessment*, 10(1), 30–48.
http://www.altanz.org/uploads/5/9/0/8/5908292/3._knoch.pdf
- Koizumi, R. (2021). The Telephone Standard Speaking Test: An outside evaluator's investigation of a rebuttal to the generalization inference. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of argument-based validation research* (pp. 154–175). Cambridge University Press.
<https://doi.org/10.1017/9781108669849.010>
- Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142–161.
<https://doi.org/10.1080/15434303.2021.2023542>
- Koizumi, R., & In'nami, Y. (2024). Predicting functional adequacy from complexity, accuracy, and fluency of second-language picture-prompted speaking. *System*, 120, 103208.
<https://doi.org/10.1016/j.system.2023.103208>
- Koizumi, R., & Watanabe, A. (2021). Rater reliability in classroom speaking assessment in a Japanese senior high school. *Annual Review of English Language Education in Japan*, 32, 129–144. https://doi.org/10.20581/arele.32.0_129
- Koizumi, R., Agawa, T., Asano, K., & In'nami, Y. (2022). Skill profiles of Japanese English learners and reasons for uneven patterns. *Language Testing in Asia*, 12, Article 53.
<https://doi.org/10.1186/s40468-022-00203-3>
- Koizumi, R., Fukazawa, M., & Inoue, C. (2024). Usability of a Speaking Assessment Portal for Japanese teachers of English. *SHIKEN: A Journal of Language Testing and Evaluation in Japan*, 28(1), 19–37.
<https://doi.org/10.37546/JALTSIG.TEVAL28.1-2>
- Koizumi, R., In'nami, Y., & Jeon, E. H. (2022). L2 speaking and its internal correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 307–338). John Benjamins.
<https://doi.org/10.1075/bpa.13>
- Koizumi, R., Maie, R., Yanagisawa, A., & In'nami, Y. (2024). Considerations to promote and accelerate Open Science: A response to Winke. *Language Testing*, 41(4), 892–897.
<https://doi.org/10.1177/02655322241239379>
- Kollias, C. (2023). *Virtual standard setting: Setting cut scores*. Peter Lang.
<https://doi.org/10.3726/b20407>
- Kongsuwanakul, K. (2022). Relating a concordance-based cloze test to the model of communicative language ability: a verbal protocol study. *Language Testing in Asia*, 12, Article 1. <https://doi.org/10.1186/s40468-021-00151-4>
- Kormos, J., & Smith, A. M. (2024). Assessment of the language skills of language learners with SpLDs. In J. Kormos, & A. M. Smith (Eds.), *Teaching languages to students with specific learning differences* (2nd ed., pp. 143–164). Multilingual Matters.
<https://doi.org/10.2307/jj.22679668.10>
- Kormos, J., Eberharder, K., Guggenbichler, E., Baumgartinger, S., Ebner, V., & Kremmel, B. (2024). First Language Literacy Skills, Listening Strategy Use, and Anxiety in Self-Paced versus Single-Play L2 Listening Tests. *Language Assessment Quarterly*, 21(4–5), 303–323. <https://doi.org/10.1080/15434303.2024.2435811>

- Koster, T., & Morgan, G. (2023). Investigating students' perceptions of effective engagement methods for online and multi-modal ELICOS classes. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 43–60). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: a meta-analysis. *Studies in Second Language Acquisition*, 44(3), 886–911. <https://doi.org/10.1017/S0272263121000395>
- Kotowicz, J., Woll, B., & Herman, R. (2021). Adaptation of the British Sign Language Receptive Skills Test into Polish Sign Language. *Language Testing*, 38(1), 132–153. <https://doi.org/10.1177/0265532220924598>
- Kroc, E., & Olvera Astivia, O. L. (2022). The importance of thinking multivariately when setting subscale cutoff scores. *Educational and Psychological Measurement*, 82(3), 517–538. <https://doi.org/10.1177/00131644211023569>
- Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self vs. peer assessment activities in EFL-speaking classes: impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia*, 13, Article 36(2023). <https://doi.org/10.1186/s40468-023-00251-3>
- Kunnan, A. J. (Ed; 2023). *The Concise Companion to Language Assessment*. Wiley.
- Kunnan, A. J., Qin, C. Y., & Zhao, C. G. (2022). Developing a scenario-based English language assessment in an Asian university. *Language Assessment Quarterly*, 19(4), 368–393. <https://doi.org/10.1080/15434303.2022.2073886>
- Kushki, A., Nassaji, H., & Rahimi, M. (2022). Interventionist and interactionist dynamic assessment of argumentative writing in an EFL program. *System*, 107, Article 102800. <https://doi.org/10.1016/j.system.2022.102800>
- Kushki, A., Rahimi, M., & Davin, K. J. (2022). Dynamic assessment of argumentative writing: mediating task response. *Assessing Writing*, 52, Article 100606. <https://doi.org/10.1016/j.asw.2022.100606>
- Kvasova, O., Shevchenko, T., Hnapovska, L., Kalinichenko, V., & Budas, I. (2022). Implications of training university teachers in developing local writing rating scales. *Studies in Language Assessment*, 11(1), 37–52.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4282559/SiLA-11.1-Kvasova-et-al.pdf
- Kwon, O., & Lee, W. (2021). A life for English language education: An interview with Oryang Kwon. *Language Assessment Quarterly*, 18(2), 188–198.
<https://doi.org/10.1080/15434303.2020.1859512>
- Kwon, S. K., & Yu, G. (2023). Investigating differences in test-takers' use of cognitive and metacognitive strategies in audio-only and video-based listening comprehension test. *System*, 114, Article 103017. <https://doi.org/10.1016/j.system.2023.103017>
- Kwon, S. K., & Yu, G. (2024). The effect of viewing visual cues in a listening comprehension test on second language learners' test-taking process and performance: An eye-tracking study. *Language Testing*, 41(3), 649–680.
<https://doi.org/10.1177/02655322241239356>
- Kyle, K., & Eguchi, M. (2023). Assessing spoken lexical and lexicogrammatical proficiency using features of word, bigram, and dependency bigram use. *Modern Language Journal*, 107, 531–564. <https://doi.org/10.1111/modl.12845>
- Kyle, K., & Eguchi, M. (2024). Evaluating NLP models with written and spoken L2 samples.

- Research Methods in Applied Linguistics*, 3(2), Article 100120.
<https://doi.org/10.1016/j.rmal.2024.100120>
- Kyle, K., Choe, A. T., Eguchi, M., LaFlair, G., & Ziegler, N. (2021). A comparison of spoken and written language use in traditional and technology-mediated learning environments. *ETS Research Report Series*, 2021(1), 1–29.
<https://doi.org/10.1002/ets2.12329>
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170.
<https://doi.org/10.1080/15434303.2020.1844205>
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781–812. <https://doi.org/10.1017/S0272263120000546>
- Kyle, K., Eguchi, M., Choe, A. T., & LaFlair, G. (2022). Register variation in spoken and written language use across technology-mediated and non-technology-mediated learning environments. *Language Testing*, 39(4), 618–648.
<https://doi.org/10.1177/02655322211057868>
- Kyllonen, P., Sevak, A., Ober, T., Choi, I., Sparks, J., & Fishtein, D. (2024). Charting the future of assessments. *ETS Research Report Series*, 2024(1), 1–62.
<https://doi.org/10.1002/ets2.12388>

L

- LaFlair, G. T. (2024). An industry perspective on Open Science: A response to Winke. *Language Testing*, 41(4), 865–871. <https://doi.org/10.1177/02655322241261716>
- Lam, B. P., Sheng, L., & Zhang, X. (2023). Children's likelihood to perform adult-like in word association test: Effects of bilingualism and distributional properties of word relationships. *Bilingualism: Language and Cognition*, 26(1), 216–230.
<https://doi.org/10.1017/S1366728922000463>
- Lam, D. M. K. (2021). Don't turn a deaf ear: A case for assessing interactive listening. *Applied Linguistics*, 42(4), 740–764. <https://doi.org/10.1093/applin/amaa064>
- Lam, D. M. K. (2021). Feedback as a learning-oriented assessment practice: Principles, opportunities, and challenges. In A. Gebriel (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 86–107). Routledge.
<https://doi.org/10.4324/9781003014102-7>
- Lam, D. M. K., Green, A., Murray, N., & Gayton, A. (2021). How are IELTS scores set and used for university admissions selection: A cross-institutional case study. *IELTS Research Reports Online Series*, 3. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/for-researchers/research-reports/online-series-2021-3>
- Lam, M. K. D., Galaczi E. D., Nakatsuhara, F., & May, L. (2023). Assessing interactional competence: Exploring ratability challenges. *Applied Pragmatics*, 5(2), 208–233.
<https://doi.org/10.1075/ap.00014.lam>
- Lam, R. (2024). E-portfolios for language assessment. In R. Lam (Ed.), *Integrating e-portfolios into L2 classrooms: Education for future* (pp. 44–56). Multilingual Matters. <https://doi.org/10.2307/jj.21995549.8>
- Lambert, C., Aubrey, S., & Leeming, P. (2021). Task preparation and second language speech production. *TESOL Quarterly*, 55(2), 331–365.

- <https://doi.org/10.1002/tesq.598>
- Lamprianou, I., Tsagari, D., & Kyriakou, N. (2021). The longitudinal stability of rating characteristics in an EFL examination: Methodological and substantive considerations. *Language Testing*, 38(2), 273–301. <https://doi.org/10.1177/0265532220940960>
- Lamprianou, I., Tsagari, D., & Kyriakou, N. (2023). Experienced but detached from reality: Theorizing and operationalizing the relationship between experience and rater effects. *Assessing Writing*, 56, Article 100713. <https://doi.org/10.1016/j.asw.2023.100713>
- Lampropoulou, L. (2023). The use and impact of pre-task planning time in the monologic task of LanguageCert speaking tests. *Language Education & Assessment*, 6(1), 1–18. <https://doi.org/10.29140/lea.v6n1.1180>
- Lane, P. (2023). From silence to silencing? Contradictions and tensions in language R=revitalization. *Applied Linguistics*, 44(5), 833–847. <https://doi.org/10.1093/applin/amac075>
- Lange, K., & Matthews, J. (2023). Analyzing trends in the aural decoding errors of Japanese EFL learners. *International Review of Applied Linguistics in Language Teaching*, 61(2), 545–570. <https://doi.org/10.1515/iral-2020-0050>
- Latif, M. M. A. (2021). Remodeling writers' composing processes: Implications for writing assessment. *Assessing Writing*, 50, Article 100547. <https://doi.org/10.1016/j.asw.2021.100547>
- Latif, M. M. A., Alsuhaibani, Z., & Alsahil, A. (2024). Matches and mismatches between Saudi university students' English writing feedback preferences and teachers' practices. *Assessing Writing*, 61, 100863. <https://doi.org/10.1016/j.asw.2024.100863>
- Latif, M. W. (2021). Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: Implications for teacher professional development. *Language Testing in Asia*, 11, Article 19. <https://doi.org/10.1186/s40468-021-00130-9>
- Latif, M. W., & Wasim, A. (2022). Teacher beliefs, personal theories and conceptions of assessment literacy—a tertiary EFL perspective. *Language Testing in Asia*, 12(1), Article 11. <https://doi.org/10.1186/s40468-022-00158-5>
- Latif, M. W., & Wasim, A. (2024). Investigating EFL instructors' approaches to classroom-based assessment culture: An explanatory sequential mixed-method approach. *Language Testing in Asia*, 14, Article 57. <https://doi.org/10.1186/s40468-024-00332-x>
- Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62–85. <https://doi.org/10.1177/0265532220929918>
- Latimer, N., & Chan, S. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: Benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Latimer, N., & Chan, S. H. C. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Lau, W. M. E., Aubrey, S., & Ellis, R. (2024). Error logs in the second language classroom: Exploring the relationship between learner engagement with written corrective feedback and improvements in writing accuracy. *System*, 123, 103329. <https://doi.org/10.1016/j.system.2024.103329>

- Laudenbach, M., Brown, D. W., Guo, Z., Ishizaki, S., Reinhart, A., & Weinberg, G. (2024). Visualizing formative feedback in statistics writing: An exploratory study of student motivation using DocuScope Write & Audit. *Assessing Writing*, 60, 100830. <https://doi.org/10.1016/j.asw.2024.100830>
- Lázaro-Ibarrola, A. (2023). Model texts in collaborative and individual writing among EFL children: noticing, incorporations, and draft quality. *International Review of Applied Linguistics in Language Teaching*, 61(2), 329-359. <https://doi.org/10.1515/iral-2020-0160>
- Le Bouthillier, J., & Kristmanson, P. (2023). Becoming a French second language teacher: Supporting confidence and competence. *Second Language Teacher Education*, 2(1), 21–42. <https://doi.org/10.1558/slte.24011>
- Lecouvet, M. (2023). Non-canonical word order as a measure of syntactic complexity in advanced L2 German. *International Review of Applied Linguistics in Language Teaching*, 61(3), 877–907. <https://doi.org/10.1515/iral-2021-0029>
- Lee, C., & Qian, H. (2022). Hybrid threshold-based sequential procedures for detecting compromised items in a computerized adaptive testing licensure exam. *Educational and Psychological Measurement*, 82(4), 782–810. <https://doi.org/10.1177/00131644211023868>
- Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, 97, Article 102461. <https://doi.org/10.1016/j.system.2021.102461>
- Lee, E. (2022). Examining the social consequences of a locally-developed placement test using test takers' attitudes. *Assessing Writing*, 53, Article 100626. <https://doi.org/10.1016/j.asw.2022.100626>
- Lee, H-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring Interaction in Video-call Paired Speaking Tests: A Look at Scores, Language, and Perceptions, *Applied Linguistics*, 45(3), 533–552. <https://doi.org/10.1093/applin/amad046>
- Lee, H. (2023). Looking into an innovative test mode in paired speaking from the perspective of scores. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 13–32). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Lee, H.-W., & Galaczi, E. (Eds.). (2023). *Exploring video-call paired speaking tests*. Research Notes, 86. Retrieved from: <https://www.cambridgeenglish.org/english-research-group/published-research/research-notes/>
- Lee, H.-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring interaction in video-call paired speaking tests: A look at scores, language, and perceptions. *Applied Linguistics*, 45(3), 533-552. <https://doi.org/10.1093/applin/amad046>
- Lee, H., & Galaczi, E. (2023). Insights into video-call paired speaking tests: Research background and design. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 5–12). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Lee, H., Patel, M., Lynch, J., & Galaczi, E. D. (2021). Development of the IELTS Video Call Speaking Test: Phase 4 operational research trial and overall summary of a four-phase test development cycle. *IELTS Partnership Research Papers*, 2021/1. IELTS Partners: British Council, Cambridge Assessment English and IDP: IELTS

- Australia. www.ielts.org/teaching-and-research/research-reports
- Lee, I. (2021). The development of feedback literacy for writing teachers. *TESOL Quarterly*, 55(3), 1048–1059. <https://doi.org/10.1002/tesq.3012>
- Lee, I. (2023). Problematising written corrective feedback: A global Englishes perspective. *Applied Linguistics*, 44(4), 791–796. <https://doi.org/10.1093/applin/amad038>
- Lee, I., Karaca, M., & Inan, S. (2023). The development and validation of a scale on L2 writing teacher feedback literacy. *Assessing Writing*, 57, Article 100743. <https://doi.org/10.1016/j.asw.2023.100743>
- Lee, J. (2024). Analysis of ESL preservice teachers' collaboration with ESL learners through a language assessment development project. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence: Case studies from around the world* (Studies in Language Testing, Vol. 56, pp. 43–55). Cambridge University Press.
- Lee, J. J., Tytko, T., & Larkin, R. (2021). (Un)attended this/these in undergraduate student writing: A corpus analysis of high- and low-rated L2 writers. *Journal of English for Academic Purposes*, 50, Article 100967. <https://doi.org/10.1016/j.jeap.2021.100967>
- Lee, J., Goldenberg, N., & Williams, D. P. (2023). Three stakeholders' reflections on language assessment literacy. *ELT Journal*, 77(4), 383–392. <https://doi.org/10.1093/elt/ccad026>
- Lee, K. (2023). Diagnosing Chinese college-level English as a Foreign Language (EFL) learners' integrated writing capability: A Log-linear Cognitive Diagnostic Modeling (LCDM) study. *Assessing Writing*, 57, Article 100730. <https://doi.org/10.1016/j.asw.2023.100730>
- Lee, M., & Jung, J. (2024). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*, 28(4), 1552–1571. <https://doi.org/10.1177/13621688211034640>
- Lee, S. (2022). Evaluating the new TOEFL ITP® speaking test: insights from field test takers. *ETS Research Report Series*, 2022(1), 1–19. <https://doi.org/10.1002/ets2.12352>
- Lee, S., & Shin, S. Y. (2021). Towards improved assessment of L2 collocation knowledge. *Language Assessment Quarterly*, 18(4), 419–445. <https://doi.org/10.1080/15434303.2021.1908295>
- Lee, S., Lim, G. S., & Basse, R. (2021). The effect of additional time on the quality of argumentation in L2 writing assessment: A mixed-methods study. *Language Assessment Quarterly*, 18(3), 253–272. <https://doi.org/10.1080/15434303.2021.1872080>
- Leonard, K., & Bonilla, N. (2022). Teachers' oral proficiency: what happens after they graduate? *Foreign Language Annals*, 55(4), 1152–1168. <https://doi.org/10.1111/flan.12657>
- Leontjev, D., & Pollari, P. (2023). Guiding and assessing development of L2 writing process: The role of peer collaboration. *Innovation in Language Learning and Teaching*, 17(2), 440–452. <https://doi.org/10.1080/17501229.2022.2058514>
- Leontjev, D., Huhta, A., & Tolvanen, A. (2023). L2 English vocabulary breadth and knowledge of derivational morphology: One or two constructs? *Language Testing*, 40(2), 300–324. <https://doi.org/10.1177/02655322221114015>
- Lestari, S. B., & Brunfaut, T. (2023). Operationalizing the reading-into-writing construct in

- analytic rating scales: Effects of different approaches on rating. *Language Testing*, 40(3), 684–722. <https://doi.org/10.1177/02655322231155561>
- Leung, C., Evans, M., & Liu, Y. (2021). English as an additional language assessment framework: Filling a void in policy and provision in school education in England. *Language Assessment Quarterly*, 18(3), 296–315. <https://doi.org/10.1080/15434303.2020.1869745>
- Li, A. W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57, Article 100746. <https://doi.org/10.1016/j.asw.2023.100746>
- Li, C., Li, W., & Lu, X. (2024). Task complexity and L2 writing performance of young learners: Contributions of cognitive and affective factors. *The Modern Language Journal*, (108)3, 741–770. <https://doi.org/10.1111/modl.12954>
- Li, H., Hunter, C. V., & Bialo, J. A. (2022). A revisit of Zumbo's third generation DIF: how are we doing in language testing?. *Language Assessment Quarterly*, 19(1), 27–53. <https://doi.org/10.1080/15434303.2021.1988953> <https://doi.org/10.1080/15434303.2021.1963253>
- Li, J. (2021). Perceived effects of CET4 test preparation, language ability, and test performance: An exploratory study of Chinese EFL learners. *Language Education & Assessment*, 4(2), 38–58. <https://doi.org/10.29140/lea.v4n2.480>
- Li, J., & Gu, P. Y. (2023). Developing classroom-based formative assessment literacy: An EFL teacher's journey. *Chinese Journal of Applied Linguistics*, 46(2), 198–218. <https://doi.org/10.1515/CJAL-2023-0204>
- Li, J., & Huang, J. (2022). The impact of essay organization and overall quality on the holistic scoring of EFL writing: perspectives from classroom English teachers and national writing raters. *Assessing Writing*, 51, Article 100604. <https://doi.org/10.1016/j.asw.2021.100604>
- Li, J., & Yongqi Gu, P. (2024). Formative assessment for self-regulated learning: Evidence from a teacher continuing professional development programme. *System*, 125, 103414. <https://doi.org/10.1016/j.system.2024.103414>
- Li, L., & Zhang, D. (Eds.). (2022). *Reading in Chinese as an additional language: learners' development, instruction, and assessment*. Routledge. <https://doi.org/10.4324/9781003029038>
- Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2), 189–218. <https://doi.org/10.1177/0265532220932481>
- Li, M., Zhang, X., & Reynolds, B. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: an ETS corpus study. *Applied Linguistics Review*, 14(4), 847–873. <https://doi.org/10.1515/applirev-2020-0129>
- Li, S. (2024). Using the Rasch model to evaluate instrument and data quality in assessing requests in L2 Chinese. In S. Li (Ed.), *Pragmatics of Chinese as a second language* (pp. 161–184). Multilingual Matters. <https://doi.org/10.2307/jj.22679704.12>
- Li, S., & Yu, H. (2024). Effects of topic familiarity on L2 writing processes and behaviors. *International Journal of Applied Linguistics*, 34(1), 348–366. <https://doi.org/10.1111/ijal.12497>
- Li, S., Wen, T., Li, X., Feng, Y., & Lin, C. (2023). Comparing holistic and analytic marking methods in assessing speech act production in L2 Chinese. *Language Testing*, 40(2), 249–275. <https://doi.org/10.1177/02655322221113917>

- Li, X., & Wang, Y. (2024). Effects of input text genre on Chinese young EFL learners' performance on continuation tasks. *Language Testing in Asia*, 14, Article 3. <https://doi.org/10.1186/s40468-024-00273-5>
- Li, Y. K., Lin, S., Liu, Y., & Lu, X. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, 56, Article 100707. <https://doi.org/10.1016/j.asw.2023.100707>
- Li, Y., & Yang, R. (2023). Assessing the writing quality of English research articles based on absolute and relative measures of syntactic complexity. *Assessing Writing*, 55, Article 100692. <https://doi.org/10.1016/j.asw.2022.100692>
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, Article 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Liang, Y., Li, Y., & Sang, Z. (2023). A study on peer mediation in dynamic assessment of translation revision competence. *Language Assessment Quarterly*, 20(1), 108–126. <https://doi.org/10.1080/15434303.2022.2153050>
- Liao, J. (2023). What skills are being assessed? Evaluating L2 Chinese essays written by hand and on a computer keyboard. *Assessing Writing*, 57, Article 100765. <https://doi.org/10.1016/j.asw.2023.100765>
- Liao, M., Patton, J., Yan, R., & Jiao, H. (2021). Mining process data to detect aberrant test takers. *Measurement: Interdisciplinary Research and Perspectives*, 19(2), 93–105. <https://doi.org/10.1080/15366367.2020.1827203>
- Liao, R. J. (2023). The use of generalizability theory in investigating the score dependability of classroom-based L2 reading assessment. *Language Testing*, 40(1), 86–106. <https://doi.org/10.1177/02655322211070840>
- Liao, R. J. T. (2021). Advancing the international recognition of the locally-produced GEPT: An interview with Jessica Wu. *Language Assessment Quarterly*, 18(4), 446–455. <https://doi.org/10.1080/15434303.2021.1919116>
- Liao, R. J. T., Ohta, R., & Lee, K. (2024). The relationship between written discourse features and integrated listening-to-write scores for adolescent English language learners. *Language Testing*, 41(1), 35–59. <https://doi.org/10.1177/02655322231167629>
- Liao, Y. F. (2022). Using the English GSAT for placement into EFL classes: Accuracy and validity concerns. *Language Testing in Asia*, 12, Article 31. <https://doi.org/10.1186/s40468-022-00181-6>
- Lim, J., Whitehead, G. E. K., & Choi, Y. (2021). Interactive e-book reading vs. paper-based reading: Comparing the effects of different mediums on middle school students' reading comprehension. *System*, 97, Article 102434. <https://doi.org/10.1016/j.system.2020.102434>
- Lin, J. (2023). ACTFL Chinese reading proficiency guidelines: Verifying the difficulty hierarchy. *Foreign Language Annals*, 56(3), 667–689. <https://doi.org/10.1111/flan.12680>
- Lin, J. (2023). Assessing the dimensionality of Chinese as a second language reading: a confirmatory factor analysis approach. *Chinese as a Second Language Research*, 12(2), 173–204. <https://doi.org/10.1515/caslar-2023-2001>
- Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*, 55(3), 1011–1023. <https://doi.org/10.1002/tesq.3055>

- Lin, R. (2023). Examining the scoring of content integration in a listening-speaking Test: A G-Theory analysis. *Language Assessment Quarterly*, 20(3), 319–338. <https://doi.org/10.1080/15434303.2023.2242334>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 123, 103344. <https://doi.org/10.1016/j.system.2024.103344>
- Lindeberg, S., Müller, N., & Samuelsson, C. (2023). Multimodality in PPA: Hand movements as resources in conversations and testing. *Journal of Interactional Research in Communication Disorders*, 14(2), 268–291. <https://doi.org/10.1558/jircd.24306>
- Ling, G., Elliot, N., Burstein, J. C., McCaffrey, D. F., MacArthur, C. A., & Holtzman, S. (2021). Writing motivation: A validation study of self-judgment and performance. *Assessing Writing*, 48, Article 100509. <https://doi.org/10.1016/j.asw.2020.100509>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liou, G., Bonner, C. V., & Tay, L. (2022). A psychometric view of technology-based assessments. *International Journal of Testing*, 22(3–4), 216–242. <https://doi.org/10.1080/15305058.2022.2070757>
- Little, D., & Figueras, N. (2022). *Reflecting on the common European framework of reference for languages and its companion volume*. Multilingual Matters.
- Liu, C., & Yu, S. (2022). Reconceptualizing the impact of feedback in second language writing: A multidimensional perspective. *Assessing Writing*, 53, Article 100630. <https://doi.org/10.1016/j.asw.2022.100630>
- Liu, C., Sun, L., He, Y., & Wu, N. (2023). The effects of task complexity and language aptitude on EFL learners' writing performance. *Assessing Writing*, 58, Article 100791. <https://doi.org/10.1016/j.asw.2023.100791>
- Liu, F., Du, J., Zhou, D. Q., & Huang, B. (2021). Exploiting the potential of peer feedback: The combined use of face-to-face feedback and e-feedback in doctoral writing groups. *Assessing Writing*, 47, Article 100482. <https://doi.org/10.1016/j.asw.2020.100482>
- Liu, G. (2022). Schema association training: Empowering students for speaking assessments. *RELC Journal*, 53(1), 216–224. <https://doi.org/10.1177/0033688220927099>
- Liu, M., Al-Hoorie, A. H., & Hiver, P. (2024). Open Access in language testing and assessment: The case of two flagship journals. *Language Testing*, 41(4), 703–728. <https://doi.org/10.1177/02655322241261417>
- Liu, S., & Lin, D. (2022). Developing and validating an analytic rating scale for a paraphrase task. *Assessing Writing*, 53, Article 100646. <https://doi.org/10.1016/j.asw.2022.100646>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78–105. <https://doi.org/10.1257/73480>
- Liu, T. & Aryadoust, V. (2022). The effect of in-class and one-on-one video feedback on EFL learners' English public speaking competency and anxiety. *Studies in Language Assessment*, 11(2), 25–57. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4396264/SiLA_11.2_Liu_an

d_Aryadoust.pdf

- Liu, T., & Aryadoust, V. (2024). Does modality matter? A meta-analysis of the effect of video input in L2 listening assessment. *System*, 120, 103191. <https://doi.org/10.1016/j.system.2023.103191>
- Liu, T., Aryadoust, V., & Foo, S. (2022). Examining the factor structure and its replicability across multiple listening test forms: Validity evidence for the Michigan English test. *Language Testing*, 39(1), 142–171. <https://doi.org/10.1177/02655322211018139>
- Liu, X., & Jane Rogers, H. (2022). Treatments of differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, 82(2), 225–253. <https://doi.org/10.1177/00131644211012050>
- Liu, X., & Read, J. (2021). Investigating the skills involved in reading test tasks through expert judgement and verbal protocol analysis: Convergence and divergence between the two methods. *Language Assessment Quarterly*, 18(4), 357–381. <https://doi.org/10.1080/15434303.2021.1881964>
- Liu, Y., & Liu, J. (2023). Comparing computer-based and paper-based rating modes in an English writing test. *Assessing Writing*, 57, Article 100771. <https://doi.org/10.1016/j.asw.2023.100771>
- Liu, Y., Cheong, C. M., & Zhang, J. (2024). The Role of Achievement Goals and Strategy Use in Secondary School students' Integrated Writing Performance. *Language Assessment Quarterly*, 21(4–5), 324–349. <https://doi.org/10.1080/15434303.2024.2435818>
- Liu, Y., Cui, Y., & Yu, H. (2023). Test review: a general academic English summative test. *Language Testing in Asia*, 13, Article 49 (2023). <https://doi.org/10.1186/s40468-023-00263-z>
- Liu, Y., Storch, N., & Morton, J. (2022). It takes two to tango: Investigating teacher-student interactions related to written corrective feedback with activity theory. *Assessing Writing*, 53, Article 100647. <https://doi.org/10.1016/j.asw.2022.100647>
- Lloyd, J., Seed, G., Wen, X. & Yue, H. (2024). Mapping the SMEEA Gaokao tests to the CEFR. *CEFR Journal – Research and Practice*, 6, 56-79. https://cefrjapan.net/images/PDF/CEFRJournal/CEFRJournal-6-4_Lloyd_Seed_Wen_Yue_2024.pdf
- Lombardi, A. (2021). More is more: Explicit intertextuality in university writing placement exam essays. *Journal of English for Academic Purposes*, 50, Article 100955. <https://doi.org/10.1016/j.jeap.2020.100955>
- López-Gopar, M. E. (2021). Citizenship in language testing: A call for respectful collaborations. *Language Testing*, 38(4), 649–653. <https://doi.org/10.1177/02655322211022061>
- Lopez, A. A., Guzman-Orth, D., Zapata-Rivera, D., Forsyth, C. M., & Luce, C. (2021). Examining the accuracy of a conversation-based assessment in interpreting English learners' written responses. *ETS Research Report Series*, 2021(1), 1–15. <https://doi.org/10.1002/ets2.12315>
- Low, A. R. L., & Aryadoust, V. (2023). Investigating test-taking strategies in listening assessment: A comparative study of eye-tracking and self-report questionnaires. *International Journal of Listening*, 37(2), 93-112. DOI: 10.1080/10904018.2021.1883433
- Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate

- students' writing performance. *Assessing Writing*, 56, Article 100725.
<https://doi.org/10.1016/j.asw.2023.100725>
- Lu, Q., Zhu, X., Zhu, S., & Yao, Y. (2024). Effects of writing feedback literacies on feedback engagement and writing performance: A cross-linguistic perspective. *Assessing Writing*, 62, 100889. <https://doi.org/10.1016/j.asw.2024.100889>
- Lu, X.-L. & Han, C. (2023). Automatic assessment of spoken-language interpreting based on machine translation evaluation metrics: A multi-scenario exploratory study. *Interpreting: International Journal of Research and Practice in Interpreting*, 25(1), 109–143. <https://doi.org/10.1075/intp.00076.lu>
- Lu, Y. (2022). *Assessing learners' competence in L2 Chinese*. Routledge.
- Luft Baker, D., McCoach, B. D., Ware, S., Coyne, M. D., & Rattan, S. M. (2022). Effects of Spanish vocabulary knowledge on the English word knowledge and listening comprehension of bilingual students. *International Journal of Bilingual Education and Bilingualism*, 25(6), 2269–2283.
<https://doi.org/10.1080/13670050.2021.1908219>
- Lukácsi, Z. (2021). Developing a level-specific checklist for assessing EFL writing. *Language Testing*, 38(1), 86–105. <https://doi.org/10.1177/0265532220916703>
- Lundgren, E., & Eklöf, H. (2023). Questionnaire-taking motivation: Using response times to assess motivation to optimize on the PISA 2018 student questionnaire. *International Journal of Testing*, 23:4, 231-256. <https://doi.org/10.1080/15305058.2023.2214647>
- Luo, Y. (2021). A comparison of common IRT model-selection methods with mixed-format tests. *Measurement: Interdisciplinary Research and Perspectives*, 19(4), 199–212.
<https://doi.org/10.1080/15366367.2021.1878779>
- Luquin, M., & García Mayo, M. D. P. (2024). A longitudinal study of the effects of model texts on EFL children's written production. *System*, 120, 103190.
<https://doi.org/10.1016/j.system.2023.103190>
- Lv, Z. (2024). Prompt effects on assessment of interactional competence in a group oral test. *International Journal of Applied Linguistics*, 34(3), 1089-1107.
<https://doi.org/10.1111/ijal.12554>
- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49, Article 100529.
<https://doi.org/10.1016/j.asw.2021.100529>
- Lydster, C. (2024). Who says “Ps get Degrees”? Examining the Profile of Undergraduate Students Maintaining High Achievement at University. *Student Success*, 15(2), 23–33. <https://doi.org/10.5204/ssj.3402>

M

- Ma, H., & Chong, S. W. (2022). Predictability of IELTS in a high-stakes context: A mixed methods study of Chinese students' perspectives on test preparation. *Language Testing in Asia*, 12, Article 2. <https://doi.org/10.1186/s40468-021-00152-3>
- Ma, H., Wang, J., & He, L. (2024). Linguistic features distinguishing students' writing ability aligned with CEFR levels. *Applied Linguistics*, 45(4), 637-657.
<https://doi.org/10.1093/applin/amad054>
- Ma, M. (2023). Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives. *Language Testing in Asia*, 13, Article 33 (2023).
<https://doi.org/10.1186/s40468-023-00249-x>

- Ma, M., & Bui, G. (2022). Implementing continuous assessment in an academic English writing course: An exploratory study. *Assessing Writing*, 53, Article 100629. <https://doi.org/10.1016/j.asw.2022.100629>
- Ma, W. (2022). What the analytic versus holistic scoring of international teaching assistants can reveal: Lexical grammar matters. *Language Testing*, 39(2), 239–264. <https://doi.org/10.1177/02655322211040020>
- Ma, W., & Winke, P. (2022). An investigation of the impact of jagged profile on L2 speaking test ratings: Evidence from rating and eye-tracking data. *Language Assessment Quarterly*, 19(4), 394–421. <https://doi.org/10.1080/15434303.2022.2078720>
- Maamuujaav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, Article 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Mabrouk, F. (2023). Tunisian EFL learners' perceptions of the use of oral presentations in assessing speaking skills: An exploratory practice study. *Studies in Applied Linguistics and TESOL*, 23(3), 18–40. <https://doi.org/10.52214/salt.v23i2.12083>
- MacGregor, D., Yen, S. J., & Yu, X. (2022). Using multistage testing to enhance measurement of an english language proficiency test. *Language Assessment Quarterly*, 19(1), 54–75. <https://doi.org/10.1080/15434303.2021.1988953>
- Macqueen, S, Pill, J., & Knoch, U. (2021). Trust the test: Score-user perspectives on the roles of language tests in professional registration and skilled migration. *Papers in Language Testing and Assessment*, 10(1), 49–69. http://www.altanz.org/uploads/5/9/0/8/5908292/4._macqueen_pill_knoch.pdf
- Macqueen, S. (2024). [Review of the book *Reconsidering Context in Language Assessment: Transdisciplinary Perspectives, Social Theories, and Validity* by J. Fox and N. Artemeva]. *Language Testing*, 41(1), 220–226. <https://doi.org/10.1177/02655322231199501>
- Madyarov, I., Movsisyan, V., Madoyan, H., Galikyan, I., & Gasparyan, R. (2021). New validity evidence on the TOEFL Junior® standard test as a measure of progress. *ETS Research Report Series*, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12334>
- Mahalingappa, L., Zong, J., & Polat, N. (2024). The impact of captioning and playback speed on listening comprehension of multilingual English learners at varying proficiency levels. *System*, 120, 103192. <https://doi.org/10.1016/j.system.2023.103192>
- Mäkipää, T. (2021). Students' and teachers' perceptions of self-assessment and teacher feedback in foreign language teaching in general upper secondary education – A case study in Finland. *Cogent Education*, 8(1), 1978622. <https://doi.org/10.1080/2331186X.2021.1978622>
- Mäkipää, T. (2023). Students' perceptions of teacher assessment practices in foreign language emergency remote teaching during Covid-19 in Finland. *Studies in Language Assessment*, 12(1), 1–25. <https://doi.org/10.58379/AHRY2151>
- Mäkipää, T., & Soltyska, A. (2024). English teachers' perceptions of feedback practices under emergency remote teaching conditions: Insights for future use. *Studies in Language Assessment (SILA)*, 13(2), 73–114. <https://doi.org/10.58379/mszy9375>
- Mäkipää, T., Hahl, K., & Luodonpää-Manni, M. (2021). Teachers' perceptions of assessment and feedback practices in Finland's foreign language classes during the Covid-19 pandemic. *CEPS Journal*, 11, 219–240. <https://doi.org/10.26529/cepsj.1108>
- Mangold, M., Lenhard, W., Schindler, J., Schulz, D., & Richter, T. (2024). Assessing

- vocabulary of bilingual German-Turkish preschool children. *Bilingualism: Language and Cognition*, 27(4), 678–689. <https://doi.org/10.1017/S1366728923000822>
- Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103, Article 102674. <https://doi.org/10.1016/j.system.2021.102674>
- Mao, Z., & Lee, I. (2022). Researching L2 student engagement with written feedback: Insights from sociocultural theory. *TESOL Quarterly*, 56(2), 788–798. <https://doi.org/10.1002/tesq.3071>
- Maplethorpe, L., Kim, H., Hunte, M. R., Vincett, M., & Jang, E. E. (2022). Student-generated questions in literacy education and assessment. *Journal of Literacy Research*, 54(1), 74–97. <https://doi.org/10.1177/1086296X221076436>
- Marinho, N. L., Witmer, S. E., Jess, N., & Roschmann, S. (2023). A meta-Analysis of accommodation effects for English learners: Considering possible moderators. *Language Assessment Quarterly*, 20(3), 296–318. <https://doi.org/10.1080/15434303.2023.2237499>
- Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *Modern Language Journal*, 107, 669–692. <https://doi.org/10.1111/modl.12866>
- Martín-Gascón, B. (2023). Developing L2 learners' metaphoric competence: a case study of figurative motion constructions. *International Review of Applied Linguistics in Language Teaching*, 61(1), 79–109. <https://doi.org/10.1515/iral-2022-0043>
- Martin, C., & Dockrell, J. E. (2024). Writing productivity development in elementary school: A systematic review. *Assessing Writing*, 60, 100834. <https://doi.org/10.1016/j.asw.2024.100834>
- Martin, J. L., & Wright, K. E. (2023). Bias in automatic speech recognition: The case of African American language. *Applied Linguistics*, 44(4), 613–630. <https://doi.org/10.1093/applin/amac066>
- Mashhadlou, H., & Izadpanah, S. (2021). Assessing Iranian EFL teachers' educational performance based on gender and years of teaching experience. *Language Testing in Asia*, 11, Article 23. <https://doi.org/10.1186/s40468-021-00140-7>
- Matsumura, L. C., Wang, E. L., Correnti, R., & Litman, D. (2023). Tasks and feedback: An exploration of students' opportunity to develop adaptive expertise for analytic text-based writing. *Assessing Writing*, 55, Article 100689. <https://doi.org/10.1016/j.asw.2022.100689>
- Matta, M., Mercer, S. H., & Keller-Margulis, M. A. (2022). Evaluating validity and bias for hand-calculated and automated written expression curriculum-based measurement scores. *Assessment in Education: Principles, Policy & Practice*, 29(2), 200–218. <https://doi.org/10.1080/0969594X.2022.2043240>
- Matthews, J. (2023). Using inter-rater discourse to trace the origins of disagreement: towards collective reflective practice in L2 assessment. *RELC Journal*, 54(1), 99–113. <https://doi.org/10.1177/0033688220977373>
- Mawer, V. (2023). Am I talking enough? Evaluating the effectiveness of quantitative speaking feedback. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 73–82). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022).

- Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, 52, Article 100608.
<https://doi.org/10.1016/j.asw.2022.100608>
- McCloy R. A., Everson H. T., & Purpura, J. E. (2023). Next Steps to Support CAT ECL 2.0: Report from a Workshop. HumRRO Technical Report 166. Alexandria, VA: Human Development Resources Research Organization.
- McDonough, K., & Uludag, P. (2021). Individual and shared assessment of ESL students' paired oral test performance: Examining rater judgments and lexico-grammatical features. In W. Crawford (Ed.), *Multiple perspectives on learner interaction: The corpus of collaborative oral tasks* (pp. 69–90). Mouton De Gruyter.
<https://doi.org/10.1515/9781501511370-005>
- McDonough, K., Uludag, P. & Neumann, H. (2021). Assessing case study exams in a business course: What role does language play? *Business and Professional Communication Quarterly*, 82(2), 116–134.
<https://doi.org/10.1177%2F23294906211012398>
- McDonough, K., Uludag, P., & Neumann, H. (2021). Instructor evaluation of business student writing: Does language play a role? *Business and Professional Communication Quarterly*, 84(2), 116–134.
<https://doi.org/10.1177/23294906211012398>
- McElwee, S., Y.F. Cheung, K., R.T. Cromie, S., Shannon, M., & Gallacher, T. (2021). Revising the BioMedical Admissions Test (BMAT) to improve impact and washback for candidates and support fair access to test preparation. *Assessment in Education: Principles, Policy & Practice*, 28(1), 77–95.
<https://doi.org/10.1080/0969594X.2019.1639133>
- McKay, T. H., Teimouri, Y., Sağdıç, A., Salen, B., Reagan, D., & Malone, M. E. (2021). The cagey C-test construct: Some evidence from a meta-analysis of correlation coefficients. *System*, 99, Article 102526.
<https://doi.org/10.1016/j.system.2021.102526>
- McLeod, M. (2023). English language proficiency requirements for migration to Canada, Australia, the United Kingdom, and the United States, and the implications for language testing research. *Canadian Journal of Applied Linguistics*, 26(1), 1–27.
<https://doi.org/10.37213/cjal.2023.32825>
- McLeod, M., & Cheng, L. (2023). The Canadian English Language Proficiency Index Program (CELPPI) Test. *Language Assessment Quarterly*, 20(3), 339–352.
<https://doi.org/10.1080/15434303.2023.2237487>
- McNamara, D. S., & Kendeou, P. (2022). The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*, 29(2), 150–182.
<https://doi.org/10.1080/0969594X.2022.2037509>
- Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2021). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 34(7), 868–897.
<https://doi.org/10.1080/09588221.2019.1645181>
- Mei, H., & Chen, H. (2024). Cognitive Diagnosis in Language Assessment: A Thematic Review. *RELC Journal*, 55(2), 529–537.
<https://doi.org/10.1177/00336882221122357>
- Meletiadiou, E., & Tsagari, D. (2022). Exploring EFL teachers' perceptions of the use of peer assessment in external exam-dominated writing classes. *Languages*, 7(1), 16.

- <https://doi.org/10.3390/languages7010016>
- Meng, Y., & Fu, H. (2023). Modeling mediation in the dynamic assessment of listening ability from the cognitive diagnostic perspective. *Modern Language Journal*, 107, 137–160. <https://doi.org/10.1111/modl.12820>
- Meng, Y., Fu, H., & Wang, C. (2024). The effectiveness of computerized listening dynamic assessment: Attribute-based mediation model. *Language Learning & Technology*, 28(1), 1–28. <https://hdl.handle.net/10125/73580>
- Menold, N., & Raykov, T. (2022). On the relationship between item stem formulation and criterion validity of multiple-component measuring instruments. *Educational and Psychological Measurement*, 82(2), 356–375. <https://doi.org/10.1177/0013164420988169>
- Mesina, V. G. (2023). The “Problem” of L2 writers in college composition placement: Identity, outcomes, and the future of directed self placement. *Studies in Applied Linguistics and TESOL*, 22(2), 23–36. <https://doi.org/10.52214/salt.v22i2.10682>
- Miao, Y. (2023). The relationship among accent familiarity, shared L1, and comprehensibility: A path analysis perspective. *Language Testing*, 40(3), 723–747. <https://doi.org/10.1177/02655322231156105>
- Mikami, Y. (2024). Classification of multiple text reading test formats by analyzing English language proficiency tests. *Language Testing in Asia*, 14, Article 12. <https://doi.org/10.1186/s40468-024-00285-1>
- Milliner, B., & Dimoski, B. (2024). The effects of a metacognitive intervention on lower-proficiency EFL learners’ listening comprehension and listening self-efficacy. *Language Teaching Research*, 28(2), 679–713. <https://doi.org/10.1177/13621688211004646>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68, Article 100963. <https://doi.org/10.1016/j.stueduc.2020.100963>
- Min, S., & Bishop, K. (2024). A shortened test is feasible: Evaluating a large-scale multistage adaptive English language assessment. *Language Testing*, 41(3), 627–648. <https://doi.org/10.1177/02655322231225426>
- Min, S., & He, L. (2022). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, 39(1), 90–116. <https://doi.org/10.1177/0265532221995475>
- Min, S., Bishop, K., & Gary Cook, H. (2022). Reading is a multidimensional construct at child-L2-English-literacy onset, but comprises fewer dimensions over time: Evidence from multidimensional IRT analysis. *Language Testing*, 39(2), 265–288. <https://doi.org/10.1177/02655322211045296>
- Min, S., Cai, H., & He, L. (2022). Application of bi-factor MIRT and higher-order CDM models to an in-house EFL listening test for diagnostic purposes. *Language Assessment Quarterly*, 19(2), 189–213. <https://doi.org/10.1080/15434303.2021.1980571>
- Min, S., Zhang, J., Li, Y., & He, L. (2022). Bridging local needs and national standards: Use of standards-based individualized feedback of an in-house EFL listening test in China. *Language Testing*, 39(3), 425–452. <https://doi.org/10.1177/02655322211070990>
- Mitchell, K. M., McMillan, D. E., Lobchuk, M. M., Nickel, N. C., Rabbani, R., & Li, J. (2021).

- Development and validation of the situated academic writing self-efficacy scale (SAWSES). *Assessing Writing*, 48, Article 100524. <https://doi.org/10.1016/j.asw.2021.100524>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.
- Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), Article 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
- Mohammadi, M., & Sanavi, R. V. (2021). Language assessment literacy: Ontogenetic and phylogenetic perspectives. In S. Hidri (Eds.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 52-65). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003016083-4/language-assessment-literacy-mojtaba-mohammadi-reza-vahdani-sanavi>
- Mohammadi, M., Abbasian, G. R., & Siyyari, M. (2022). Adaptation and validation of a critical thinking scale to measure the 3D critical thinking ability of EFL readers. *Language Testing in Asia*, 12, Article 24. <https://doi.org/10.1186/s40468-022-00173-6>
- Mohebbi, H. (2021). 25 years on, the correction debate continues: An interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), Article 3. <https://doi.org/10.1186/s40862-021-00110-9>
- Mohsen, M. A. (2024). L1 versus L2 writing processes: What insight can we obtain from a keystroke logging program? *Language Teaching Research*, 28(6), 2251-2275. <https://doi.org/10.1177/13621688211041292>
- Molokopeeva, T., & Simard, D. (2024). Interaction between levels of text representation and working memory during L2 reading comprehension: What about it?. *International Journal of Applied Linguistics*, 34(2), 568-585. <https://doi.org/10.1111/ijal.12516>
- Monfils, L. F., & Manna, V. F. (2021). Time to achieving a designated criterion score level: A survival analysis study of test taker performance on the TOEFL iBT® test. *Language Testing*, 38(1), 154–176. <https://doi.org/10.1177/0265532220940709>
- Monster, I., Tellings, A., Burk, W. J., Keuning, J., Segers, E., & Verhoeven, L. (2021). Assessing children's incremental word knowledge in the upper primary grades. *Language Testing*, 38(4), 536–557. <https://doi.org/10.1177/0265532220961541>
- Montasseri, Z., & Ahmadi, A. (2024). Validity argument of an Interactional Competence checklist: Perspectives and many-facets Rasch analysis. *Studies in Language Assessment (SILA)*, 13(1), 69-107. <https://doi.org/10.58379/MEZB3039>
- Monteiro, K., Crossley, S., Botarleanu, R.-M., & Dascălu, M. (2023). L2 and L1 semantic context indices as automated measures of lexical sophistication. *Language Testing*, 40(3), 576-606. <https://doi.org/10.1177/02655322221147924>
- Mootoosamy, K., & Aryadoust, V. (2024). Transitioning from communicative competence to multimodal and intercultural competencies: A systematic review. *Societies*, 14(115). <https://doi.org/10.3390/soc14070115>
- Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking anxiety and task complexity effects on second language speech. *International Journal of Applied Linguistics*, 34(1), 292-315. <https://doi.org/10.1111/ijal.12494>
- Moradi, E., Ghabanchi, Z., & Pishghadam, R. (2022). Reading comprehension test fairness across gender and mode of learning: Insights from IRT-based differential item

- functioning analysis. *Language Testing in Asia*, 12, Article 39.
<https://doi.org/10.1186/s40468-022-00192-3>
- Moradkhan, D., & Shahmirzadi, N. (2012). Utilizing measures of (CAF) complexity, accuracy and fluency for interpreter evaluation. *Translation Studies (Motale'at-e-Tarjome Quarterly)*, 10(39), 87–104, (ISSN: 1735–0212).
- Moraes, I. T. (2024). Concepções e práticas de avaliação em diálogo em cursos de formação de professores/as de inglês em instituições de ensino superior do Paraná [Assessment Concepts and Practices in Dialogue in English Teacher Education Programs in Higher Education Institutions in the State of Paraná]. *Revista X*, 19(1), 327-352. <https://doi.org/10.5380/rvx.v19i2.94796>
- Moraes, I. T. & Batista, E. G. (2020). Letramento em avaliação para professores de línguas estrangeiras para crianças: orientações teórico-práticas [Assessment literacy for foreign language teachers of children: theoretical and practical guidelines]. *Revista Horizontes de Linguística Aplicada*, 19(2), 15-42.
<https://doi.org/10.26512/rhla.v19i2.26804>
- Morea, N., Kasprowicz, R. E., Morrison, A., & Silvestri, C. (2024). Diverse population, homogenous ability: The development of a new receptive vocabulary size test for young language learners in England using Rasch analysis. *Research Methods in Applied Linguistics*, 3(3), 100166. <https://doi.org/10.1016/j.rmal.2024.100166>
- Morrison, A., & Sepulveda-Escobar, P. (2024). Chilean English as a Foreign Language Teacher Educators' Conceptions and Practices of Online Assessment. *REL C Journal*, 55(2), 394–407. <https://doi.org/10.1177/00336882221079059>
- Morton, T., & Nashaat-Sobhy, N. (2024). Exploring Bases of achievement in content and Language Integrated assessment in a bilingual education program. *TESOL Quarterly*, 58(1), 5–31. <https://doi.org/10.1002/tesq.3207>
- Mostafa, T., Crossley, S. A., & Kim, Y. (2021). Predictors of English as second language learners' oral proficiency development in a classroom context. *International Journal of Applied Linguistics*, 31(3), 526–548. <https://doi.org/10.1111/ijal.12358>
- Motteram, J., Spiby, R., Bellhouse, G., & Sroka, K. (2023). Implementation of an accommodations policy for candidates with diverse needs in a large-scale testing system. *Language Testing*, 40(4), 904-932.
<https://doi.org/10.1177/02655322231166587>
- Mujtaba, S. M., Reynolds, B. L., Parkash, R., & Singh, M. K. M. (2021). Individual and collaborative processing of written corrective feedback affects second language writing accuracy and revision. *Assessing Writing*, 50, Article 100566.
<https://doi.org/10.1016/j.asw.2021.100566>
- Müller, A. & Han, W. (2022). *IELTS writing band scores 5.5–7.5: Grammatical error rates, stakeholder perceptions, and risk* (IELTS Research Reports Online Series, No. 1/22). International English Language Testing System.
<https://www.ielts.org/-/media/research-reports/muller-and-han-2022.ashx>
- Müller, A., & Brenner, A. (2024). Benchmarking English standards across professions and professional university degrees IELTS Research Reports Online Series, No. 4/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at: <https://ielts.org/researchers/our-research/research-reports>
- Mulloly, A. (2023). Operational considerations for video-call delivery of paired format speaking tests. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 87–96). Cambridge University Press & Assessment.

- <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Mullooly, A., & Glasson, N. (2023). Investigating language functions elicited by in-person and online paired speaking tests. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 33–52). Cambridge.
<https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Mumford, S., & Atay, D. (2021). Teachers' perspectives on the causes of rater discrepancy in an English for Academic Purposes context. *Assessing Writing*, 48, Article 100527. <https://doi.org/10.1016/j.asw.2021.100527>
- Murphy Odo, D. (2022). The Effect of Automatic Text Simplification on L2 Readers' Text Comprehension. *Applied Linguistics*, 44(6), 1030–1046. <https://doi.org/10.1093/applin/amac057>
- Murphy, D., & Sedivy, S. K. (2024). The speaking proficiency outcomes of face-to-face and online intensive summer LCTL programs. *Foreign Language Annals*, 57(4), 872–899. <https://doi.org/10.1111/flan.12758>
- Muszyńska, K., Łuniewska, M., Dynak, A., Kolak, J., Lohrum, R., Otwinowska, A., ... Haman, E. (2024). Bilinguals' knowledge of 'home' and 'school' words revisited: evidence from Polish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 28(1), 73–91. <https://doi.org/10.1080/13670050.2024.2399639>
- Myburgh-Smit J. & Weideman, A. (2021). How early should we measure academic literacy? The usefulness of an appropriate test of academic literacy for Grade 10 students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 117–131). Multilingual Matters. <https://doi.org/10.21832/9781788926218-007>
- Myint Maw, T. M., Clenton, J., & Higginbotham, G. (2022). Investigating whether a flemma count is a more distinctive measurement of lexical diversity. *Assessing Writing*, 53, Article 100640. <https://doi.org/10.1016/j.asw.2022.100640>

N

- Nagle, C. (2022). Rethinking pronunciation posttesting. *Journal of Second Language Pronunciation*, 8(2), 161–167. <https://doi.org/10.1075/jslp.22019.nag>
- Nagle, C. L., Huensch, A., & Zárate-Sánchez, G. (2023). Exploring phonetic predictors of intelligibility, comprehensibility, and foreign accent in L2 Spanish speech. *Modern Language Journal*, 107, 202–221. <https://doi.org/10.1111/modl.12827>
- Nagle, C., & Rehman, I. (2021). Doing L2 speech research online: Why and how to collect online ratings data. *Studies in Second Language Acquisition*, 43(4), 916–939. <https://doi.org/10.1017/S0272263121000292>
- Nahatame, S. (2023). Predicting processing effort during L1 and L2 reading: The relationship between text linguistic features and eye movements. *Bilingualism: Language and Cognition*, 26(4), 724–737. <https://doi.org/10.1017/S136672892200089X>
- Naismith, B., Attali, Y., & LaFlair, G. T. (2024). The impact of task duration on the scoring of independent writing responses of adult L2-English writers. *Assessing Writing*, 62, 100895. <https://doi.org/10.1016/j.asw.2024.100895>
- Nakamura, K. (2022). Investigating the factor structure of the Test of English for Academic Purposes (TEAP) and its relation to test takers' perceived test task value. *Language Testing in Asia*, 12, Article 35. <https://doi.org/10.1186/s40468-022-00183-4>

- Nakamura, S., Phung, L., & Reinders, H. (2021). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*, 43(2), 428–441.
<https://doi.org/10.1017/S0272263121000437>
- Nakanishi, Y., & Takeuchi, O. (2024). Validating the foreign language enjoyment scale for young learners: An exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100167.
<https://doi.org/10.1016/j.rmal.2024.100167>
- Nakata, T., Tada, S., Mclean, S., & Kim, Y. A. (2021). Effects of distributed retrieval practice over a semester: Cumulative tests as a way to facilitate second language vocabulary learning. *TESOL Quarterly*, 55(1), 248–270.
<https://doi.org/10.1002/tesq.596>
- Nakatsuhara, F., Inoue, C., & Khabbazzbashi, N. (2021). Measuring L2 speaking. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.285–294). Routledge.
<https://doi.org/10.4324/9781351034784-31>
- Nakatsuhara, F., Inoue, C., & Taylor, L. (2021). Comparing rating modes: Analysing live, audio and video ratings of IELTS Speaking Test performance. *Language Assessment Quarterly*, 18(2), 83–106.
<https://doi.org/10.1080/15434303.2020.1799222>
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2021). Video-conferencing speaking tests: Do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice*, 28(4), 369–388.
<https://doi.org/10.1080/0969594X.2021.1951163>
- Nakatsuhara, F., Inoue, C., Nitta, R., & Aoyama, T. (2023). Towards reforming the university entrance exam system in Japan: a case study of Linguaskill General. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/685092-research-notes-84.pdf>
- Nakatsuhara, F., May, L., & Khabbazzbashi, N. (2023). *Exploring the speaking construct in academic settings in a digital age*. British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262806705843094>
- Nakatsuhara, F., May, L., Inoue, C., Willcox-Ficzere, E., Westbrook, C., & Spiby, R. (2021). *Exploring the potential for assessing interactional and pragmatic competence in semi-direct speaking tests* (Report No. VS/2021/001). British Council.
https://www.britishcouncil.org/sites/default/files/nakatsuhara_et_al_2021_-_assessing_interactional_and_pragmatic_competence_in_semi-direct_speaking_tests_v2.pdf
- Namaziandost, E., Kargar Behbahani, H., & Heydarnejad, T. (2024). Tapping the alphabets of learning-oriented assessment: Self-assessment, classroom climate, mindsets, trait emotional intelligence, and academic engagement are in focus. *Language Testing in Asia*, 14, Article 21. <https://doi.org/10.1186/s40468-024-00293-1>
- Namkung, Y., & Kim, Y. (2024). Learner engagement in collaborative writing: The effects of SCMC mode, interlocutor familiarity, L2 proficiency, and task repetition. *System*, 121, 103251. <https://doi.org/10.1016/j.system.2024.103251>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, Article 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Nawas, A., Darmawan, I.N., & Maadad, N. (2023). Indonesian secular vs. Madrasah

- schools: assessing the discrepancy in English reading and listening tests. *Language Testing in Asia*, 13, Article 52 (2023). <https://doi.org/10.1186/s40468-023-00266-w>
- Nayernia, A., & Mohebbi, H. (2023). Teachers' online language assessment literacy. In Z. Tajeddin & C. Griffiths (Eds.), *Language Education Programs: Perspectives on Policies and Practices (Volume 34)* (pp. 121-140). Springer. https://doi.org/10.1007/978-3-031-38754-8_8
- Nayernia, A., Nosrati, R., & Mohebbi, H. (2022). The factors contributing to language teachers' effectiveness in an EFL learning context: A questionnaire validation study. *Profile: Issues in Teachers' Professional Development*, 24(2), 63–79. <https://doi.org/10.15446/profile.v24n2.93571>
- Nazari, M., & Molana, K. (2023). "Predators of emotions": The role of school assessment policies in English language teachers' emotion labor. *TESOL Quarterly*, 57(4), 1226–1255. <https://doi.org/10.1002/tesq.3188>
- Neittaamäki, R., & Lamprianou, I. (2024). All types of experience are equal, but some are more equal: The effect of different types of experience on rater severity and rater consistency. *Language Testing*, 41(3), 606–626. <https://doi.org/10.1177/02655322241239362>
- Neittaamäki, R., & Lamprianou, I. (2024). Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, 41(3), 584–605. <https://doi.org/10.1177/02655322241239363>
- Neumann, H., *Kozak, S., & *Gil, L. (2023). The impact of academic acculturation and language proficiency on international students' university experience and academic success: A longitudinal study. *Canadian Modern Language Review*, 79, 121-140. <https://doi.org/10.3138/cmlr-2021-0055>
- Newman, D. A., Tang, C., Song, Q. C., & Wee, S. (2022). Dropping the GRE, keeping the GRE, or GRE-optional admissions? Considering tradeoffs and fairness. *International Journal of Testing*, 22(1), 43-71. <https://doi.org/10.1080/15305058.2021.2019750>
- Nguyen L., Yuan Z., & Seed G. (2022). Building educational technologies for code-switching: Current practices, difficulties and future directions. *Languages*, 7, Article 220. <https://doi.org/10.3390/languages7030220>
- Nguyen, B. T. T., & Nguyen, L. Q. (2024). Measuring productive derivative knowledge of vietnamese EFL learners: The role of headword scoring, vocabulary breadth, and headword familiarity. *System*, 123, 103322. <https://doi.org/10.1016/j.system.2024.103322>
- Nguyen, B. T. T., Nguyen, T. T. B., Phan, Q. N. (2022). Vietnamese IELTS learners' perceptions of IELTS Writing Task 2 and their performance-based lexical resource. IELTS Research Reports Online Series, No. 5/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/nguyen-nguyen--phan-oct-2022.ashx>
- Nguyen, L. Q., & Le, H. V. (2022). Improving L2 learners' IELTS task 2 writing: The role of model essays and noticing hypothesis. *Language Testing in Asia*, 12, Article 58. <https://doi.org/10.1186/s40468-022-00206-0>
- Nguyen, L. Q., Nguyen, B. T. T., & Phuong, H. Y. (2024). Exploring the use of model texts as a feedback instrument in expository writing: EFL learners' noticing, incorporations, and text quality. *Assessing Writing*, 62, 100890. <https://doi.org/10.1016/j.asw.2024.100890>
- Nguyen, L., & Barrot, J. S. (2024). Detecting and assessing AI-generated and

- human-produced texts: The case of second language writing teachers. *Assessing Writing*, 62, 100899. <https://doi.org/10.1016/j.asw.2024.100899>
- Nguyen, T. H. H., Nguyen, B. T. T., Hoang, G. T. L., Pham, N. T. H., & Dang, T. T. C. (2024). Computer-delivered vs. face-to-face score comparability and test takers' perceptions: The case of the two English speaking proficiency tests for Vietnamese EFL learners. *Language Testing in Asia*, 14, Article 6. <https://doi.org/10.1186/s40468-024-00277-1>
- Nguyen, T. M. H., Gu, P., & Coxhead, A. (2024). Argument-based validation of Academic Collocation Tests. *Language Testing*, 41(3), 459–505. <https://doi.org/10.1177/02655322231198499>
- Nguyen, T. T., Richardson, T., Nguyen, A. N., Vu, T. N., & Dang, T. T. H. (2023). A systematic review of potential opportunities and challenges to the use of portfolios in Vietnam as an assessment tool. *Innovation in Language Learning and Teaching*, 17(5), 894–908. <https://doi.org/10.1080/17501229.2023.2194038>
- Nicholas, A., & Blake, J. (2024). Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing. *Research Methods in Applied Linguistics*, 3(3), Article 100164. <https://doi.org/10.1016/j.rmal.2024.100164>
- Nicklin, C., & Vitta, J. P. (2022). Assessing Rasch measurement estimation methods across R packages with yes/no vocabulary test data. *Language Testing*, 39(4), 513–540. <https://doi.org/10.1177/02655322211066822>
- Niessen, A. S. M., & Neumann, M. (2022). Using personal statements in college admissions: An investigation of gender bias and the effects of increased structure. *International Journal of Testing*, 22(1), 5-20. <https://doi.org/10.1080/15305058.2021.2019749>
- Nieto, D., & Nguyen, A. (2023). Strategizing assessment to support emerging bilingual students: a focus on the home language. *Multicultural Learning and Teaching*, 18(2), 239-250. <https://doi.org/10.1515/mlt-2021-0005>
- Niimi, N., & Matsuura, N. (2022). Assessing Japanese junior high school students' English achievement through computer-based testing in the classroom: A case of integrated reading-into-writing continuous task. *Language Testing in Asia*, 12, Article 38. <https://doi.org/10.1186/s40468-022-00189-y>
- Nikmard, F., & Tavassoli, K. (2021). A three-parameter logistic IRT calibration of the items of the TEFL MA admission test as a high-stakes test in Iran. *Iranian Journal of Applied Linguistics (IJAL)*, 24(2), 135-161. <http://ijal.khu.ac.ir/article-1-3173-en.html>
- Nikmard, F., & Tavassoli, K. (2023). The impact of test length on raters' mental processes during scoring test-takers' writing performance. *Journal of Language Horizons*, 7(1), 159-182. <https://doi.org/10.22051/lghor.2022.37340.1545>
- Nikmard, F., Tavassoli, K., & Pourdana, N. (2023). Designing and validating a scale for evaluating the sources of unreliability of a high-stakes test. *Language Testing in Asia*, 13, Article 2 (2023). <https://doi.org/10.1186/s40468-023-00215-7>
- Nikolov, M., & Timpe-Laughlin, V. (2021). Assessing young learners of foreign language abilities. *Language Teaching*, 54(1), 1–37. <https://doi.org/10.1017/S0261444820000294>
- Nishizawa, H. (2023). Construct validity and fairness of an operational listening test with World Englishes. *Language Testing*, 40(3), 493-520. <https://doi.org/10.1177/02655322221137869>
- Nishizawa, H. (2024). Authenticity of academic lecture passages in high-stakes tests: A

- temporal fluency perspective. *Language Testing*, 41(4), 792–816.
<https://doi.org/10.1177/02655322241262453>
- Nishizawa, H., Isbell, D. R., & Suzuki, Y. (2022). Review of the Japanese-language proficiency test. *Language Testing*, 39(3), 494–503.
<https://doi.org/10.1177/02655322221080898>
- Niu, R., Shan, P., & You, X. (2021). Complementation of multiple sources of feedback in EFL learners' writing. *Assessing Writing*, 49, Article 100549.
<https://doi.org/10.1016/j.asw.2021.100549>
- Noori, M. (2022). Road not taken" in language testing: Sociocultural implications of test and teaching contents. *TESOL Quarterly*, 56(4), 1486-1503.
<https://doi.org/10.1002/tesq.3189>
- Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of IELTS preparation materials from a critical perspective. *Language Assessment Quarterly*, 18(4), 382–397.
<https://doi.org/10.1080/15434303.2021.1883618>
- Noroozi, S., & Karami, H. (2022). A scrutiny of the relationship between cognitive load and difficulty estimates of language test items. *Language Testing in Asia*, 12, Article 13.
<https://doi.org/10.1186/s40468-022-00163-8>
- Noroozi, S., & Karami, H. (2024). A Rasch-based validation of the University of Tehran English Proficiency Test (UTEPT). *Language Testing in Asia*, 14, Article 18.
<https://doi.org/10.1186/s40468-024-00290-4>
- Norouzian, R. (2021). Interrater reliability in second language meta-analyses: The case of categorical moderators. *Studies in Second Language Acquisition*, 43(4), 896–915.
<https://doi.org/10.1017/S0272263121000061>
- Norris, J.M., Sasayama, S. and Kim, M. (2023), Simulating real-world context in an email writing task: Implications for task-based language assessment. *ETS Research Report Series*, 2023, 1-22. <https://doi.org/10.1002/ets2.12366>
- North, B., & Piccardo, E. (2023). Aligning Language Frameworks: An Example with the CLB and CEFR. *Language Assessment Quarterly*, 20(2), 143-165.
<https://doi.org/10.1080/15434303.2023.21842666>
- Nour, P., Esfandiari, R., & Zarei, A. A. (2021). Development and validation of a metamemory maturity questionnaire in the context of English as a foreign language. *Language Testing in Asia*, 11, Article 24.
<https://doi.org/10.1186/s40468-021-00141-6>
- Nowruzi, M. (2021). A study of EFL teachers' classroom grading practices in secondary schools and private institutes: A mixed methods approach. *Language Testing in Asia*, 11, Article 29. <https://doi.org/10.1186/s40468-021-00145-2>
- Nurjain, A., Salazar-Espinoza, D. E., Saenko, N., & Bina, E. (2023). Learner-oriented assessment matters: testing the effects of academic buoyancy, reflective thinking, and learner enjoyment in self-assessment and test-taking anxiety management of the EFL learners. *Language Testing in Asia*, 13, Article 30 (2023).
<https://doi.org/10.1186/s40468-023-00247-z>

O

- O'Grady, S. (2023). Adapting multiple-choice comprehension question formats in a test of second language listening comprehension. *Language Teaching Research*, 27(6),

- 1431-1455. <https://doi.org/10.1177/1362168820985367>
- O'Reilly, D., & Marsden, E. (2023). Elicited metaphoric competence in a second language: a construct associated with vocabulary knowledge and general proficiency?. *International Review of Applied Linguistics in Language Teaching*, 61(2), 287-327. <https://doi.org/10.1515/iral-2020-0054>
- O'Shannessy, C., Davis, V., Bartlett, J., Nelson, A., & Foster, D. (2023). Developing the Little Kids' Word List app, a fair assessment tool of communicative development for young Aboriginal children in multilingual families in Central Australia. *Studies in Language Assessment*, 12(2), 286-320.
- O'Sullivan, B. (2023). Integrating a Remote Proctoring Solution: the case of Aptis Remote. In H. Işık, E. Kantarcioğlu, A. Üner, & C. Thomas (Eds.), *Re-Vision and Re-Form in English for Academic Purposes Contexts after the Pandemic* (pp. 75-82). Cambridge Scholars.
- O'Sullivan, B. (2023). Language Testing. In L. Wei, Z. Hua, & J. Simpson (Eds.), *Routledge Handbook of Applied Linguistics (2nd ed., pp.136-149)*. Routledge. <https://doi.org/10.4324/9781003082644>
- O'Sullivan, B. (2023). Reflections on the application and validation of technology in language testing. *Language Assessment Quarterly*, 20(4-5), 501-511. <https://doi.org/10.1080/15434303.2023.2291486>
- O'Sullivan, B. (2024). Assessing Speaking. In Antony Kunnan (ed), *The Concise Companion to Language Assessment*, (pp.143-153). Wiley. DOI: 10.1002/9781118411360.wbcla084
- O'Sullivan, B. Saville, N. (2024). Foreword. In: Field, J. *Insights into Assessing Academic Listening: The Case of IELTS*. SiLT 53, Cambridge: Cambridge University Press and Assessment. <https://www.cambridge.org/cf/cambridgeenglish/teacher-development/insights-assessing-academic-listening-case-ielts/insights-assessing-academic-listening-case-ielts-paperback>
- O'Sullivan, B., & Chalhoub-Deville, M. (2021). Language testing for migrants: Co-constructing validation. *Language Assessment Quarterly*, 18(5), 547–557. <https://doi.org/10.1080/15434303.2021.1986513>
- O'Sullivan, B., & Chalhoub-Deville, M. (2023). Validity: An integrated arguments approach. In: Dobric, N., Cesnik, H., and Harsch, C (Eds). *Festschrift in honor of Günther Sigott: Advanced Methods in Language Testing*. Peter Lang.
- O'Sullivan, B., & Cheng, L. (2022). Lessons from the Chinese imperial examination system. *Language Testing in Asia*, 12, Article 52. <https://doi.org/10.1186/s40468-022-00201-5>
- O'Sullivan, B., Breakspear, T., & Bayliss, W. (2023). Validating an AI-Driven Scoring System: The Model Card Approach. In Karim Sadeghi & Dan Douglas (eds), *Fundamental Considerations in Technology Mediated Language Assessment*, (pp.115-134). New York: Routledge. DOI: 10.4324/9781003292395-10
- O'Sullivan, B., Dunn, K., & Berry, V. (2021). Test preparation: An international comparison of test takers' preferences. *Assessment in Education: Principles, Policy & Practice*, 28(1), 13–36. <https://doi.org/10.1080/0969594X.2019.1637820>
- Occhino, C., Lidster, R., Geer, L. C., Listman, J., & Hauser, P. C. (2024). Development of the American Sign Language Fingerspelling and Numbers Comprehension Test (ASL FaN-CT). *Language Testing*, 41(1), 135–161.

- <https://doi.org/10.1177/02655322231179494>
- Ockey, G. J., & Chukharev-Hudilainen, E. (2021). Human versus computer partner in the paired oral discussion test. *Applied Linguistics*, 42(5), 924–944. <https://doi.org/10.1093/applin/amaa067>
- Ockey, G. J., & Neiriz, R. (2021). Evaluating technology-mediated second language oral communication assessment delivery models. *Assessment in Education: Principles, Policy & Practice*, 28(4), 350–368. <https://doi.org/10.1080/0969594X.2021.1976106>
- Ockey, G. J., Chukharev-Hudilainen, E., & Hirsch, R. R. (2023). Assessing interactional competence: ICE versus a human partner. *Language Assessment Quarterly*, 20(4-5), 377–398. <https://doi.org/10.1080/15434303.2023.2237486>
- Ockey, G. J., Muhammad, A. A., Prasetyo, A. H., Elnegahy, S., Kochem, T., Neiriz, R., Kim, H., & Beck, J. (2021). Iowa State University's English placement test of oral communication in times of COVID-19. *Language Assessment Quarterly*, 18(1), 26–35. <https://doi.org/10.1080/15434303.2020.1862122>
- Ockey, G. J., Vo, S., & Baghestani, S. (2021). Establishing appropriate cut scores of a standardized test for a local placement context. *TESOL Quarterly*, 55(3), 719–742. <https://doi.org/10.1002/tesq.3007>
- Ogawa, C. (2022). CAF indices and human ratings of oral performances in an opinion-based monologue task. *Language Testing in Asia*, 12, Article 4. <https://doi.org/10.1186/s40468-022-00154-9>
- Oh, S. (2022). The use of spelling and reference tools in second language writing: Their impact on students' writing performance and process. *Journal of Second Language Writing*, 57, Article 100916. <https://doi.org/10.1016/j.jslw.2022.100916>
- Oliver, G. P., & Gregory, K. M. (2024). Integrated Performance Assessments: providing equitable instruction and assessment for ELLs/MLLs. *TESOL Quarterly*, 58(4), 1815–1830. <https://doi.org/10.1002/tesq.3324>
- Oliver, R., Chen, H., & Dovchin, S. (2024). Review of selected research in applied linguistics published in Australia (2015–2022). *Language Teaching*, 57(3), 341–376. doi:10.1017/S0261444824000077
- Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, 40(3), 521–547. <https://doi.org/10.1177/02655322221139162>
- Orabah, S. S. B., Bijani, H., & Ismail, S. M. (2022). Assessing English language teachers' understanding and practices of student-centered learning in Oman. *Language Testing in Asia*, 12, Article 41. <https://doi.org/10.1186/s40468-022-00184-3>
- Ortiz, M., and Díaz, C. (2022). *The effect of screencast feedback on the performance of the IELTS essay writing component among pre-service teachers of English: An intervention study in Chile*. IELTS Research Reports Online Series, No. 3/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/ortiz-and-diaz-2022-june.ashx>
- Osidak, V., Kvasova, O., Kavytska, T., & Drobotun, V. (2024). Book Creator as a tool of formative assessment in the TEFL classroom: Emergency remote teaching during wartime. *Studies in Language Assessment (SILA)*, 13(2), 1–37. <https://doi.org/10.58379/RBXT4093>
- Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51, Article 100603.

- <https://doi.org/10.1016/j.asw.2021.100603>
- Owen, N., Shrestha, P. N., & Hultgren, A. K. (2021). Researching academic reading in two contrasting English as a medium of instruction contexts at a university level. *ETS Research Report Series*, 2021(1), 1–28. <https://doi.org/10.1002/ets2.12317>
- Owen, N., Shrestha, P., & Bax, S. (2021). *Researching lexical thresholds and lexical profiles across the Common European Framework of References for Languages (CEFR) levels assessed in the Aptis test* (Report No. AR-G/2021/1). British Council. https://www.britishcouncil.org/sites/default/files/owen_shrestha_and_bax_layout.pdf
- Ożańska-Ponikwia, K., Piechurska-Kuciel, E., & Skalaćka, K. (2023). Emotional intelligence as a mediator in the relationship between neuroticism and L2 achievement. *Applied Linguistics Review*, 14(1), 67–86. <https://doi.org/10.1515/applirev-2020-0008>
- Ozer, O. (2024). AI language models: A breach of academic integrity in online language learning? *Studies in Language Assessment (SILA)*, 13(1), 237–262. <https://doi.org/10.58379/EZYS5680>

P

- Paek, I., Liang, X., & Lin, Z. (2021). Regarding item parameter invariance for the Rasch and the 2-parameter logistic models: An investigation under finite non-representative sample calibrations. *Measurement: Interdisciplinary Research and Perspectives*, 19(1), 39–54. <https://doi.org/10.1080/15366367.2020.1754703>
- Palmour, L. (2023). Assessing oral presentations: An analysis of score-reaching dialogue between EAP practitioners. *Journal of English for Academic Purposes*, 62, Article 101210. <https://doi.org/10.1016/j.jeap.2022.101210>
- Palmour, L. (2024). Assessing speaking through multimodal oral presentations: The case of construct underrepresentation in EAP contexts. *Language Testing*, 41(1), 9–34. <https://doi.org/10.1177/02655322231183077>
- Pan, J., & Marsden, E. (2024). Developing internet-based Tests of Aptitude for Language Learning (TALL): An open research endeavour. *Language Testing*, 41(4), 817–827. <https://doi.org/10.1177/02655322241241849>
- Pan, R., & Lu, X. (2023). The design and cognitive validity verification of reading-to-write tasks in L2 Chinese writing assessment. *Assessing Writing*, 56, Article 100699. <https://doi.org/10.1016/j.asw.2023.100699>
- Panahi, A., & Mohebbi, H. (2022). Cognitive diagnostic assessment of IELTS listening: Providing feedback from its internal structure. *Language Teaching Research Quarterly*, 29, 147–160. <https://doi.org/10.32038/ltrq.2022.29.10>
- Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly*, 39, 372–425. <https://doi.org/10.32038/ltrq.2024.39.23>
- Panahi, A., Mohebbi, H., & Oxford, R. L. (2024). Rebecca Oxford's 50 years of contributions to language education and related fields: A systematic review. *Language Teaching Research Quarterly*, 41, 169–230.
- Pang, F., & Skehan, P. (2021). Performance profiles on second language speaking tasks. *The Modern Language Journal*, 105(1), 371–390. <https://doi.org/10.1111/modl.12699>
- Papageorgiou, S. (2024). Can language test providers do more to support Open Science? A response to Winke. *Language Testing*, 41(4), 860–864.

- <https://doi.org/10.1177/02655322241232361>
- Papageorgiou, S., & Manna, V. F. (2021). Maintaining access to a large-scale test of academic language proficiency during the pandemic: The launch of TOEFL iBT Home Edition. *Language Assessment Quarterly*, 18(1), 36–41. <https://doi.org/10.1080/15434303.2020.1864376>
- Papageorgiou, S., Davis, L., Ohta, R., & Garcia Gomez, P. (2022). Mapping TOEFL® Essentials™ test scores to the Canadian Language Benchmarks. *ETS Research Report Series*, 2022(1), 1–42. <https://doi.org/10.1002/ets2.12357>
- Papi, M., Wolff, D., Nakatsukasa, K., & Bellwoar, E. (2021). Motivational factors underlying learner preferences for corrective feedback: Language mindsets and achievement goals. *Language Teaching Research*, 25(6), 858–877. <https://doi.org/10.1177/13621688211018808>
- Paquot, M., Rubin, R., & Vandeweerd, N. (2022). Crowdsourced adaptive comparative judgment: A community-based solution for proficiency rating. *Language Learning*, 72, 853–885. <https://doi.org/10.1111/lang.12498>
- Park, H.I., Solon, M., Dehghan-Chaleshtori, M., & Ghanbar, H. (2022). Proficiency reporting practices in research on second language acquisition: Have we made any progress?. *Language Learning*, 72, 198–236. <https://doi.org/10.1111/lang.12475>
- Park, Y., Lee, S., & Shin, S.-Y. (2022). Developing a local academic English listening test using authentic unscripted audio-visual texts. *Language Testing*, 39(3), 401–424. <https://doi.org/10.1177/02655322221076024>
- Parker, S. (2023). Building connection: An EAP course that facilitates engagement with the university community. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 15–42). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Parkinson, J., Watterson, G., Sawczuc da Silva, A., & Kithulgoda, E. (2023). Rhetorical features of high-graded and low-graded engineering case study responses. *Journal of English for Academic Purposes*, 61, Article 101206. <https://doi.org/10.1016/j.jeap.2022.101206>
- Patekar, J. (2021). A look into the practices and challenges of assessing young EFL learners' writing in Croatia. *Language Testing*, 38(3), 456–479. <https://doi.org/10.1177/0265532221990657>
- Patra, I., Alazemi, A., Al-Jamal, D., & Gheisari, A. (2022). The effectiveness of teachers' written and verbal corrective feedback (CF) during formative assessment (FA) on male language learners' academic anxiety (AA), academic performance (AP), and attitude toward learning (ATL). *Language Testing in Asia*, 12, Article 19. <https://doi.org/10.1186/s40468-022-00169-2>
- Pearson, W. (2021). The sale of IELTS testing in India: Motivations and implications. *ELT Journal*, 75(4), 515–517. <https://doi.org/10.1093/elt/ccab053>
- Pearson, W. S. (2022). Test Review: LanguageCert IESOL B1 (Achiever) SELT. *Language Assessment Quarterly*, 19(5), 526–537. <https://doi.org/10.1080/15434303.2022.2103420>
- Pearson, W. S. (2022). The mediating effects of student beliefs on engagement with written feedback in preparation for high-stakes English writing assessment. *Assessing Writing*, 52, Article 100611. <https://doi.org/10.1016/j.asw.2022.100611>
- Pelleriti, M. (2023). Item writing and dyslexic students: Promoting best practices. *Iperstoria*,

21. 217–231. <https://iperstoria.it/article/view/1285/1315>
- Peng, C. X. (2024). Beyond accuracy gains: Investigating the impact of individual and collaborative feedback processing on L2 writing development. *Assessing Writing*, 61, 100876. <https://doi.org/10.1016/j.asw.2024.100876>
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang H. (2023). Predicting Chinese EFL Learners' Human-rated Writing Quality in Argumentative Writing Through Multidimensional Computational Indices of Lexical Complexity. *Assessing Writing*, 56, Article 100722. <https://doi.org/10.1016/j.asw.2023.100722>
- Peng, Y., Yan, W., & Cheng, L. (2021). Hanyu Shuiping Kaoshi (HSK): A multi-level, multi-purpose proficiency test. *Language Testing*, 38(2), 326–337. <https://doi.org/10.1177/0265532220957298>
- Peng, Y., Zheng, Y., Sun, J., Jiang, Y., Lin, J., & Zhang, H. (2024). Modeling relationships among large-grained, fine-grained absolute syntactic complexity and assessed L2 writing quality: An SEM approach. *Assessing Writing*, 61, 100875. <https://doi.org/10.1016/j.asw.2024.100875>
- Pérez Castillejo, S. (2023). Prior processing, foreign language classroom anxiety, and L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 61(2), 519-544. <https://doi.org/10.1515/iral-2021-0091>
- Perry, S. (2024). Podcasting as a form of assessment: Increasing student motivation in academic English-speaking assessment. *Language Testing in Asia*, 14, Article 15. <https://doi.org/10.1186/s40468-024-00288-y>
- Pesau, H. G., Immanuel, A. S., Sulastri, A., & van Lijstelaar, G. (2023). The role of daily spoken language on the performance of language tests: The Indonesian experience. *Bilingualism: Language and Cognition*, 26(3), 538-549. <https://doi.org/10.1017/S136672892200075X>
- Peterson, S. S., Altidor, A., & Kerwood, J. (2021). Young children's written and verbal responses in a dynamic assessment context. *Assessing Writing*, 49, Article 100543. <https://doi.org/10.1016/j.asw.2021.100543>
- Phakiti, A. (2024). Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study. IELTS Research Reports Online Series, No. 1/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at <https://ielts.org/researchers/our-research/research-reports>
- Phakiti, A., & Isaacs, T. (2021). Classroom assessment and validity: Psychometric and edumetric approaches. *The European Journal of Applied Linguistics and TEFL*, 10(1), 3–21.
- Phuoc, V. D., & Barrot, J. S. (2022). Complexity, accuracy, and fluency in L2 writing across proficiency levels: A matter of L1 background? *Assessing Writing*, 54, Article 100673. <https://doi.org/10.1016/j.asw.2022.100673>
- Phuong, H. Y., Phan, Q. T. & Le, T. T. (2023). The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: a Vietnamese contextual study. *Language Testing in Asia*, 13, Article 42 (2023). <https://doi.org/10.1186/s40468-023-00256-y>
- Pill J., & Smart, C. (2021). Raters: Behavior and training. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 135–144). Routledge. <https://doi.org/10.4324/9781351034784-15>
- Pill, J., & Harding, L. (2019). A most engaging scholar: Tim McNamara and the role of language testing expertise. In C. Roever & G. Wigglesworth (Eds.), *Social*

- perspectives on language testing: Papers in honour of Tim McNamara (pp. 217–228). Peter Lang. <https://doi.org/10.3726/b15199>
- Pill, J., Bandman, J., Bottini, R., Brunfaut, T., Davidson, N., Dávila Pérez, G., Harding, L., Jung, Y., Lestari, S. B., Ramos Gálvez, C., & Rossi, O. (2024). Shaping a language testing curriculum: Insights from an oral history of a Master's programme. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence. Volume 1: Research and reflections from the field* (pp. 168–191). Cambridge University Press. <https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- Piyumi Udesinee, W. A., Knutsson, O., Barbutiu, S. M., & Jayathilake, C. (2024). Re-designing a regulatory scale for dynamic assessment in the synchronous text chat environment in collaboration with teachers. *Computer Assisted Language Learning*, 37(7), 1527-1553. <https://doi.org/10.1080/09588221.2022.2092153>
- Pinar, Y., Ünal, F., & Pinar, N. K. (2021). Language proficiency analysis of Turkish children from high-income and highly educated families aged 5 in Antalya. *Language Testing in Asia*, 11, Article 28. <https://doi.org/10.1186/s40468-021-00138-1>
- Poehner, M. E., & Inbar-Lourie, O. (2024). Indeterminacy, disruption, and (more) intelligent approaches to classroom language assessment: Learning from the COVID-19 pandemic. Introduction to the special issue on rethinking classroom language assessment in the wake of the COVID-19 pandemic. *Studies in Language Assessment Journal*, 13(2), vi–xx.
- Poehner, M. E., & Lantolf, J. P. (2023). Advancing L2 dynamic assessment: Innovations in Chinese contexts. *Language Assessment Quarterly*, 20(1), 1-19. <https://doi.org/10.1080/15434303.2022.2158465>
- Poehner, M. E., & Yu, L. (2022). Dynamic assessment of L2 writing: Exploring the potential of rubrics as mediation in diagnosing learner emerging abilities. *TESOL Quarterly*, 56(4), 1191-1217. <https://doi.org/10.1002/tesq.3098>
- Poetsch, S. (2023). Unguided. Assessing young EAL/D learners' achievements in mainstream curriculum areas. *Studies in Language Assessment*, 12(2), 28-58.
- Pomarolli, G., Artoni, D., Boschiero, M., & Piccinin, S. (2023). Inclusive and accessible teaching of Russian L2 to absolute beginners: A comparison between TBLT and PPP. *Instructed Second Language Acquisition*, 7(2), 166-193. <https://doi.org/10.1558/isla.26047>
- Pourdana, N., & Asghari, S. (2021). Different dimensions of teacher and peer assessment of EFL learners' writing: Descriptive and narrative genres in focus. *Language Testing in Asia*, 11, Article 6. <https://doi.org/10.1186/s40468-021-00122-9>
- Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, 12, Article 7. <https://doi.org/10.1186/s40468-022-00156-7>
- Prabowo, M. Y. (2022). Assessing equivalency between paper-and-pencil and computer-based English competency test. In R. E. Capel & Q. S. Hoon (Eds.), *Conference proceedings iELT-CON2021* (pp. 42–51). Penang English Language Learning and Teaching Association. https://pellta.org/uploads/8/1/8/9/81899300/ielt-con_2021_proceedings.pdf
- Prabowo, M. Y., & Rahmadian, S. (2022). Computer-based English competency assessment for scholarship selection: Challenges, strategies, and implementation in the Ministry of Finance. *Jurnal Sositoknologi*, 21(1), 84–96. <https://doi.org/10.5614/sostek.itbj.2022.21.1.9>

- Prasad, K., Aladini, A., Normurodova, N. Z., & Belton, B. (2023). Take language assessment easy: the mediator impacts of self-assessment, test-taking skills in predicting student evaluation apprehension, foreign language learning self-esteem, and language achievement in online classes. *Language Testing in Asia*, 13, Article 34 (2023). <https://doi.org/10.1186/s40468-023-00246-0>
- Prichard, C., & Atkins, A. (2021). Evaluating the vocabulary coping strategies of L2 readers: An eye tracking study. *TESOL Quarterly*, 55(2), 593–620. <https://doi.org/10.1002/tesq.3005>
- Pun, J., & Li, W. K. (2024). A structural equation investigation of linguistic features as indices of writing quality in assessed secondary-level EMI learners' scientific reports. *Assessing Writing*, 62, 100897. <https://doi.org/10.1016/j.asw.2024.100897>
- Pun, J., Tan, S., & Li, X. (2024). Discourse competence in Hong Kong secondary students' disciplinary research writing. *Assessing Writing*, 61, 100872. <https://doi.org/10.1016/j.asw.2024.100872>
- Purpura, J. E. (2024). An Interview with Timothy Francis McNamara. *Language Assessment Quarterly*, 21(2). 123-124.
- Purpura, J. E. (2024). In Memoriam: Timothy Francis McNamara. *Language Assessment Quarterly*. 21(2). 125-140.
- Purpura, J. E. (2024). Learning-oriented assessment. In A. J. Kunnan (Ed.), *Companion to Language Assessment* (pp. 100-124). Oxford, UK: John Wiley & Sons.
- Purpura, J. E., & Kunnan, A. J. (Eds.) (2024). *The writings of L. F. Bachman: "Assuring that what we count counts."* New York: Routledge.
- Purpura, J. E., & Oh, S. (2024). Assessing the linguistic resources of meaningful communication. In A. J. Kunnan (Ed.), *Companion to Language Assessment* (pp. 100-124). Oxford, UK: John Wiley & Sons.
- Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of an online language placement exam. *Language Assessment Quarterly*, 18(1), 42–50. <https://doi.org/10.1080/15434303.2020.1867145>
- Purpura, J.E (2021). A rationale for using a scenario-based assessment to measure competency-based, situated second and foreign language proficiency. In M. Masperi, C Cervini, Y Bardière (Eds.), *Évaluation des acquisitions langagières: du formatif au certificatif* (pp. 54-96). MediAzioni. <http://www.mediazioni.sitlec.unibo.it>.
- Pusey, K., & Butler, Y. G. (2023). Investigating the ecological validity of second language writing assessment tasks. *System*, 119, Article 103174. <https://doi.org/10.1016/j.system.2023.103174>
- Pusey, K., & Butler, Y. G. (2024). Amplifying test-taker voices in the validation of L2 writing assessment tasks. *Assessing Writing*, 59, 100790. <https://doi.org/10.1016/j.asw.2023.100790>

Q

- Qayyum, R. (2024). A teacher's inquiry into diagnostic assessment in an EAP writing course. *Assessing Writing*, 61, 100848. <https://doi.org/10.1016/j.asw.2024.100848>
- Qian, L. (2023). Use of lexical features in high-stakes tests: Evidence from the perspectives of complexity, accuracy and fluency. *Assessing Writing*, 57, Article 100758. <https://doi.org/10.1016/j.asw.2023.100758>
- Qin, C. Y., & Baker, B. (2025). Review of the Canadian English Language Proficiency Index

- Program (CELP). *Language Testing*, 42(1), 100-113.
<https://doi.org/10.1177/02655322241291764>
- Qin, L., & Nian, Y. (2024). Unpacking Chinese lower-proficiency EFL learners' conceptual development of English subjunctives with group dynamic assessment: Efficacy, sustainability and reciprocity. *System*, 121, 103248.
<https://doi.org/10.1016/j.system.2024.103248>
- Qin, T., & van Compernelle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, 25(2), 55–74.
<http://hdl.handle.net/10125/73433>
- Qiu, Y., & Aryadoust, V. (2024). The predictive value of gaze behavior and mouse-clicking in testing listening proficiency: A sensor technology study. *System*, 126, 103440.
<https://doi.org/10.1016/j.system.2024.103440>
- Quaid, E., & Barrett, A. (2021). Interpretations of spoken utterance fluency in simulated and face-to-face oral proficiency interviews. *Language Education & Assessment*, 4(1), 1-18. <https://doi.org/10.29140/lea.v4n1.385>
- QUEVEDO-CAMARGO, G., & SCARAMUCCI, M. V. R. (2016) Methodology in washback studies. In: GITSAKI, Christine; COOMBE, Christine. (Eds.). *Current Trends in Language Evaluation, Assessment & Testing: Research Perspectives*. 1ed. Cambridge: Cambridge Scholars Publishing, v. 1, p. 287-304.
- QUEVEDO-CAMARGO, G., TONELLI, J. A. R., FURTOSO, V. A. B. (2024). Avaliação e formação de professores de línguas: contextos, necessidades, desafios e avanços [Assessment and language teacher education: needs, challenges and advances]. Special volume. *REVISTA X*, v. 19, p. 217-221.
<https://revistas.ufpr.br/revistax/issue/view/3629>
- QUEVEDO-CAMARGO, G. (2020) Formação de professores de línguas adicionais e letramento em avaliação: breve panorama e desafios para os cursos de licenciatura em LEM no Brasil [Additional language teacher education and assessment literacy: a brief overview and challenges for undergraduate courses in Brazil]. *Calidoscópio (UNISINOS)*, v. 18, p. 435-459. DOI: <https://doi.org/10.4013/cld.2020.182.10>
- QUEVEDO-CAMARGO, G. (2021) Avaliação online: um guia para professores. [Online assessment: a guide for teachers] Letraria. E-book, 63p. Available at: <https://www.letraria.net/avaliacao-online-um-guia-para-professores/>
- QUEVEDO-CAMARGO, G. & SANTOS, Gabriela M. (2020) Validade de conteúdo das provas de um curso de compreensão oral para fins acadêmicos: relato de uma experiência [Content validity of tests in a listening comprehension course for academic purposes: report of an experience]. *Revista Olhares e Trilhas*, v. 22, p. 289-308. DOI: <https://doi.org/10.14393/OT2020v22.n.2.55520>
- QUEVEDO-CAMARGO, G. & SCARAMUCCI, M. V. R. (2018) O conceito de letramento em avaliação de línguas: origem de relevância para o contexto brasileiro [The concept of language assessment literacy: origin and relevance for the Brazilian context]. *Linguagem: estudos e pesquisas*, v. 22, p. 225-245. DOI: <https://doi.org/10.5216/lep.v22i1.54474>
- QUEVEDO-CAMARGO, G., & PINHEIRO, L. A. (2021) Ética na avaliação de línguas adicionais: da postura docente ao instrumento avaliativo [Ethics in the assessment of additional languages: from the teaching stance to the assessment instrument]. *Estudos em Avaliação Educacional (Online)*, v. 32. DOI: <https://doi.org/10.18222/eae.v32.8641>

- QUEVEDO-CAMARGO, G., & SOUSA, A. A. C. P. (2022). Letramento em avaliação para professores de inglês e a perspectiva glocal: um microestudo [Language assessment literacy for language teachers and the glocal perspective: a micro-study]. *Revista Do GEL*, 19(1), 223–245. <https://doi.org/10.21165/gel.v19i1.3371>
- QUEVEDO-CAMARGO, G., & TONELLI, J. A. R. (2021) Contribuições de um curso EAD para o letramento em avaliação de professores de línguas adicionais para crianças [Contributions of a distance-mode course on assessment literacy for teachers of additional languages to young learners]. *Revista do GEL*, v. 18, p. 230-248. DOI: <https://doi.org/10.21165/gel.v18i1.3094>
- QUEVEDO-CAMARGO, G., & TONELLI, J. R. A. (2024). Inglês e avaliação no ensino fundamental I: Propostas para o PNE 2025-2035 [English and assessment in primary school I: proposals for the National Plan of Education 2025-2035]. *Estudos Em Avaliação Educacional*, 35, e10547. <https://doi.org/10.18222/eae.v35.10547>

R

- Rad, H. S., & Alipour, R. (2024). Unlocking writing success: Building assessment literacy for students and teachers through effective interventions. *Assessing Writing*, 59, 100804. <https://doi.org/10.1016/j.asw.2023.100804>
- Rafatbakhsh, E., Ahmadi, A., Moloodi, A., & Mehrpour, S. (2021). Development and validation of an automatic item generation system for English idioms. *Educational Measurement: Issues and Practice*, 40(2), 49–59. <https://doi.org/10.1111/emip.12401>
- Rafi, A. (2024). Creativity, criticality and translanguaging in assessment design: perspectives from Bangladeshi higher education. *Applied Linguistics Review*, 15(5), 1887-1909. <https://doi.org/10.1515/applirev-2023-0086>
- Rafi, F., & Pourdana, N. (2023). E-diagnostic assessment of collaborative and individual oral tiered task performance in differentiated second language instruction framework. *Language Testing in Asia*, 13, Article 6 (2023). <https://doi.org/10.1186/s40468-023-00223-7>
- Rahimi, M. (2021). A comparative study of the impact of focused vs. comprehensive corrective feedback and revision on ESL learners' writing accuracy and quality. *Language Teaching Research*, 25(5), 687–710. <https://doi.org/10.1177/1362168819879182>
- Rahimi, M. (2024). Effects of integrating motivational instructional strategies into a process-genre writing instructional approach on students' engagement and argumentative writing. *System*, 121, 103261. <https://doi.org/10.1016/j.system.2024.103261>
- Rahman, K. A., Hasan, M. K., Namaziandost, E., & Ibna Seraj, P. M. (2021). Implementing a formative assessment model at the secondary schools: Attitudes and challenges. *Language Testing in Asia*, 11, Article 18. <https://doi.org/10.1186/s40468-021-00136-3>
- Rahman, K. A., Ibna Seraj, P.M., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11, Article 12. <https://doi.org/10.1186/s40468-021-00129-2>

- Rahmati, T., & Nushi, M. (2023). EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography. *Iranian Journal of Language Teaching Research*, 11(3), 31-52. <https://doi.org/10.30466/ijltr.2023.121405>
- Rahmati, T., & Sadeghi, K. (2024). Technological innovations used in researching language assessment: A scoping review. In K. Sadeghi (Ed.), *Routledge handbook of technological advances in researching language learning* (pp. 118-129). <https://doi.org/10.4324/9781003459088-11>
- Rahmati, T., Kazemian, M., Khonamri, F., & Daryaback, A.R. (2023). A chaos complexity perspective on EFL learners' sensitivity dependence on initial conditions in vocabulary tests, *The Journal of Asia TEFL*, 20(4), 895-903. <http://dx.doi.org/10.18823/asiatefl.2023.20.4.9.895>
- Rahmati, T., Kazemian, M., Khonamri, F., & Daryabak, A. (2023). A chaos complexity perspective on EFL learners' sensitivity dependence on initial conditions in vocabulary tests. *The Journal of Asia TEFL*, 20(4), 895-903. <http://dx.doi.org/10.18823/asiatefl.2023.20.4.9.895>
- Rambiritch, A., Alston, L. & Graham, M. (2021). Diagnosing with care: The academic literacy needs of theology students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 170-195). Multilingual Matters. <https://doi.org/10.21832/9781788926218-012>
- Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, 55(1), 105–133. <https://doi.org/10.1002/tesq.579>
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, Article 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Randall, J., Poe, M., Slomp, D., & Oliveri, M. E. (2024). Our validity looks like justice. Does yours? *Language Testing*, 41(1), 203–219. <https://doi.org/10.1177/02655322231202947>
- Randall, T. S., & Urbanski, K. (2023). Development of a Computerized Dynamic Assessment Program for Second Language Grammar Instruction and Assessment. *Language and Sociocultural Theory*, 10(1), 50-81. <https://doi.org/10.1558/lst.21006>
- Randez, R. A., & Cornell, C. (2023). Advancing equity in language assessment for learners with disabilities. *Language Testing*, 40(4), 984-999. <https://doi.org/10.1177/02655322231169442>
- Ranger, J., Kuhn, J., & Pohl, S. (2021). Effects of motivation on the accuracy and speed of responding in tests: The speed-accuracy tradeoff revisited. *Measurement: Interdisciplinary Research and Perspectives*, 19(1), 15–38. <https://doi.org/10.1080/15366367.2020.1750934>
- Rasooli, A., DeLuca, C. & Cheng, L. (2022). Beginning teacher candidates' approaches to grading and assessment conceptions—Implications for teacher education in assessment. *Educational Research for Policy and Practice*, 22, 63–90. <https://doi.org/10.1007/s10671-022-09320-5>
- Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2024). Classroom Assessment Fairness Inventory: A New Instrument to Support Fairness in Classroom Assessment. *Assessment in Education*, 30-5-6. <https://doi.org/10.1080/0969594X.2023.2255936>
- Rassaei, E. (2023). Implementing mobile-mediated dynamic assessment for teaching

- request forms to EFL learners. *Computer Assisted Language Learning*, 36(3), 257–287. <https://doi.org/10.1080/09588221.2021.1912105>
- Rastgou, A. (2022). How feedback conditions broaden or constrain knowledge and perceptions about improvement in L2 writing: A 12-week exploratory study. *Assessing Writing*, 53, Article 100633. <https://doi.org/10.1016/j.asw.2022.100633>
- Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education: Principles, Policy & Practice*, 29(3), 383-394. <https://doi.org/10.1080/0969594X.2022.2103516>
- Raykov, T., & Marcoulides, G.A. (2023) Evaluating the discrepancy between scale reliability and Cronbach's coefficient alpha using latent variable modeling, *Measurement: Interdisciplinary Research and Perspectives*, 21(1), 29-37, DOI: 10.1080/15366367.2022.2031485
- Razavipour, K. (2023). Classroom writing assessment and feedback practices: A new materialist encounter. *Assessing Writing*, 57, Article 100760. <https://doi.org/10.1016/j.asw.2023.100760>
- Razavipour, K., & Raji, B. (2022). Reliability of measuring constructs in applied linguistics research: a comparative study of domestic and international graduate theses. *Language Testing in Asia*, 12, Article 16. <https://doi.org/10.1186/s40468-022-00166-5>
- Read, J. (2022). Test Review: The International English Language Testing System (IELTS). *Language Testing*, 39(4), 679–694. <https://doi.org/10.1177/02655322221086211>
- Read, J. (2023). Towards a new sophistication in vocabulary assessment. *Language Testing*, 40(1), 40-46. <https://doi.org/10.1177/02655322221125698>
- Reed, D. J. and Stansfield, C. W. (2023). The MLAT-Elementary: History, adaptations and current uses. In E. Wen, P. Skehan, & R. Sparks (Eds.), *Language aptitude: Theory and practice* (pp. 25-46). Cambridge University Press & Assessment
- Reed, D. K., Ma, J., & Gerde, H. K. (2023). Resiliency and vulnerability in early grades writing performance during the COVID-19 pandemic. *Assessing Writing*, 57, Article 100741. <https://doi.org/10.1016/j.asw.2023.100741>
- Reifman, J. B. (2024). Reading the reader through raciolinguistic ideologies: An investigation of the evidence students present in self-placement. *Assessing Writing*, 59, 100792. <https://doi.org/10.1016/j.asw.2023.100792>
- Révész, A., & Brunfaut, T. (2021). Validating assessments for research purposes. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 21–32). Routledge. <https://doi.org/10.4324/9781351034784-4>
- Révész, A., Bunting, L., Florea, A., Gilabert, R., Hård af Segerstad, Y., Miha, I. P., & Vasalou, A. (2021). The effects of multiple-exposure textual enhancement on child L2 learners' development in derivational morphology: A multi-site study. *TESOL Quarterly*, 55(3), 901–930. <https://doi.org/10.1002/tesq.3036>
- Revesz, A., Michel, M., Lu, X., Kourtali, Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, 58, Article 100927. <https://doi.org/10.1016/j.jslw.2022.100927>
- Reyes, N. Á., Carrasco, D., Escribano, R., Espinosa, M. J., Figueroa, J., & Castillo, C. (2024). Effects of a genre and topic knowledge activation device on a standardized writing test performance. *Assessing Writing*, 62, 100898.

- <https://doi.org/10.1016/j.asw.2024.100898>
- Rezadoust Siah Khaleh Sar, H., Khalili Sabet, M., Mahdavi Zafarghandi, A., & Hassaskhah, J. (2024). Iranian teachers' and learners' attitudes towards the use of e-portfolio in second language speaking assessment. *Language Testing in Asia*, 14, Article 61. <https://doi.org/10.1186/s40468-024-00335-8>
- Rezaei, A., & Barkaoui, K. (2021). Peer and teacher assessment of L2 writing in high- and low-stakes conditions. *International Journal of Applied Linguistics*, 172(2), 199–228. <https://doi.org/10.1075/itl.20006.rez>
- Rezagah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Language Testing Journal*, 12(2), 76-94, <https://doi.org/10.22034/ijlt.2022.157126>
- Rezai, A. (2022). Fairness in classroom assessment: development and validation of a questionnaire. *Language Testing in Asia*, 12, Article 17. <https://doi.org/10.1186/s40468-022-00162-9>
- Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: Hearing from Iranian university teachers. *Language Testing in Asia*, 11, Article 26. <https://doi.org/10.1186/s40468-021-00142-5>
- Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: insights from Iranian university teachers. *Language Testing in Asia*, 12, Article 8. <https://doi.org/10.1186/s40468-022-00157-6>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2021). Linguistic complexity in teachers' assessment of German essays in high stakes testing. *Assessing Writing*, 50, Article 100561. <https://doi.org/10.1016/j.asw.2021.100561>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2024). The interplay of task characteristics, linguistic complexity, and language proficiency in High-Stakes English as a Foreign language writing. *TESOL Quarterly*, 58(2), 775–801. <https://doi.org/10.1002/tesq.3254>
- Rios, J. A., & Ihlenfeldt, S. D. (2021). State assessment score reporting practices for English learner parents. *Educational Measurement: Issues and Practice*, 40(3), 31–41. <https://doi.org/10.1111/emip.12424>
- Riswanto, Teferi, H., & Ahmed Abdel-Al Ibrahim, K. (2023). Cultivating EFL learners' productive skills by employing dynamic and non-dynamic assessments: Attitude in focus. *Language Testing in Asia*, 13, Article 18 (2023). <https://doi.org/10.1186/s40468-023-00228-2>
- Ritonga, M., Farhangi, F., Ajanil, B., & Farid Khafaga, A. (2022). Interventionist vs. interactionist models of dynamic assessment (DA) in the EFL classroom: impacts on speaking accuracy and fluency (SAF), foreign language classroom anxiety (FLCA), and foreign language learning motivation (FLLM). *Language Testing in Asia*, 12, Article 43. <https://doi.org/10.1186/s40468-022-00195-0>
- Ritonga, M., Shaban, A. A., Al-Rashidi, A. H., & Chilani, N. (2023). Engagement in on-line language assessment: Are test-taking skills, self-assessment, resilience, and autonomy critical?. *Language Testing in Asia*, 13, Article 25 (2023). <https://doi.org/10.1186/s40468-023-00236-2>
- Ritonga, M., Tazik, K., Omar, A., & Saberi Dehkordi, E. (2022). Assessment and language improvement: the effect of peer assessment (PA) on reading comprehension, reading motivation, and vocabulary learning among EFL learners. *Language Testing in Asia*, 12, Article 36. <https://doi.org/10.1186/s40468-022-00188-z>

- Robles-García, P., McLean, S., Stewart, J., Shin, J. young, & Sánchez-Gutiérrez, C. H. (2024). The Development and Initial Validation of O-WSVLT, a Meaning-Recall Online L2 Spanish Vocabulary Levels Test. *Language Assessment Quarterly*, 21(2), 181–205. <https://doi.org/10.1080/15434303.2024.2311724>
- Rock, K. (2022). Constructing a data-based analytic rubric for an academic blog post. *Assessing Writing*, 51, Article 100602. <https://doi.org/10.1016/j.asw.2021.100602>
- Roever, C., & Ikeda, N. (2022). What scores from monologic speaking tests can(not) tell us about interactional competence. *Language Testing*, 39(1), 7–29. <https://doi.org/10.1177/02655322211003332>
- Roever, C., & Ikeda, N. (2024). The relationship between L2 interactional competence and proficiency. *Applied Linguistics*, 45(4), 676-698. <https://doi.org/10.1093/applin/amad053>
- Rolstad, K., & MacSwan, J. (2024). Bilingual language assessment: A persistent case of bias. In A. K. Kibler, A. Walqui, G. C. Bunch, & C. J. Faltis (Eds.), *Equity in multilingual schools and communities: Celebrating the contributions of Guadalupe Valdés* (pp. 54-66). Multilingual Matters. <https://doi.org/10.2307/jj.22452326.9>
- Roohr, K., Olivera-Aguilar, M., Bochenek, J., & Belur, V. (2022), Exploring GRE® and TOEFL® Score Profiles of International Students Intending to Pursue a Graduate Degree in the United States. ETS Research Report Series, 2022(1), 1–27. <https://doi.org/10.1002/ets2.12343>
- Roschmann, S., Witmer, S. E., & Volker, M. A. (2021). Examining provision and sufficiency of testing accommodations for English learners. *International Journal of Testing*, 21(1), 32–55. <https://doi.org/10.1080/15305058.2021.1884872>
- Roslan, R., Nishio, Y., & Jawawi, R. (2022). Analyzing English language teacher candidates' assessment literacy: A case of Bruneian and Japanese universities. *System*, 111, Article 102934. <https://doi.org/10.1016/j.system.2022.102934>
- Rossi, O., & Brunfaut, T. (2021). Text authenticity in listening assessment: Can item writers be trained to produce authentic-sounding texts? *Language Assessment Quarterly*, 18(4), 398–418. <https://doi.org/10.1080/15434303.2021.1895162>
- Rousse-Malpat, A., Koote, L., Steinkrauss, R., & Verspoor, M. (2024). Parlez-vous français? Effects of structure-based versus dynamic-usage-based approaches on oral proficiency. *Language Teaching Research*, 28(5), 1893-1917. <https://doi.org/10.1177/13621688211040298>
- Ruegg, R., & Prommas, P. (2024). Comparing automated analysis and human analysis of hedging in academic texts. *Studies in Language Assessment (SILA)*, 13(1), 263-287. <https://doi.org/10.58379/SLVL8621>
- Ruegg, R., Hoang, H., & Petersen, N. (2024). The influence of English language proficiency test scores on the academic success of ESL undergraduate students. *Educational Research and Evaluation*, 29(3-4), 111-129. <https://doi.org/10.1080/13803611.2024.2314533>
- Rukthong, A. (2021). MC listening questions vs. integrated listening-to-summarize tasks: What listening abilities do they assess? *System*, 97, Article 102439. <https://doi.org/10.1016/j.system.2020.102439>
- Rüsseler, J., Arendt, D., Münte, T. F., Mohammadi, B., & Boltzmann, M. (2021). Literacy affects brain structure—What can we learn for language assessment in low literates? *Language Assessment Quarterly*, 18(5), 492–507. <https://doi.org/10.1080/15434303.2021.1931231>

- Russell, B., Barron, C., Kim, H., & Jang, E. E. (2022). A mixed-method investigation into international university students' experience with academic language demands. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2022.934692>
- Ryan, M., Khosronejad, M., Barton, G., Myhill, D., & Kervin, L. (2022). Reflexive writing dialogues: Elementary students' perceptions and performances as writers during classroom experiences. *Assessing Writing*, 51, Article 100592. <https://doi.org/10.1016/j.asw.2021.100592>
- Ryu, J., Ae Kim, Y., Eum, S., Park, S., Chun S., & Yang, S. (2022). The assessment of memes as digital multimodal composition in L2 classrooms. *Journal of Second Language Writing*, 57, Article 100914. <https://doi.org/10.1016/j.jslw.2022.100914>

S

- Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes*, 61, Article 101208. <https://doi.org/10.1016/j.jeap.2022.101208>
- Sadeghi, K. (Ed.). (2022). *Technology-assisted language assessment in diverse contexts: Lessons from the transition to online testing during COVID-19*. Routledge. <https://doi.org/10.4324/9781003221463>
- Sadeghi, K. & Douglas, D. (Eds.). (2023). *Fundamental considerations in technology mediated language assessment*. Routledge. <https://doi.org/10.4324/9781003292395>
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50–61. <https://doi.org/10.1016/j.asw.2017.09.003>
- Sadoughi, M., & Hejazi, S. Y. (2022). Development and validation of Foreign Language Teacher Support Scale (FLTSS). *Language Testing in Asia*, 12, Article 30. <https://doi.org/10.1186/s40468-022-00186-1>
- Saeed, M. A., & Alharbi, M. A. (2023). Towards fostering Saudi EFL learners' collaborative engagement and feedback literacy in writing. *Assessing Writing*, 56, Article 100721. <https://doi.org/10.1016/j.asw.2023.100721>
- Saeed, M. A., AbuSa'aleek, A., & Alharbi, M. A. (2024). Examining teacher's evaluative language in written, audio and screencast feedback on EFL learners' writing from the appraisal framework: A linguistic perspective. *Assessing Writing*, 61, 100871. <https://doi.org/10.1016/j.asw.2024.100871>
- Saeli, H., & Rahmati, P. (2023). Developments in learners' affective engagement with written peer feedback: The affordances of in situ translanguaging. *Assessing Writing*, 58, Article 100788. <https://doi.org/10.1016/j.asw.2023.100788>
- Safari, F., & Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *Journal of Second Language Writing*, 60, Article 101007. <https://doi.org/10.1016/j.jslw.2023.101007>
- Saglam, A. L. G., & Tsagari, D. (2022). Perceptions towards the consequential validity of integrated language proficiency assessment'. *Languages*, 7(1), 65. <https://doi.org/10.3390/languages7010065>
- Sagre, A., Herazo, J. D., & Davin, K. J. (2022). Contradictions in teachers' classroom dynamic assessment implementation: An activity system analysis. *TESOL Quarterly*, 56(1), 154-177. <https://doi.org/10.1002/tesq.3046>

- Şahin, A. (2021). The Rater Performance Categorization System (RPCS) for intensive English programs. *Shanlax International Journal of Education*, 9(3), 225–241. <https://doi.org/10.34293/education.v9i3.3986>
- Saif, S., Ma, J., May, L., & Cheng, L. (2021). Complexity of test preparation across three contexts: Case studies from Australia, Iran and China. *Assessment in Education: Principles, Policy & Practice*, 28(1), 37–57. <https://doi.org/10.1080/0969594X.2019.1700211>
- Saito, H., Sawaki, Y., & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and use for Japanese university admission. *Language Assessment Quarterly*, 19(4), 422–448. <https://doi.org/10.1080/15434303.2022.2083965>
- Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, 55(3), 866–900. <https://doi.org/10.1002/tesq.3027>
- Saito, K., & Liu, Y. (2022). Roles of collocation in L2 oral proficiency revisited: Different tasks, L1 vs. L2 raters, and cross-sectional vs. longitudinal analyses. *Second Language Research*, 38(3), 531–554. <https://doi.org/10.1177/0267658320988055>
- Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 46(4), 1206–1230. <https://doi.org/10.1017/S027226312200047X>
- Saito, K., Macmillan, K., Kachlicka, M., Kuniyara, T., & Minematsu, N. (2023). Automated assessment of second language comprehensibility: Review, training, validation, and generalization studies. *Studies in Second Language Acquisition*, 45(1), 234–263. <https://doi.org/10.1017/S0272263122000080>
- Salaberry, M. R. & Weideman, A. (2024). Context, construct and ethics. In M. R., Salaberry, A. Weideman, A. & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 3-27). Routledge. <https://doi.org/10.4324/9781003384922-2>
- Salamoura, A., & Morgan, S. (2021). Learning-oriented assessment from a teacher's perspective: Insights from teachers' action research. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp.182–206). Routledge. <https://doi.org/10.4324/9781003014102-13>
- Sánchez-Pérez, M. D. M. (2021). Predicting content proficiency through disciplinary-literacy variables in English-medium writing. *System*, 97, Article 102463. <https://doi.org/10.1016/j.system.2021.102463>
- Sarac, M., & Loken, E. (2023). Examining patterns of omitted responses in a large-scale English language proficiency test. *International Journal of Testing*, 23(1), 56-72. <https://doi.org/10.1080/15305058.2022.2070756>
- Sarte, K. M., & Gnevsheva, K. (2022). Noun phrasal complexity in ESL written essays under a constructed-response task: Examining proficiency and topic effects. *Assessing Writing*, 51, Article 100595. <https://doi.org/10.1016/j.asw.2021.100595>
- Sasayama, S., Garcia Gomez, P., & Norris, J. M. (2021). Designing efficient L2 writing assessment tasks for low-proficiency learners of English. *ETS Research Report Series*, 2021(1), 1–31. <https://doi.org/10.1002/ets2.12341>
- Sato, T. (2022). Assessing critical thinking through L2 argumentative essays: an

- investigation of relevant and salient criteria from raters' perspectives. *Language Testing in Asia*, 12, Article 9. <https://doi.org/10.1186/s40468-022-00159-4>
- Sato, T. (2024). Assessing the content quality of essays in content and language integrated learning: Exploring the construct from subject specialists' perspectives. *Language Testing*, 41(2), 316–337. <https://doi.org/10.1177/02655322231190058>
- Saville, N. (2021). Learning-oriented assessment: Basic concepts and frameworks in using assessment to support language learning. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 13–33). Routledge. <https://doi.org/10.4324/9781003014102-3>
- Savski, K. (2023). CEFR and the ELT practitioner: Empowerment or enforcement?. *ELT Journal*, 77(1), 62–71. <https://doi.org/10.1093/elt/ccac013>
- Sawaki, Y. (2024). [Review of the book *Validity Argument in Language Testing: Case Studies of Validation Research* by C. A. Chapelle and E. Voss]. *Language Testing*, 41(1), 227–229. SAGE Publications. <https://doi.org/10.1177/02655322231193705>
- Sawaki, Y., Ishii, Y., & Oi, Y. (2024). The efficacy of automated feedback on summary content in an online formative assessment module for undergraduate L2 writing instruction in Japan. *Studies in Language Assessment (SILA)*, 13(2), 38–72. <https://doi.org/10.58379/mjfp3942>
- Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, 14, Article 50. <https://doi.org/10.1186/s40468-024-00325-w>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14, Article 49. <https://doi.org/10.1186/s40468-024-00321-0>
- Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to Generate Reading Comprehension Items. *Educational Measurement: Issues and Practice*, 43(1), 5–18. <https://doi.org/10.1111/emip.12590>
- Sayyadi, A., & Rezvani, R. (2021). Questioning in TOEFL iBT speaking test: A case of washback and construct underrepresentation. *Language Testing in Asia*, 11, Article 21. <https://doi.org/10.1186/s40468-021-00137-2>
- Schildt, L., Deygers, B., & Weideman, A. (2024). Language testers and their place in the policy web. *Language Testing*, 41(2), 338–356. <https://doi.org/10.1177/02655322231191133>
- Schissel, J. L., & Khan, K. (2021). Responsibilities and opportunities in language testing with respect to historicized forms of socio-political discrimination: A matter of academic citizenship. *Language Testing*, 38(4), 640–648. <https://doi.org/10.1177/02655322211028590>
- Schissel, J., Korne, H., & López-Gopar, M. (2021). Grappling with translanguaging for teaching and assessment in culturally and linguistically diverse contexts: Teacher perspectives from Oaxaca, Mexico. *International Journal of Bilingual Education and Bilingualism*, 24(3), 340–356. <https://doi.org/10.1080/13670050.2018.1463965>
- Schmidgall, J., & Powers, D. E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC® Speaking test scores from linguistic laypersons. *Language Testing*, 38(2), 302–325.

- <https://doi.org/10.1177/0265532220941803>
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge® tests. *ETS Research Report Series*, 2021(1), 1–22. <https://doi.org/10.1002/ets2.12335>
- Schmidgall, J., Huo, Y., Cid, J., & Wei, Y. (2024). Investigating fairness claims for a general-purposes assessment of English proficiency for the international workplace: Do full-time employees have an unfair advantage over full-time students? *ETS Research Report Series*, 2024(1), 1–20. <https://doi.org/10.1002/ets2.12380>
- Schmitt, Romain & Saif, Shahrzad. 2022. L'utilisation du TFLM comme mesure de compétence langagière des candidats aux programmes de formation initiale en enseignement au Québec est-elle justifiée? Un argumentaire d'utilisation des évaluations. In N. Talbot et S. Eleuch (Eds.), *L'évaluation des compétences: une pluralité de défis*, Éditions JFD.
- Schnoor, B., Hartig, J., Klinger, T., Naumann, A., & Usanova, I. (2023). Measuring the development of general language skills in English as a foreign language—Longitudinal invariance of the C-test. *Language Testing*, 40(3), 796-819. <https://doi.org/10.1177/02655322231159829>
- Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency, and task effects in Spanish Dual Language Immersion. *Language Learning & Technology*, 25(1), 53–72. <https://hdl.handle.net/10125/73425>
- Screti, F. (2023). Using google docs for guided academic writing assessments: Students' perspectives. *Language Learning in Higher Education*, 13(1), 153-174. <https://doi.org/10.1515/cercles-2023-2007>
- Sebolai, K. (2021). Generic academic literacy testing: A logical precursor for faculty-specific language teaching and assessment. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 151-169). Multilingual Matters.
- Sebolai, K. (2022). Determining the faculty specific academic literacies needs of first year university students: a mixed methods approach. *Journal for Language Teaching*, 56(2), 1-20. <https://doi.org/10.56285/jltVol56iss2a5702>
- Sebolai, K. (2022). The Academic Development approach to academic literacy in higher education South Africa: a disconnect between teaching and assessment. *Journal for Language Teaching*, 56(2), 1-16. <https://doi.org/10.56285/jltVol56iss2a5387>
- Seed, G., Salamoura, A. & Saville, N. (2024). New approaches to the assessment of English as an additional language. In N. A. Nazari & A. Mehdi Riazi (Eds.), *Adaptable English Language Teaching: Advances and Frameworks for Responding to New Circumstances* (pp.79-95). Routledge.
- Seedhouse, P., & Satar, M. (2021). VEO IELTS Project Report: Which specific features of candidate talk do examiners orient to when taking scoring decisions? *IELTS Research Reports Online Series*, 5. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/for-researchers/research-reports/online-series-2021-5>
- Selvi, A. F., Saracoğlu, E., & Çalıřkan, E. (2023). When inclusivity means playing safe: Ideological discourses and representations in English testing materials. *RELC Journal*, 54(3), 773-787. <https://doi.org/10.1177/00336882211008662>

- Seymour, J., & Angelo, D. (2023). Seeing the positives in assessment. Contributing to a “literature of doing” school-based Aboriginal language revival programs. *Studies in Language Assessment*, 12(2), 168-204.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4820543/SiLA_12_2_Seymour_and_Angelo.pdf
- Shafipoor, M., Ravand, H., & Maftoon, P. (2021). Test-level and item-level model fit comparison of general vs. specific diagnostic classification models: A case of true DCM. *Language Testing in Asia*, 11, Article 33.
<https://doi.org/10.1186/s40468-021-00148-z>
- Shahmirzadi, N. (2023). Validation of a language center placement test: Differential item functioning. *International Journal of Language Testing*, 13(1), 1-17.
<https://doi.org/10.22034/IJLT.2022.336779.1151>
- Shahmirzadi, N. (2024). Simulation study: Seeking fairness in a high-stakes placement test. *Language Testing in Focus: An International Journal*. 10, 18–28, (e-ISSN: 2717-9087). <https://doi.org/10.32038/ltf.2024.10.02>
- Shahmirzadi, N., & Marashi, H. (2023). Cognitive diagnostic assessment of reading comprehension for high-stakes tests: Using GDINA model. *Language Testing in Focus: An International Journal*, (8), 1–16. <https://doi.org/10.32038/ltf.2023.08.01>
- Shahmirzadi, N., Siyyari, M., Marashi, H., & Gramipour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA. *Journal of Foreign Languages Research*, 10(1), 152–165, (ISC), (ISSN: 2588–4123, Online ISSN: 2588–7521). DOI: 10.22059/JFLR.2019.282328.634
- Shahzadi, A., & Ducasse, A. M. (2022). Language assessment literacy of teachers in an English medium of instruction university: Implications for ELT training in Pakistan. *Studies in Language Assessment*, 11(1), 92–118.
- Shamsaldeen, F., Wang, J., & Ahn, S. (2024). Evaluating the fairness of a high-stakes college entrance exam in Kuwait. *Language Testing in Asia*, 14, Article 27.
<https://doi.org/10.1186/s40468-024-00301-4>
- Shamshiri, F., Esfahani, F. R., & Hosseini, S. E. (2023). Models of assessment in the classroom: A comparative research of CALL-based vs. traditional assessment on vocabulary learning among Iranian EFL learners. *Language Testing in Asia*, 13(1), Article 43. <https://doi.org/10.1186/s40468-023-00259-9>
- Shan, Z., Yang, H., & Xu, H. (2023). The mediating role of curriculum configuration on teacher’s L2 writing assessment literacy and practices in embedded French writing. *Assessing Writing*, 57, Article 100742. <https://doi.org/10.1016/j.asw.2023.100742>
- Shang, Y., Aryadoust, V., & Hou, Z. (2024). A meta-analysis of the reliability of second language listening tests (1991–2022). *Brain Sciences*, 14(8), 746.
<https://doi.org/10.3390/brainsci14080746>
- Shehata, A. (2024). Arabic pronunciation instruction: Learner cognition and speech comprehensibility. *Journal of Second Language Pronunciation*, 10(2), 179-203.
<https://doi.org/10.1075/jslp.23015.she>
- Shen, B., & Bai, B. (2024). Enhancing Chinese university students’ writing performance and self-regulated learning (Srl) writing strategy use through a strategy-based intervention. *System*, 122, 103249. <https://doi.org/10.1016/j.system.2024.103249>
- Sherkuziyeva, N., Imamutdinovna Gabidullina, F., Ahmed Abdel-Al Ibrahim, K., & Bayat, S. (2023). The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners’ oral proficiency, writing performance, and

- test anxiety. *Language Testing in Asia*, 13, Article 15.
<https://doi.org/10.1186/s40468-023-00227-3>
- Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361-388.
- Shi, G., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*. Advance online publication.
<https://doi.org/10.1017/S0958344023000265>
- Shi, H., & Aryadoust, V. (2023). A systematic review of automated writing evaluation within the argument-based validation framework. *Education and Information Technology*, 28, 771-795. <https://link.springer.com/article/10.1007/s10639-022-11200-7>
- Shi, X., Ma, X., Du, W., & Gao, X. (2024). Diagnosing Chinese EFL learners' writing ability using polytomous cognitive diagnostic models. *Language Testing*, 41(1), 109-134.
<https://doi.org/10.1177/02655322231162840>
- Shi, X., Wang, X., & Zhang, W. (2024). Exploring the relationships between ASS indices and CAF and the impact on Chinese college students' oral English performance. *Language Testing in Asia*, 14, Article 30.
<https://doi.org/10.1186/s40468-024-00306-z>
- Shi, Z., & Li, S. (2024). From mock interviews to real job interviews: Assessing the efficacy of business Chinese instruction. *Foreign Language Annals*, 57(4), 962-980.
<https://doi.org/10.1111/flan.12759>
- Shi, Z., Liu, F., Lai, C., & Jin, T. (2022). Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback. *Language Learning & Technology*, 26(2), 106-128.
<https://doi.org/10.125/73481>
- Shin, D. (2024). Critical discursive approaches to evaluating policy-driven testing: Social impact as a target for validation. *Language Testing*, 41(1), 162-180.
<https://doi.org/10.1177/02655322231163863>
- Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts? *Language Learning & Technology*, 27(3), 27-40.
<https://hdl.handle.net/10125/73530>
- Shin, D., & Lee, J. H. (2024). AI-powered automated item generation for language testing. *ELT Journal*, 78(4), 446-452. <https://doi.org/10.1093/elt/ccae016>
- Shin, D., Kwon, S. K., & Lee, Y. (2021). The effect of using online language-support resources on L2 writing performance. *Language Testing in Asia*, 11, Article 4.
<https://doi.org/10.1186/s40468-021-00119-4>
- Shin, D., Park, S., & Cho, E. (2023). A review study on discourse-analytical approaches to language testing policy in the South Korean context. *Language Testing in Asia*, 13, Article 44. <https://doi.org/10.1186/s40468-023-00257-x>
- Shin, J. (2022). Investigating and optimizing score dependability of a local ITA speaking test across language groups: A generalizability theory approach. *Language Testing*, 39(2), 313-337. <https://doi.org/10.1177/02655322211052680>
- Shin, J., & Gierl, M. J. (2021). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 38(2), 247-272. <https://doi.org/10.1177/0265532220937830>
- Shin, J., & Gierl, M. J. (2022). Generating reading comprehension items using automated processes. *International Journal of Testing*, 22(3-4), 289-311.
<https://doi.org/10.1080/15305058.2022.2070755>

- Shin, J., Rodríguez-Fuentes, R. A., Swatek, A. M., & Ginther, A. (2022). Aptis test review. *Language Testing*, 39(1), 172–187. <https://doi.org/10.1177/02655322211032873>
- Shin, S. Y., Lee, S., & Lidster, R. (2021). Examining the effects of different English speech varieties on an L2 academic listening comprehension test at the item level. *Language Testing*, 38(4), 580–601. <https://doi.org/10.1177/0265532220985432>
- Shin, S., & Lee, S. (2024). Effects of the type of audio-visual input and listenability on young L2 learners' listening comprehension. *TESOL Quarterly*, 58(4), 1580–1608. <https://doi.org/10.1002/tesq.3298>
- Shoahosseini, R., Baghaei, P., Khodabakhshzadeh, H., & Ashraf, H. (2024). C-Test construct validity: Evidence from nonparametric item response theory. *Language Testing in Asia*, 14, Article 10. <https://doi.org/10.1186/s40468-024-00279-z>
- Shohamy, E. (2022). Critical language testing, multilingualism and social justice. *TESOL Quarterly*, 56(4), 1445–1457. <https://doi.org/10.1002/tesq.3185>
- Shrestha, P. N. (2022). Examining evaluative language used in assessment feedback on business students' academic writing. *Assessing Writing*, 54, Article 100664. <https://doi.org/10.1016/j.asw.2022.100664>
- Siekmann, L., Parr, J. M., & Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing*, 54, Article 100672. <https://doi.org/10.1016/j.asw.2022.100671>
- Silva, A. L. B. C. (2023). Aviation English assessment for military and civil pilots in Brazil: Similarities and differences. *Applied Linguistics Papers*, 27(2), 53–62. <https://doi.org/10.32612/uw.25449354.2023.1.pp.53-62>
- Silva, A. L. B. C. (2023). Language proficiency test design for a group of military pilots: Criteria, construct and purpose. *Journal of Teaching English for Academic and Specific Purposes*, 11(1), 1–9. <https://doi.org/10.22190/JTESAP230120002S>
- Silva, A. L. B. C. & Raymundo, N. A. (2023). Test preparation issues in the aeronautical context in Brazil. In P. Tosqui-Lucks & J. C. Santana (Eds.), *Aviation English – A global perspective: Analysis, teaching, assessment* (pp. 172–182). Bookerfield Editors. <https://bookerfield.com/catalogo/10.53268/BKF22080500>
- Sims, M. L. (2023). Shifting perceptions of socially just writing assessment: Labor-based contract grading and multilingual writing instruction. *Assessing Writing*, 57, Article 100731. <https://doi.org/10.1016/j.asw.2023.100731>
- Sinha, T. S., & Nassaji, H. (2022). ESL learners' perception and its relationship with the efficacy of written corrective feedback. *International Journal of Applied Linguistics*, 32(1), 41–56. <https://doi.org/10.1111/ijal.12378>
- Sippel, L., & Martin, I. A. (2024). Is corrective feedback during telecollaboration beneficial? The effects of peer and teacher corrections on L2 writing proficiency. *Journal of Second Language Writing*, 64, Article 101098. <https://doi.org/10.1016/j.jslw.2024.101098>
- Skaggs, G. (2021). *Test development and validation*. SAGE Publications.
- Skar, G. B., & Huebner, A. (2022). Predicting first grade students' writing proficiency. *Assessment in Education: Principles, Policy & Practice*, 29(2), 219–237. <https://doi.org/10.1080/0969594X.2022.2057424>
- Skehan, P., Bui, G., Wang, Z., & Shum, S. (2024). Re-examining accuracy measures in second language task-based spoken performance. *Research Methods in Applied Linguistics*, 3(1), Article 100098. <https://doi.org/10.1016/j.rmal.2024.100098>

- Sohrabi, Z., Ghanbari, N., & Abbasi, A. (2022). Iranian EFL teachers' perceptions of writing assessment literacy: a countrywide study. *Language Testing in Asia*, 12, Article 23. <https://doi.org/10.1186/s40468-022-00172-7>
- Sok, S., & Shin, H. W. (2023). Investigating the dimensions and determinants of children's narrative writing in Korean. *Assessing Writing*, 57, Article 100740. <https://doi.org/10.1016/j.asw.2023.100740>
- Sok, S., Shin, H. W., & Do, J. (2021). Exploring which test-taker characteristics predict young L2 learners' performance on listening and reading comprehension tests. *Language Testing*, 38(3), 378–400. <https://doi.org/10.1177/0265532221991134>
- Solon, M., Park, H. I., Dehghan-Chaleshtori, M., Carver, C., & Long, A. Y. (2022). Exploring an elicited imitation task as a measure of heritage language proficiency. *Studies in Second Language Acquisition*, 44(4), 1095–1123. <https://doi.org/10.1017/S0272263121000905>
- Song, Y., Sparks, J. R., & Guzman-Orth, D. (2021). Investigating student performance on an assessment tool to support the development of English learners' argumentation skills. *Computer Assisted Language Learning*, 35(9), 2291–2311. <https://doi.org/10.1080/09588221.2021.1879160>
- Sparks, R. L. (2021). Identification and characteristics of strong, average, and weak foreign language readers: The simple view of reading model. *The Modern Language Journal*, 105(2), 507–525. <https://doi.org/10.1111/modl.12711>
- Spiby, R., & Clark, T. (2023). Responding to a transformed assessment landscape: The example of IELTS. In H. I. Işık, E. Kantarcıoğlu, A. Üner & C. Thomas (Eds.), *Re-Vision and Re-Form in English for academic purposes contexts after the pandemic* (pp. 83–93). Newcastle: Cambridge Scholars Publishing.
- Spino, L. L., Echevarría, M. M., & Wu, Y. (2022). Assisted self-assessment to optimize the OPIc test experience. *Foreign Language Annals*, 55(3), 853–876. <https://doi.org/10.1111/flan.12645>
- Stansfield, C.W., Gao, J., & Rivers, W.P. (2010). A concurrent validity study of self-assessments and the federal interagency roundtable oral proficiency interview. *Russian Language Journal / Русский Язык*, 60, 299–315. <http://www.jstor.org/stable/43669189>
- Steiss, J., Krishnan, J., Kim, Y-S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, Article 100600. <https://doi.org/10.1016/j.asw.2021.100600>
- Stephenson, M. & Hall, G. (2021). Organizing talk in group speaking tests: Learning from high-scoring students. *ELT Journal*, 75(1), 42–54. <https://doi.org/10.1093/elt/ccaa052>
- Stewart, J., Gyllstad, H., Nicklin, C., & McLean, S. (2024). Establishing meaning recall and meaning recognition vocabulary knowledge as distinct psychometric constructs in relation to reading proficiency. *Language Testing*, 41(1), 89–108. <https://doi.org/10.1177/02655322231162853>
- Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, 43(2), 462–471. <https://doi.org/10.1017/S0272263121000437>
- Stewart, J., Vitta, J. P., Nicklin, C., McLean, S., Pinchbeck, G. G., & Kramer, B. (2022). The relationship between word difficulty and frequency: a response to Hashimoto (2021).

- Language Assessment Quarterly, 19(1), 90-101.
<https://doi.org/10.1080/15434303.2021.1992629>
- Steyn, S. (2021). Pathways to parity between parallel tests of language ability: Lessons from a project. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 132-147). Multilingual Matters.
- Stoeckel, T., Ha, H. T., Nguyen, D. T. B., & Nicklin, C. (2024). Can an extended-matching second-language vocabulary test format bridge the gap between meaning-recognition and meaning-recall?. *Research Methods in Applied Linguistics*, 3(2), Article 100109. <https://doi.org/10.1016/j.rmal.2024.100109>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>
- Su, H., & Lu, X. (2022). Assessing pragmatic performance in advanced L2 academic writing through the lens of local grammars: A case study of “exemplification.” *Assessing Writing*, 54, Article 100668. <https://doi.org/10.1016/j.asw.2022.100668>
- Su, X., & Lee, I. (2024). Is the magic in the mix? The development and validation of the L2 teachers’ blended assessment literacy scale. *System*, 123, 103354. <https://doi.org/10.1016/j.system.2024.103354>
- Su, Y., & Chen, X. (2024). Examining refusal and acceptance sequences employing a data-driven binary rating approach. *Research Methods in Applied Linguistics*, 3(3), Article 100161. <https://doi.org/10.1016/j.rmal.2024.100161>
- Su, Y., & Shin, S.-Y. (2024). Comparing two formats of data-driven rating scales for classroom assessment of pragmatic performance with roleplays. *Language Testing*, 41(2), 357–383. <https://doi.org/10.1177/02655322231210217>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Suárez, M. M. (2022). A need for an aptitude test for young learners in Catalan: The case of the Modern Language Aptitude Test - Elementary in Catalan. *Language Teaching Research Quarterly*, 31, 101-118. <https://doi.org/10.32038/ltrq.2022.31.08>
- Suárez, M.-d.-M., & Stansfield, C. W. (2023). Young learners’ bilingual status and cognitive development in foreign language aptitude testing. *ITL - International Journal of Applied Linguistics*, 174(2), 291-315. <https://doi.org/https://doi.org/10.1075/itl.22017.sua>
- Sudina, E., Vernon, T., Foster, H., Del Villano, H., Hernandez, S., Beck, D., & Plonsky, L. (2021). Development and initial validation of the L2-teacher grit scale. *TESOL Quarterly*, 55(1), 156–184. <https://doi.org/10.1002/tesq.581>
- Suhan, M., Papageorgiou, S., & Wolf, M. K. (2024). Mapping the scores of the TOEFL Primary® Writing test to the Common European Framework of Reference (CEFR) levels (Research Memorandum ETS RM-24-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-03.pdf>
- Sun, T. & Kim, S. Y. (2021). Evaluating six approaches to handling zero-frequency scores under equipercentile equating. *Measurement: Interdisciplinary Research and Perspectives*, 19(4), 213–235. <https://doi.org/10.1080/15366367.2020.1855034>
- Sun, Y., Lan, G., & Zhang, L. (2023). Pedagogical values of translingual practices in improving student feedback literacy in academic writing. *Assessing Writing*, 56,

- Article 100715. <https://doi.org/10.1016/j.asw.2023.100715>
- Sun, Z., Xu, P., & Wang, J. (2023). Dynamic assessment of the learning potential of Chinese as a second language. *Language Assessment Quarterly*, 20(1), 127-142. <https://doi.org/10.1080/15434303.2022.2151911>
- Sundqvist, P., & Uztosun, M. S. (2024). Extramural English in Scandinavia and Asia: scale development, learner engagement, and perceived speaking ability. *TESOL Quarterly*, 58(4), 1638–1665. <https://doi.org/10.1002/tesq.3296>
- Sung, H., Cho, S., & Kyle, K. (2024). An Empirical Evaluation of Lexical Diversity Indices in L2 Korean Writing Assessment. *Language Assessment Quarterly*, 21(2), 159–180. <https://doi.org/10.1080/15434303.2024.2311728>
- Sureeyatanapas, P., Sureeyatanapas, P., Panitanarak, U., Kraisriwattana, J., Sarootyanapat, P., & O'Connell, D. (2024). The analysis of marking reliability through the approach of gauge repeatability and reproducibility (GR&R) study: A case of English-speaking test. *Language Testing in Asia*, 14, Article 1. <https://doi.org/10.1186/s40468-023-00271-z>
- Suto, I., Williamson, J., Ireland, J., & Macinska, S. (2023). On reducing errors in assessment instruments. *Research Papers in Education*, 38(3), 357-377. <https://doi.org/10.1080/02671522.2021.1968940>
- Suvorov, R., & He, S. (2022). Visuals in the assessment and testing of second language listening: A methodological synthesis. *International Journal of Listening*, 36(2), 80-99. <https://doi.org/10.1080/10904018.2021.1941028>
- Suvorov, R., & Li, Z. (2023). *Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS*. British Council, Cambridge Assessment English, and IDP: IELTS Australia. <https://search.informit.org/doi/epdf/10.3316/informit.262825338814353>
- Suvorov, R., Karatay, Y., & Hegelheimer, V. (2024). Computer-Assisted Language Testing. In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment*. Wiley.
- Suzuki, R., & Yamane, S. (2024). Japanese EFL learners' comprehension and perceptions of three different accents in listening assessment. *International Journal of Applied Linguistics*, 34(4), 1497-1513. <https://doi.org/10.1111/ijal.12584>
- Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, 105(2), 435–463. <https://doi.org/10.1111/modl.12706>
- Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, 43(3), 663–676. <https://doi.org/10.1017/S0272263120000704>
- Suzuki, Y., & Koizumi, R. (2021). Using equivalent test forms in SLA pretest-posttest design research. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 457–467). Routledge. <https://doi.org/10.4324/9781351034784-49>
- Suzuki, Y., Eguchi, M., & de Jong, N. (2022). Does the reuse of constructions promote fluency development in task repetition? A usage-based perspective. *TESOL Quarterly*, 56(4), 1290-1319. <https://doi.org/10.1002/tesq.3103>
- Suzumura, N. (2022). Content analysis of test taker responses on an AP Japanese computer-simulated conversation test: A mixed methods approach for a validity argument. *Language Assessment Quarterly*, 19(5), 480-502.

<https://doi.org/10.1080/15434303.2022.2130326>

- Syquia, J., & Leeming, P. (2024). Assessing interactional competence through group discussion: A mixed methods validation. *Research Methods in Applied Linguistics*, 3(3), Article 100144. <https://doi.org/10.1016/j.rmal.2024.100144>
- Szabo, C., Stickler, U., & Adinolfi, L. (2021). Predicting the academic achievement of multilingual students of English through vocabulary testing. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1531–1542. <https://doi.org/10.1080/13670050.2020.1814196>

I

- Tabari, M. A., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing*, 57, Article 100749. <https://doi.org/10.1016/j.asw.2023.100749>
- Tabari, M. A., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing*, 58, Article 100783. <https://doi.org/10.1016/j.asw.2023.100783>
- Tabari, M. A., Wang, Y., & Miller, M. (2024). Understanding the SSARC model of task sequencing: Assessing L2 writing development. *Assessing Writing*, 62, 100893. <https://doi.org/10.1016/j.asw.2024.100893>
- Taguchi, N. (2022). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, 56(1), 308–335. <https://doi.org/10.1002/tesq.3070>
- Taguchi, N. (2023). Using immersive virtual reality for the assessment of intercultural conflict mediation. *Language Learning & Technology*, 27(3), 108–128. <https://hdl.handle.net/10125/73534>
- Tai, H., Liu, C., & Chen, Y. (2023). Validation of a multiple-choice implicature test: insights from Chinese EFL learners' cognitive processes. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1481–1506. <https://doi.org/10.1515/iral-2021-0190>
- Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, 39(4), 649–678. <https://doi.org/10.1177/02655322211057040>
- Talebinamvar, M., & Zarrabi, F. (2022). Clustering students' writing behaviors using keystroke logging: a learning analytic approach in EFL writing. *Language Testing in Asia*, 12, Article 6. <https://doi.org/10.1186/s40468-021-00150-5>
- Tan, L. Y., McLean, S., Kim, Y. A., & Vitta, J. P. (2024). Rasch modelling vs. item facility: Implications on the validity of assessments of Asian EFL/ESL vocabulary knowledge and lexical sophistication modelling. *Language Testing in Asia*, 14, Article 55. <https://doi.org/10.1186/s40468-024-00327-8>
- Tan, T. X., Fan, X., Braunstein, L. B., & Lane-Holbert, M. (2022). Linguistic, cultural and substantive patterns in L2 writing: A qualitative illustration of MisLevy's sociocognitive perspective on assessment. *Assessing Writing*, 51, Article 100574. <https://doi.org/10.1016/j.asw.2021.100574>
- Tanaka, M., & Ross, S. J. (2023). Impact of self-construal on rater severity in peer assessments of oral presentations. *Assessment in Education: Principles, Policy & Practice*, 30(2), 203–220. <https://doi.org/10.1080/0969594X.2023.2189564>
- Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in

- second language Chinese. *The Modern Language Journal*, 105(3), 740–759.
<https://doi.org/10.1111/modl.12725>
- Tang, Y., & Ma, X. (2023). An interventionist dynamic assessment approach to college English writing in China. *Language Assessment Quarterly*, 20(1), 44-65.
<https://doi.org/10.1080/15434303.2022.2155165>
- Tao, X., & Aryadoust, V. (2024). A multidimensional analysis of a high-stakes English listening test: A corpus-Based Approach. *Education Sciences*, 14(2), 137.
<https://doi.org/10.3390/educsci14020137>
- Tarlani-Aliabadi, H., Tazik, K., & Azizi, Z. (2022). Exploring the role of language knowledge and background knowledge in reading comprehension of specific-purpose tests in higher education. *Language Testing in Asia*, 12, Article 48.
<https://doi.org/10.1186/s40468-022-00198-x>
- Tashmuradova, B., Prasad, K. D. V., Ahmed Abdel-Al Ibrahim, K., & Sarabani, L. (2023). The comparative effects of consciousness raising tasks and dynamic assessment on morphological awareness: The case of pre-intermediate EFL learners. *Language Testing in Asia*, 13, Article 19. <https://doi.org/10.1186/s40468-023-00231-7>
- Tavakoli, P. & Cooke, S. (2024). *Comprehensibility in Language Assessment: A Broader Perspective*. University of Toronto Press.
- Tavakoli, P., Kendon, G., Mazhurnaya, S., & Ziomek, A. (2023). Assessment of fluency in the test of English for educational purposes. *Language Testing*, 40(3), 607-629.
<https://doi.org/10.1177/02655322231151384>
- Tavassoli, K., & Rahmatollahi, M. (2024). Differential impact of asynchronous and synchronous computer-assisted dynamic assessment on higher-order and lower-order writing skills. *Journal of Language Horizons*, 8(1), 33-61.
<https://doi.org/10.22051/lghor.2023.41130.1720>
- Tavassoli, K., & Sorat, Z. (2023). Iranian EFL teachers' oral/aural skills language assessment literacy: Instrument development and validation. *International Journal of Language Testing*, 13(2), 56-76. <https://doi.org/10.22034/IJLT.2023.376466.1219>
- Tavassoli, K., Bashiri, L., & Pourdana, N. (2022). Experienced and novice L2 raters' cognitive processes while rating integrated and independent writing tasks. *Journal of Language and Education*, 8(4), 169-181. <https://doi.org/10.17323/jle.2022.13466>
- Tavassoli, K., Oskouiefar, M., & Ghamoushi, M. (2023). The effect of mobile-assisted learning-oriented assessment on EFL learners' writing ability. *Iranian Journal of Applied Linguistics*, 26(2), 6-6. <http://ijal.khu.ac.ir/article-1-3233-en.html>
- Taylor, L. (2023). Reframing the discourse and rhetoric of language testing and assessment for the public square. *Language Testing*, 40(1), 47-53.
<https://doi.org/10.1177/02655322221127421>
- Taylor, L., & Banerjee, J. (2023). Accommodations in language testing and assessment: Safeguarding equity, access, and inclusion. *Language Testing*, 40(4), 847-855.
<https://doi.org/10.1177/02655322231186221>
- Taylor, L., & Banerjee, J. (2023). Language assessment accommodations: Issues and challenges for the future. *Language Testing*, 40(4), 1000-1006.
<https://doi.org/10.1177/02655322231186222>
- Taylor, L., & Pill, J. (2024). Assessing health and other professionals. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 371–382). Wiley-Blackwell.
- Tayyebi, M., Abbasabady, M. M., & Abbassian, G. R. (2022). Examining classroom writing assessment literacy: a focus on in-service EFL teachers in Iran. *Language Testing in*

- Asia, 12, Article 12. <https://doi.org/10.1186/s40468-022-00161-w>
- Tenca, E., Hartle, S., & Piccinin, S. (2023). Teaching oral skills in Higher Education (HE) language learning: Piloting a model for inclusive, accessible English Language Teaching (ELT). *Instructed Second Language Acquisition*, 7(2), 130-165. <https://doi.org/10.1558/isla.26019>
- Teng, L. S. (2022). Explicit strategy-based instruction in L2 writing contexts: A perspective of self-regulated learning and formative assessment. *Assessing Writing*, 53, Article 100645. <https://doi.org/10.1016/j.asw.2022.100645>
- Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing*, 59, 100811. <https://doi.org/10.1016/j.asw.2024.100811>
- Teng, M. F., & Teng, L. S. (2024). Validating the multi-dimensional structure of self-efficacy beliefs in peer feedback for L2 writing: A bifactor-exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100136. <https://doi.org/10.1016/j.rmal.2024.100136>
- Teng, M. F., & Zhan, Y. (2023). Assessing self-regulated writing strategies, self-efficacy, task complexity, and performance in English academic writing. *Assessing Writing*, 57, Article 100728. <https://doi.org/10.1016/j.asw.2023.100728>
- Teng, M. F., & Zhang, L. J. (2024). Validating a new method for assessing young foreign language learners' metacognitive knowledge. *Research Methods in Applied Linguistics*, 3(2), Article 100121. <https://doi.org/10.1016/j.rmal.2024.100121>
- Teng, M. F., Wang, C., & Zhang, L. J. (2022). Assessing self-regulatory writing strategies and their predictive effects on young EFL learners' writing performance. *Assessing Writing*, 51, Article 100573. <https://doi.org/10.1016/j.asw.2021.100573>
- Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s. *System*, 100, Article 102567. <https://doi.org/10.1016/j.system.2021.102567>
- Thai, T., & Sheehan, S. (2022). The processes of rating L2 speaking performance using an analytic rating scale – A qualitative exploration. *Language Education and Assessment*, 5(1), 34–51. <https://doi.org/10.29140/lea.v5n1.777>
- Thi, N. K., Nikolov, M., & Simon, K. (2023). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 17(3), 690–705. <https://doi.org/10.1080/17501229.2022.2122476>
- Thomson, R. I., & Isaacs, T. (2022). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research* (pp. 24-44). Springer.
- Thwaites, P., Kollias, C., & Paquot, M. (2024). Is CJ a valid, reliable form of L2 writing assessment when texts are long, homogeneous in proficiency, and feature heterogeneous prompts? *Assessing Writing*, 60, 100843. <https://doi.org/10.1016/j.asw.2024.100843>
- Tian, W., Lê, H. D., & Thúy Nguyễn, N. (2023). Vietnamese university EFL teachers' practices and literacy in classroom assessment: A sociocultural perspective. *System*, 116, Article 103066. <https://doi.org/10.1016/j.system.2023.103066>
- Tian, W., Louw, S., & Khan, M. K. (2021). Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes. *System*, 103,

- Article 102682. <https://doi.org/10.1016/j.system.2021.102682>
- Timpe-Laughlin, V., & Cho, Y. (2021). Reflecting on assessing young foreign language learners. *Language Testing*, 38(3), 343–355.
<https://doi.org/10.1177/0265532221998264>
- Timpe-Laughlin, V., & Youn, S. J. (2021). Measuring L2 pragmatics. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.254–264). Routledge.
<https://doi.org/10.4324/9781351034784-28>
- Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2024). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*, 37(1-2), 149-178.
<https://doi.org/10.1080/09588221.2022.2032184>
- Tinkle, T., Godfrey, J., Menon, A. R., Moos, A., Romaine, L., & Sprouse, M. (2022). (In)Equities in directed self-placement. *Assessing Writing*, 54, Article 100671.
<https://doi.org/10.1016/j.asw.2022.100671>
- Tokunaga, H. (2021). *Moving from IBM SPSS to R and RStudio: A statistics companion*. SAGE.
- TONELLI, J. A. R., & QUEVEDO-CAMARGO, G. (2018) Teaching EFL to children: reflections on (future) teachers' language assessment literacy. *Revista Horizontes de Lingüística Aplicada*, v. 17, p. 63-91. DOI:
<https://doi.org/10.26512/rhla.v17i1.9209>
- Toprak, T. E., & Cakir, A. (2021). Examining the L2 reading comprehension ability of adult ELLs: Developing a diagnostic test within the cognitive diagnostic assessment framework. *Language Testing*, 38(1), 106–131.
<https://doi.org/10.1177/0265532220941470>
- Torregrossa, J., Listanti, A., Bongartz, C., & Narinis T. (2024). Adding discourse to sentence repetition tasks: Under which conditions does bilingual children's performance improve?. *Research Methods in Applied Linguistics*, 3(2), Article 100107.
<https://doi.org/10.1016/j.rmal.2024.100107>
- Treadaway, M., & Read, J. (2024). Setting standards for a diagnostic test of aviation English for student pilots. *Language Testing*, 41(3), 557–583.
<https://doi.org/10.1177/02655322231224051>
- Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, 5, 1-15. <https://doi.org/10.22599/jesla.74>
- Trofimovich, P., Isaacs, T., Kennedy, S., & Tsunemoto, A. (2022). Speech comprehensibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and L2 speaking* (pp. 174-187). Routledge.
- Trofimovich, P., Tekin, O., & McDonough, K. (2021). Task engagement and comprehensibility in interaction: Moving from what second language speakers say to what they do. *Journal of Second Language Pronunciation*, 7(3), 435–461.
<https://doi.org/10.1075/jslp.21006.tro>
- Tsagari, D. (2021). Gauging the assessment literacy levels of English teachers in Norway. *The European Journal of Applied Linguistics and TEFL*, 10(1), 161–191.
- Tsagari, D. (2022). Inclusive and accessible assessment. In D. Tsagari, J. Nijakowska, & E. Guz. (Eds.), *Good practices in inclusive language education and foreign language teacher training - universal design in education and increasing accessibility*,

especially for marginalized groups.

https://scaled.uw.edu.pl/wp-content/uploads/sites/115/2022/11/SCALED_Output_1_Report.pdf

- Tsagari, D. & H. Bøhn (eds.) (2022) Recent Developments in Language Testing and Assessment (Special Issue) Issue 7. *Languages* (open access journal) https://www.mdpi.com/journal/languages/special_issues/Language_Testing_Assessment?fbclid=IwAR2_A6SIGEKehpZ95o5AQvPmM3kzi6O3rqypHjoTPA22Lldl6F_dHJLpnas
- Tsagari, D., & Demetriou, T. (2022). Investigating the parameters accounting for language proficiency: The case of an English language test used for academic purposes. *Research Papers in Language Teaching and Learning*, 12(1), 89-110. https://rpltl.eap.gr/images/2022/01_loannou__Tsagari.pdf
- Tsagari, D., & Giannikas, C. (2021). The impact of cert-mania of English language learning and teaching: the Cypriot case. *The European Journal of Applied Linguistics and TEFL*, 10(1), 193–215.
- Tsagari, D., & Vogt, K. (2022). Introduction to the special issue: Contextualising language assessment literacy. *Studies in Language Assessment*, 11(1), ii-vii. https://www.altaanz.org/uploads/5/9/0/8/5908292/0._sila_11_1__front_matter_introduction__tsagari_vogt.pdf
- Tsagari, D., & Vogt, K. (eds.) (2022) Contextualizing Language Assessment Literacy (Special Issue). *Studies in Language Assessment (SiLA)*, formerly *Papers in Language Testing and Assessment (PLTA)*, 11(1). <https://www.altaanz.org/current-issue.html>
- Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238. <https://doi.org/10.1177/02655322211050600>
- Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Language Annals*, 55(1), 135– 154. <https://doi.org/10.1111/flan.12571>
- Tsunemoto, A., Uludag, P., McDonough, K., & Isaacs, T. (2023). Research forum: The relationship between Japanese EFL learners' perceived fluency and temporal speech measures in a read-aloud task. *JALT Journal*, 45, 107–121. <https://doi.org/10.37546/JALTJJ45.1-5>
- Turner, C. E. (2023). Assessment and evaluation: mixed methods research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Wiley Online Library. <https://doi.org/10.1002/9781405198431.wbeal0040.pub2>

U

- Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing*, 52, Article 100609. <https://doi.org/10.1016/j.asw.2022.100609>
- Uludag, P., Crawford, W. J., & McDonough, K. (2021). English for academic purposes writers' use of reporting verbs in argumentative and cause-and-effect essay exams. *BC TEAL Journal*, 6(1), 42–58. <https://doi.org/10.14288/bctj.v6i1.388>

- Uludag, P., McDonough, K., & Payant, C. (2021). Does prewriting planning positively impact English L2 students' integrated writing performance? *Canadian Journal of Applied Linguistics*, 24(3), 166–185. <https://doi.org/10.37213/cjal.2021.31313>
- Uludag, P., McDonough, K., & Trofimovich, P. (2022). Exploring shared and individual assessment of paired oral interactions. *Studies in Language Assessment*, 11(2), 1–24. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4258357/2022-Early-View_Uludag-et-al.pdf
- Uskokovic, B., & Schirm, S. (2024). Assessment of Interactional Competence in L2 German: Integrating an innovative rubric to help language teachers effectively evaluate pragmatics. *System*, 125, 103400. <https://doi.org/10.1016/j.system.2024.103400>
- Uyar, G., & Balaman, U. (2024). The Role of Gazing Behaviors in Navigating Paired Role-Play Interactional Competence Assessment Tasks. *Language Assessment Quarterly*, 21(3), 266–286. <https://doi.org/10.1080/15434303.2024.2375651>
- Uysal, H. (2022). Standardized tests and within-group segregation: The not-so-optimal ESL classroom for long-term English learners. *TESOL Quarterly*, 56(4), 1471-1485. <https://doi.org/10.1002/tesq.3187>
- Uzun, K. (2021). Performance prediction strengths of noun and verb phrases in L2 writing: Comparison of density and complexity variables. *Assessing Writing*, 50, Article 100572. <https://doi.org/10.1016/j.asw.2021.100572>

V

- Vakili, S., & Ebadi, S. (2022). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning*, 35(3), 345-380. <https://doi.org/10.1080/09588221.2019.1698616>
- Valentine, K. A., Truckenmiller, A. J., Troia, G. A., & Aldridge, S. (2021). What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change? *Assessing Writing*, 50, Article 100567. <https://doi.org/10.1016/j.asw.2021.100567>
- Valkova, R. B. (2024). Remote Proctoring in language assessment: Exploring the impact on test-takers' scores and perceptions. *Studies in Language Assessment (SILA)*, 13(1), 141-198. <https://doi.org/10.58379/FCZM2855>
- Van Dyk, T. (2021). Postscript: What the data tell us: an overview of language assessment research in South Africa's multilingual context. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 217–235). Multilingual Matters.
- Van Dyk, T., & Weideman, A. (2022). Taaltoetsing en die assessering van taalvermoë [Language testing and the assessment of language ability]. In W. Carstens & T. van Dyk (Eds.), *Toegepaste Taalkunde in Afrikaans [Applied linguistics in Afrikaans]* (pp. 125-1420). Van Schaik. <https://toegepaste-taalkunde.co.za/hoofstuk-7/>
- Van Dyk, T., Murre, P., & Kotzé, H. (2021). Does one size fit all? Some considerations for test translation. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp.

- 52–74). *Multilingual Matters*.
- Vanbuel, M., & Deygers, B. (2024). Assessing Receptive Skills in an Adult L2 Learner Population with Diverse Educational Backgrounds. *Language Assessment Quarterly*, 21(4–5), 379–403. <https://doi.org/10.1080/15434303.2024.2438151>
- Vanderwoude, A., & Shefarevich, L. (2024). Investigating the quality of source integration in an academic writing exam. *Studies in Language Assessment (SILA)*, 13(1), 37–68. <https://doi.org/10.58379/CUUO1529>
- Vandeweerd, N., Housen, A., & Paquot, M. (2023). Proficiency at the lexis–grammar interface: Comparing oral versus written French exam tasks. *Language Testing*, 40(3), 658–683. <https://doi.org/10.1177/02655322231153543>
- Vasu, K. A., Mei Fung, Y., Nimehchisalem, V., & Md Rashid, S. (2022). Self-regulated learning development in undergraduate ESL writing classrooms: Teacher feedback versus self-assessment. *RELC Journal*, 53(3), 612–626. <https://doi.org/10.1177/0033688220957782>
- Vidal Rodeiro, C., & Macinska, S. (2022). Equal opportunity or unfair advantage? The impact of test accommodations on performance in high-stakes assessments. *Assessment in Education : Principles, Policy & Practice*, 29(4), 462–481. <https://doi.org/10.1080/0969594X.2022.2121680>
- Viera, C., & Arispe, K. (2022). Can Spanish programs evaluate oral proficiency gains for all learners? An empirical analysis of traditional and nontraditional Spanish majors' oral proficiency development. *Foreign Language Annals*, 55(3), 832–852. <https://doi.org/10.1111/flan.12635>
- Villa Larenas, S., & Brunfaut, T. (2023). But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy. *Language Testing*, 40(3), 463–492. <https://doi.org/10.1177/02655322221134218>
- Villa Larenas, S., & Brunfaut, T. (2024). How do teacher educators learn about and teach language assessment? In B. Baker, & L. Taylor (Eds.), *Language assessment literacy and competence: Volume 1: Research and reflections from the field* (pp. 66–87). (Studies in Language Testing; Vol. 55). Cambridge University Press. <https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- Vo, Y., Rickels, H., Welch, C., & Dunbar, S. (2023). Human scoring versus automated scoring for English learners in a statewide evidence-based writing assessment. *Assessing Writing*, 56, Article 100719. <https://doi.org/10.1016/j.asw.2023.100719>
- Vogt, K. & Tzagari, D. (2022). Language assessment literacy and test validation: Highlighting the role of language teachers. *Studies in Language Assessment*, 11(1), 1–15. https://arts.unimelb.edu.au/__data/assets/pdf_file/0006/4282557/SiLA-11.1-Vogt_and_Tzagari.pdf
- Vogt, K., & Antia, B. E. (2024). *Multilingual assessment: Finding the nexus?* Peter Lang. <https://doi.org/10.3726/b21992>
- Vogt, K., & Tzagari, D. (2024). Insights from emergency remote language assessment for a post-pandemic world. *Studies in Language Assessment (SILA)*, 13(2), 115–146. <https://doi.org/10.58379/jjzp4250>
- Vogt, K., Bøhn, H., & Tzagari, D. (2024). Language assessment literacy. *Language Teaching*, 57(3), 325–340. doi:10.1017/S0261444824000090
- Voss, E. (2023). Proctoring remote language assessments. In K. Sadeghi (Eds.),

Fundamental considerations in technology mediated language assessment.
Routledge. <https://doi.org/10.4324/9781003292395-15>

- Voss, E. (2024). Language assessment and artificial intelligence. In A. Kunnan (Ed.) *The Concise Companion to Language Assessment* (pp. 112–125). John Wiley & Sons.
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4-5), 520-532.
<https://doi.org/10.1080/15434303.2023.2288256>
- Vössing, L., & Kern, F. (2023). The co-construction of pragmatic competencies in different settings: The case of two children with autism spectrum disorder. *Journal of Interactional Research in Communication Disorders*, 14(2), 195-219.
<https://doi.org/10.1558/jircd.24423>

W

- Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 17(1), 47–71. <https://doi.org/10.1080/17501229.2021.1914062>
- Wagner, E. (2022). Assessing listening. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing* (2nd Ed.) (pp. 223–235). Routledge.
- Wagner, E. (2022). Assessing second language listening. In H. Mohebbi & C. Coombe (Eds.), *Research questions in language education and applied linguistics* (pp. 299–304). Springer. https://doi.org/10.1007/978-3-030-79143-8_53
- Wagner, E., & Krylova, A. (2021). Temple University's ITA placement test in times of COVID-19. *Language Assessment Quarterly*, 18(1), 12–16.
<https://doi.org/10.1080/15434303.2020.1862849>
- Wagner, E., Batty, A. O. & Galaczi, E. (Eds.). *The Routledge Handbook of Second Language Acquisition and Listening* (1st ed.). Routledge.
- Wagner, E., Liao, Y. F., & Wagner, S. (2021). Authenticated spoken texts for L2 listening tests. *Language Assessment Quarterly*, 18(3), 205–227.
<https://doi.org/10.1080/15434303.2020.1860057>
- Wallace, M.P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72(1), 5-44. <https://doi.org/10.1111/lang.12424>
- Wang, Z., & Zhang, J. (2023). Mediation and learner reciprocity: Applying dynamic assessment to the oral proficiency interview framework. *Language and Sociocultural Theory*, 10(1), 82-105. <https://doi.org/10.1558/lst.22181>
- Wang, D. & East, M. (2024). Integrating translanguaging into assessment: students' responses and perceptions. *Applied Linguistics Review*, 15(5), 1911–1937.
<https://doi.org/10.1515/applirev-2023-0087>
- Wang, D., Zhang, L., & Huang, Y. (2024). Tracking the development of logical metaphor usage in argumentative writing: A longitudinal study with EFL learners. *System*, 122, 103292. <https://doi.org/10.1016/j.system.2024.103292>
- Wang, I. K., & Cohen, A. D. (2022). Self-access strategy instruction for academic writing vocabulary: What learners actually do. *Applied Linguistics*, 44(6), 976–1009.
<https://doi.org/10.1093/applin/amac056>
- Wang, J., Zheng, Y., & Zou, Y. (2024). Face validity and washback effects of the shortened

- PTE Academic: Insights from teachers in Mainland China. *Language Testing in Asia*, 14, Article 32. <https://doi.org/10.1186/s40468-024-00302-3>
- Wang, L., & Fan, J. (2021). *Working towards a proficiency scale of Business English writing: A mixed-methods approach*. Springer. <https://doi.org/10.1007/978-981-16-5449-7>
- Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing*, 50, Article 100571. <https://doi.org/10.1016/j.asw.2021.100571>
- Wang, M., Shen, H., Boland, J. E., & Wei, H. (2021). Visual context modulates L2 long-term structural priming for the Chinese ba construction. *The Modern Language Journal*, 105(2), 526–551. <https://doi.org/10.1111/modl.12715>
- Wang, P. & Cheng, L. (2024). The Impact of TOEFL iBT Test Preparation on Chinese Test-Takers' Perceived Test Value. *Critical Inquiry in Language Studies*.
- Wang, S. (2021). *Consequential validity of the Chinese Proficiency Test (HSK) from macro and micro perspectives*. Cambridge Scholars Publishing.
- Wang, T-H., Sun, Y., & Huang, N-W. (2023). Implementation of web-based dynamic assessment in improving low English achiever's learning effectiveness. *Computer Assisted Language Learning*, 36(8), 1450–1476. <https://doi.org/10.1080/09588221.2021.1998129>
- Wang, W., Kingston, N. M., Davis, M. H., Tiemann, G. C., Tonks, S., & Hock, M. (2021). Applying evidence-centered design in the development of a multidimensional adaptive reading motivation measure. *Educational Measurement: Issues and Practice*, 40(4), 91–100. <https://doi.org/10.1111/emip.12468>
- Wang, X., & Wang, J. (2024). Comparing Chinese L2 writing performance in paper-based and computer-based modes: Perspectives from the writing product and process. *Assessing Writing*, 61, 100849. <https://doi.org/10.1016/j.asw.2024.100849>
- Wang, Y., & Jiang, J. (2024). Exploring the development of noun phrase complexity in L2 English writings across two genres. *Assessing Writing*, 62, 100892. <https://doi.org/10.1016/j.asw.2024.100892>
- Wang, Y., & Ren, W. (2024). L2 grit and pragmatic comprehension among Chinese postgraduate learners: Enjoyment and anxiety as mediators. *International Journal of Applied Linguistics*, 34(1), 224–241. <https://doi.org/10.1111/ijal.12489>
- Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, 53, Article 100641. <https://doi.org/10.1016/j.asw.2022.100641>
- Wang, Y., & Xie, Q. (2022). Diagnostic assessment of novice EFL learners' discourse competence in academic writing: a case study. *Language Testing in Asia*, 12, Article 47. <https://doi.org/10.1186/s40468-022-00197-y>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, Article 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Wang, Y., Xu, M., Jiang, Y., & Jin, T. (2023). Connecting form with function: Model texts for bilingual learners' narrative writing. *Assessing Writing*, 57, Article 100753. <https://doi.org/10.1016/j.asw.2023.100753>
- Ward, E. (2021). An integrated mixed-methods study of contract grading's impact on adolescents' perceptions of stress in high school English: a pilot study. *Assessing Writing*, 48, Article 100508. <https://doi.org/10.1016/j.asw.2020.100508>
- Ward, E. (2021). Easing stress: Contract grading's impact on adolescents' perceptions of

- workload demands, time constraints, and challenge appraisal in high school English. *Assessing Writing*, 48, Article 100526. <https://doi.org/10.1016/j.asw.2021.100526>
- Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programmes. *Journal of English for Academic Purposes*, 61, Article 101203. <https://doi.org/10.1016/j.jeap.2022.101203>
- Watson, S. (2022). *Formative speaking assessment: Beginning with story* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Watzinger-Tharp, J., Tharp, D. S., & Rubio, F. (2021). Sustaining dual language immersion: Partner language outcomes in a statewide program. *The Modern Language Journal*, 105(1), 194–217. <https://doi.org/10.1111/modl.12694>
- Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/S0272263121000449>
- Wei, R., & Zhao, X. (2024). Effects of task-based language teaching on functional adequacy in L2 writing. *Assessing Writing*, 60, 100838. <https://doi.org/10.1016/j.asw.2024.100838>
- Weideman, A. (2021). A skills-neutral approach to academic literacy assessment. In A. Weideman, J. Read, & T. du Plessis, T. (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 22-51). Multilingual Matters.
- Weideman, A. (2022). Context, construct, and validation: A perspective from South Africa. *Language Assessment Quarterly*, 19(2), 124-141. <https://doi.org/10.1080/15434303.2020.1860991>
- Weideman, A. (2024). Yardsticks for the future of language assessment: Disclosing the meaning of measurement. In M. R. Salaberry, A. Weideman, A., & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 220-234). Routledge. <https://doi.org/10.4324/9781003384922-12>
- Weideman, A. & Van Dyk, T. (2023). Achieving technical economy: A modification of cloze procedure. *Language Teaching Research Quarterly*, 37, 144-160. <https://doi.org/10.32038/ltrq.2023.37.07>
- Weideman, A. 2023. The practicality of theory: Reciprocity, assessment and applied linguistics. *Stellenbosch Papers in Linguistics Plus*, 66, 77-196. <https://doi.org/10.5842/66-1-935>
- Weideman, A., & Deygers, B. (2024). Validity and validation: An alternative perspective. In M. R. Salaberry, A. Weideman, & W-L. Hsu.(Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 28-50). Routledge. <https://doi.org/10.4324/9781003384922-3>
- Weideman, A., Read, J. & Du Plessis, T. (Eds.). (2021). *Assessing academic literacy in a multilingual society: Transition and transformation* (Volume 84). Multilingual Matters. <https://doi.org/10.21832/WEIDEM6201>.
- Weiler, T. (2021). *Testing lexicogrammar: An investigation into the construct tested in the Language in Use section of the Austrian Matura in English*. Peter Lang.
- Weir, C. & Chan, S. (2019). *Research and practice in assessing academic English: The case of IELTS*. Cambridge University Press.
- Weisi, H., & Zandi, M. (2024). A mixed-method concurrent transformative study of metadiscourse markers employed by L2 speakers: Does proficiency level matter? *Language Testing in Asia*, 14, Article 60.

- <https://doi.org/10.1186/s40468-024-00330-z>
- Weisleder, A., Friend, M., Tsui, A. S. M., & Marchman, V. A. (2024). Using parent report to measure vocabulary in young bilingual children: A scoping review. *Language Learning*, 74(2), 468-505. <https://doi.org/10.1111/lang.12617>
- Weng, F., Zhao, C. G., & Chen, S. (2024). Effects of peer feedback in English writing classes on EFL students' writing feedback literacy. *Assessing Writing*, 61, 100874. <https://doi.org/10.1016/j.asw.2024.100874>
- Westbrook, C. (2023). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61, Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Wheeler, J. M., Engelhard, G., & Wang, J. (2022). Exploring rater accuracy using unfolding models combined with topic models: Incorporating supervised latent Dirichlet allocation. *Measurement: Interdisciplinary Research and Perspectives*, 20(1), 34–46. <https://doi.org/10.1080/15366367.2021.1915094>
- Wicaksono, B. H., Ismail, S. M., Sultanova, S. A., & Abeda, D. (2023). I like language assessment: EFL learners' voices about self-assessment, self-efficacy, grit tendencies, academic resilience, and academic demotivation in online instruction. *Language Testing in Asia*, 13, Article 37 (2023). <https://doi.org/10.1186/s40468-023-00252-2>
- Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2023). High-level listening comprehension in advanced English as a second language: Effects of the first language and inhibitory control. *Bilingualism: Language and Cognition*, 26(5), 865-879. <https://doi.org/10.1017/S1366728923000135>
- Williams, K. (2024). Exploring English for academic purposes instructors' perceptions of speech fluency through developing and piloting a rating scale for a paired conversational task. *System*, 121, 103266. <https://doi.org/10.1016/j.system.2024.103266>
- Wilsenach, C., & Schaefer, M. (2022). Development and initial validation of productive vocabulary tests for isiZulu, Siswati and English in South Africa. *Language Testing*, 39(4), 567–592. <https://doi.org/10.1177/02655322211063785>
- Wilson, J., & Huang, Y. (2024). Validity of automated essay scores for elementary-age English language learners: Evidence of bias?. *Assessing Writing*, 60, 100815. <https://doi.org/10.1016/j.asw.2024.100815>
- Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, 29(2), 183–199. <https://doi.org/10.1080/0969594X.2022.2025762>
- Wilson, J., Potter, A., Cordero, T. C., & Myers, M. C. (2023). Integrating goal-setting and automated feedback to improve writing outcomes: A pilot study. *Innovation in Language Learning and Teaching*, 17(3), 518–534. <https://doi.org/10.1080/17501229.2022.2077348>
- Wind, S. A. (2022). *Exploring rating scale functioning for survey research*. SAGE.
- Wind, S. A. (2023). A sequential approach to detecting differential rater functioning in sparse rater-mediated assessment networks. *Language Testing*, 40(2), 209-226. <https://doi.org/10.1177/02655322221092388>
- Winke, P. (2024). Sharing, collaborating, and building trust: How Open Science advances language testing. *Language Testing*, 41(4), 845–859.

- <https://doi.org/10.1177/02655322231211159>
- Winke, P. M. (2023). Forty years of Language Testing, and the changing paths of publishing. *Language Testing*, 40(1), 3-7.
<https://doi.org/10.1177/02655322221136802>
- Winke, P., & Brunfaut, T. (2021). Perspectives on “knowing” a second language: What are we seeking to measure? In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.1–8). Routledge.
<https://doi.org/10.4324/9781351034784-1>
- Winke, P., & Brunfaut, T. (Eds.). (2021). *The Routledge handbook of second language acquisition and language testing*. Routledge.
<https://doi.org/10.4324/9781351034784>
- Winke, P., Zhang, X., & Pierce, S. (2023). A closer look at a marginalized test method: Self-assessment as a measure of speaking proficiency. *Studies in Second Language Acquisition*, 45(2), 416-441. doi:10.1017/S0272263122000079
- Wolf, M. K. (2022). Interconnection between constructs and consequences: a key validity consideration in K–12 English language proficiency assessments. *Language Testing in Asia*, 12, Article 44. <https://doi.org/10.1186/s40468-022-00194-1>
- Wolf, M. K. (2024). Assessment of young language learners. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 312-325). Wiley-Blackwell.
- Wolf, M. K., & Lopez, A. (2022). Developing a technology-based classroom assessment of academic reading skills for English language learners and teachers: validity evidence for formative use. *Languages*, 7(2), Article 71.
<https://doi.org/10.3390/languages7020071>
- Wolf, M. K., & Oh, S. (2024). Investigating the types and use of feedback in middle-school English language learners’ academic writing. *Language Learning & Technology*, 28(1), 1-25. <https://hdl.handle.net/10125/73583>
- Wolf, M. K., Bailey, A. L., & Ballard, L. (2023). Aligning English language proficiency assessments to standards: Conceptual and technical issues. *TESOL Quarterly*, 57(2), 670-685. <https://doi.org/10.1002/tesq.3199>
- Wolf, M. K., Lopez, A. A., & Lee, J. (2023). An investigation of the use of standardized and local assessments for young EAL students. In G. Brooks, J. Clenton, & S. Fraser (Eds.), *EAL research for the classroom: Practical and pedagogical implications* (pp. 164-184). Routledge. <https://doi.org/10.4324/9781003274889-13>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). Exploring the provision of AI-based feedback for TOEFL Junior Writing practice tasks (Research Memorandum No. RM 24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>
- Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). Design framework for the TOEFL Primary® Writing test (Research Memorandum ETS RM-24-02). ETS.
<https://www.ets.org/Media/Research/pdf/RM-24-02.pdf>
- Wolf, M. K., Yoo, H., & Ballard, L. (2024). The relationship between English language proficiency and academic performance among English learners: A subgroup analysis by home language background. *Learning and Individual Differences*, 109, 102388. <https://doi.org/10.1016/j.lindif.2023.102388>
- Wolf, M. K., Yoo, H., Guzman-Orth, D., & Abedi, J. (2022). Investigating the effects of test accommodations with process data for English learners in a mathematics assessment. *Educational Assessment*, 27(1), 27-45.
<https://doi.org/10.1080/10627197.2021.1982693>

- Wood, C., Hoge, R., Schatschneider, C., & Castilla-Earls, A. (2021). Predictors of item accuracy on the Test de Vocabulario en Imagenes Peabody for Spanish-English speaking children in the United States. *International Journal of Bilingual Education and Bilingualism*, 24(8), 1178–1192.
<https://doi.org/10.1080/13670050.2018.1547266>
- Woods, K., Hashimoto, B., & Brown, E. K. (2023). A multi-measure approach for lexical diversity in writing assessments: Considerations in measurement and timing. *Assessing Writing*, 55, Article 100688. <https://doi.org/10.1016/j.asw.2022.100688>
- Wu, C., & Gao, F. (2023). Comparing the Intelligibility of Different Varieties of English Through Predicting Learners Comprehension: A Phonetic Experimental Approach. *Chinese Journal of Applied Linguistics*, 46(1), 19-30.
<https://doi.org/10.1515/CJAL-2023-0102>
- Wu, J., & Lu, X. (2024). Noun phrase complexity and second language Chinese proficiency: An analysis of narratives by Korean learners of Chinese. *Assessing Writing*, 59, 100810. <https://doi.org/10.1016/j.asw.2024.100810>
- Wu, M. Y., Steinkrauss, R., & Lowie, W. (2023). The reliability of single task assessment in longitudinal L2 writing research. *Journal of Second Language Writing*, 59, 100950. <https://doi.org/10.1016/j.jslw.2022.100950>
- Wu, P., & Lei, C. (2023). Developing feedback literacy through dialogue-supported performances of multi-draft writing in a postgraduate class. *Assessing Writing*, 57, Article 100759. <https://doi.org/10.1016/j.asw.2023.100759>
- Wu, P., Yu, S., & Luo, Y. (2023). The development of teacher feedback literacy in situ: EFL writing teachers' endeavor to human-computer-AWE integral feedback innovation. *Assessing Writing*, 57, Article 100739. <https://doi.org/10.1016/j.asw.2023.100739>
- Wu, S. L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, 44(1), 271-300. <https://doi.org/10.1017/S0272263121000103>
- Wu, X. M., Zhang, L. J., & Dixon, H. R. (2021). Implementing Assessment for Learning (AfL) in Chinese university EFL classes: Teachers' values and practices. *System*, 101, Article 102589. <https://doi.org/10.1016/j.system.2021.102589>
- Wu, Y.-j. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology*, 25(2), 192–214. <http://hdl.handle.net/10125/73440>
- Wudthayagorn, J. (2022). An exploration of the English Exit Examination policy in Thai public universities. *Language Assessment Quarterly*, 19(2), 107-123.
<https://doi.org/10.1080/15434303.2021.1937174>

X

- Xi, X. (2023). Advancing Language Assessment with AI and ML-Learning into AI is Inevitable, but Can Theory Keep Up? *Language Assessment Quarterly*, 20(4-5), 357-376. <https://doi.org/10.1080/15434303.2023.2291488>
- Xi, X., & Norris, J. M. (Eds.). (2021). *Assessing English for higher education admissions*. Routledge. <https://doi.org/10.4324/9781351142403>
- Xia, D., Chen, Y., & Pae, H. (2023). Lexical and grammatical collocations in beginning and intermediate L2 argumentative essays: a bigram study. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1421-1453.
<https://doi.org/10.1515/iral-2021-0188>

- Xia, T., Chen, X., Parsaei, H. R., & Qiu, F. (2023). An intelligent vocabulary size measurement method for second language learner. *Language Testing in Asia*, 13, Article 45 (2023). <https://doi.org/10.1186/s40468-023-00258-w>
- Xie, Q. (2022). An online database of English assessment tasks for developing student teachers' assessment literacy. *International Journal of Mobile Learning and Organization*, 16, Article 4. <https://doi.org/10.1504/IJMLO.2022.10048804>
- Xie, Q. (2024). Construct representation and predictive validity of integrated writing tasks: A study on the writing component of the Duolingo English Test. *Assessing Writing*, 61, 100846. <https://doi.org/10.1016/j.asw.2024.100846>
- Xie, Q. Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing*, 57, Article 100755. <https://doi.org/10.1016/j.asw.2023.100755>
- Xie, Q., & Lei, Y. (2022). Diagnostic assessment of L2 academic writing product, process and self-regulatory strategy use with a comparative dimension. *Language Assessment Quarterly*, 19(3), 231-263. <https://doi.org/10.1080/15434303.2021.1903470>
- Xu, J. (2022). Incremental intelligence matters: How L2 writing mindsets impact feedback orientation and self-regulated learning writing strategies. *Assessing Writing*, 51, Article 100593. <https://doi.org/10.1016/j.asw.2021.100593>
- Xu, J., & Fan, Y. (2024). Task complexity, L2 proficiency and EFL learners' L1 use in task-based peer interaction. *Language Teaching Research*, 28(2), 346-365. <https://doi.org/10.1177/13621688211004633>
- Xu, J., & Qiu, X. (2024). Engaging L2 Learners in Information-gap Tasks: How Task Type and Topic Familiarity Affect Learner Engagement. *RELC Journal*, 55(1), 29–45. <https://doi.org/10.1177/00336882211061628>
- Xu, J., & Wang, Y. (2023). Individual differences in L2 writing feedback-seeking behaviors: The predictive roles of various motivational constructs. *Assessing Writing*, 56, Article 100698. <https://doi.org/10.1016/j.asw.2023.100698>
- Xu, J., & Wang, Y. (2024). Do I need feedback or avoid it in L2 writing? Impacts of self-efficacy and shyness on feedback-seeking behavior. *Assessing Writing*, 59, 100805. <https://doi.org/10.1016/j.asw.2023.100805>
- Xu, J., & Zheng, Y. (2024). Understanding the interconnectedness of L2 learners' writing assessment literacy and engagement: A longitudinal survey. *System*, 123, 103318. <https://doi.org/10.1016/j.system.2024.103318>
- Xu, J., Jones, E., Laxton, V., & Galaczi, E. (2021). Assessing L2 English speaking using automated scoring technology: Examining automarker reliability. *Assessment in Education: Principles, Policy & Practice*, 28(4), 411–436. <https://doi.org/10.1080/0969594X.2021.1979467>
- Xu, J., Schmidt, E., Galaczi, E., & Somers, A. (2024). Automarking in language assessment: Key considerations for best practice. Cambridge Papers in English Language Education. Cambridge University Press & Assessment. <https://cambunipress.prod.acquia-sites.com/sites/default/files/media/documents/cambridge-research-papers-in-english-language-education.pdf>
- Xu, J., Zheng, Y., & Braund, H. (2023). Voices from L2 learners across different languages: Development and validation of a student writing assessment literacy scale. *Journal of Second Language Writing*, 60, 100993. <https://doi.org/10.1016/j.jslw.2023.100993>
- Xu, Q., Sawaki, Y., & Zhu, Y. (2024). Linguistic complexity measures of short narrative writings for modeling overall Chinese proficiency. *System*, 121, 103220.

- <https://doi.org/10.1016/j.system.2024.103220>
- Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2023). A multidimensional approach to assessing the effects of task complexity on L2 students' argumentative writing. *Assessing Writing*, 55, Article 100690. <https://doi.org/10.1016/j.asw.2022.100690>
- Xu, Y., Chen, J., & He, L. (2023). How Personal, Experiential, and Contextual Factors Mediate EFL Teachers' Conceptions of Assessment: A Narrative Study. *Chinese Journal of Applied Linguistics*, 46(2), 237-252. <https://doi.org/10.1515/CJAL-2023-0206>
- Xu, Z., Zhang, L., & Parr, J. (2023). Incorporating peer feedback in writing instruction: examining its effects on Chinese English-as-a-foreign-language (EFL) learners' writing performance. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1337-1364. <https://doi.org/10.1515/iral-2021-0078>

Y

- Yamashita, T. (2024). An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, 3(3), Article 100133. <https://doi.org/10.1016/j.rmal.2024.100133>
- Yan, Q. (2024). Exploring Chinese university EFL students' perceptions of formative assessment: A qualitative study. *System*, 125, 103391. <https://doi.org/10.1016/j.system.2024.103391>
- Yan, X., & Chuang, P.-L. (2023). How do raters learn to rate? Many-facet Rasch modeling of rater performance over the course of a rater certification program. *Language Testing*, 40(1), 153-179. <https://doi.org/10.1177/02655322221074913>
- Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219-246. <https://doi.org/10.1177/0265532220929924>
- Yan, X., & Fan, J. (2022). Reliability and dependability. In Fulcher, G., & Harding, L. (Eds.), *The Routledge Handbook of Language Testing* (pp. 477-494). Routledge.
- Yan, X., & Lin, J. (2023). Chinese character matters!: An examination of linguistic accuracy in writing performances on the HSK test. *Assessing Writing*, 57, Article 100767. <https://doi.org/10.1016/j.asw.2023.100767>
- Yan, X., & Staples, S. (2023). *Investigating the cognitive and social aspects of IELTS speaking performances across proficiency levels: Comparing the CAF-based and register-linguistic analyses*. British Council, Cambridge Assessment English, and IDP: IELTS Australia. <https://search.informit.org/doi/epdf/10.3316/informit.262843971785611>
- Yan, X., Chuang, P.-L., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment*, 13(1), 108-140. <https://doi.org/10.58379/DLDI3148>
- Yan, X., Chuang, P., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment (SILA)*, 13(1), 108-140. <https://doi.org/10.58379/DLDI3148>
- Yan, X., Dimova, S., & Ginther, A. (Eds.). (2023). *Local language testing: Practice across contexts*. Springer.

- Yan, X., Kim, H. R., & Kim, J. Y. (2021). Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency. (CAF). features of speaking performances on the Aptis test. *Language Testing*, 38(4), 485–510.
<https://doi.org/10.1177/0265532220951508>
- Yan, Z., King, R. B., & Haw, J. Y. (2021). Formative assessment, growth mindset, and achievement: Examining their relations in the East and the West. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 676–702.
<https://doi.org/10.1080/0969594X.2021.1988510>
- Yang, C. (2023). Using eye-tracking and retrospective verbal reports to explore the cognitive processes of banked gap-filling: a case study featuring methodological triangulation. *Language Testing in Asia*, 13, Article 22 (2023).
<https://doi.org/10.1186/s40468-023-00234-4>
- Yang, R., & May, L. (2023). Understanding and Implementing Learning-Oriented Assessment: Evidence From Three Chinese Junior High School EFL Teachers. *Chinese Journal of Applied Linguistics*, 46(2), 308-325.
<https://doi.org/10.1515/CJAL-2023-0210>
- Yang, S. (2021). Diagnosing reading problems for low-level Chinese as second language learners. *System*, 97, Article 102433. <https://doi.org/10.1016/j.system.2020.102433>
- Yang, X., & Osborne, C. (2023). The development and validation of a C-test and a pseudo-character test for online CFL assessment. *Journal of China Computer-Assisted Language Learning*, 3(1), 101-131.
<https://doi.org/10.1515/jccall-2022-0019>
- Yang, Y. (2023). The Correlation Between English and French Proficiency of Chinese Trilingual Learners. *Chinese Journal of Applied Linguistics*, 46(3), 404-423.
<https://doi.org/10.1515/CJAL-2023-0305>
- Yang, Y., & Qian, D. D. (2023). Enhancing EFL learners' reading proficiency through dynamic assessment. *Language Assessment Quarterly*, 20(1), 20-43.
<https://doi.org/10.1080/15434303.2022.2132160>
- Yang, Y., Yap, N. T., & Ali, M. A. (2023). Predicting EFL expository writing quality with measures of lexical richness. *Assessing Writing*, 57, Article 100762.
<https://doi.org/10.1016/j.asw.2023.100762>
- Yang, Z., & Yu, C.* (2024). Is the use of multiple-choice items and a holistically-scored paragraph translation task fair? Examining a large-scale translation subtest. *Asia Pacific Education Review*. Online First. <https://doi.org/10.1007/s12564-024-09993-y>
- Yao, Y., Guo, N. S., Li, C., & McCampbell, D. (2021). How university EFL writers' beliefs in writing ability impact their perceptions of peer assessment: Perspectives from implicit theories of intelligence. *Assessment and Evaluation in Higher Education*, 46(1), 151-167. <https://doi.org/10.1080/02602938.2020.1750559>
- Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, 50, Article 100568.
<https://doi.org/10.1016/j.asw.2021.100568>
- Yao, Y., Zhu, X., Zhu, S., & Jiang, Y. (2023). The impacts of self-efficacy on undergraduate students' perceived task value and task performance of L1 Chinese integrated writing: A mixed-method research. *Assessing Writing*, 55, Article 100687.
<https://doi.org/10.1016/j.asw.2022.100687>
- Yao, Y., Zhu, X., & Zhan, J. (2024). The associations among growth mindsets, the ideal L2

- writing self, and L2 writing enjoyment and their impacts on L2 English writing performance: A gender difference perspective. *Assessing Writing*, 60, 100832. <https://doi.org/10.1016/j.asw.2024.100832>
- Yasuda, S. (2024). Does “more complexity” equal “better writing”? Investigating the relationship between form-based complexity and meaning-based complexity in high school EFL learners’ argumentative writing. *Assessing Writing*, 61, 100867. <https://doi.org/10.1016/j.asw.2024.100867>
- Yeager, R. R., Park, G., & Liao, R. J. T. (2024). Notetaking as Validity Evidence: A Mixed-Methods Investigation of Question Preview in Listening Assessment. *Journal of English for Academic Purposes*, 68(3), 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Yeager, R., Park, G., & Liao, R. J. (2024). Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment. *Journal of English for Academic Purposes*, 68, Article 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Yeldham, M., & Chen, R. T.-H. (2024). Conducting L2 listener verbal reports: Common missteps by novice researchers and ways to address them. *Research Methods in Applied Linguistics*, 3(2), Article 100113. <https://doi.org/10.1016/j.rmal.2024.100113>
- Yen, P. H., & Thao, L. T. (2024). Exploring the implementation and perception of competency-based assessment practices among Vietnamese EFL instructors. *Language Testing in Asia*, 14, Article 26. <https://doi.org/10.1186/s40468-024-00300-5>
- Yenkimaleki, M., van Heuven, V. J., & Hosseini, M. (2024). The effect of providing feedback and feedforward in prosody instruction for developing listening comprehension skills by interpreter trainees. *Foreign Language Annals*, 57(1), 184-200. <https://doi.org/10.1111/flan.12725>
- Yeo, M. (2023). Peer Review Circles: The Value of Working in Triadic Groups and Oral Observational Peer Feedback in the Development of Student Feedback Literacy. *Chinese Journal of Applied Linguistics*, 46(2), 270-288. <https://doi.org/10.1515/CJAL-2023-0208>
- Yeom, S., & Llosa, L. (2024). Comparability of reading tasks in high-stakes English proficiency tests and university courses in Korea. *Language Testing in Asia*, 14, Article 8. <https://doi.org/10.1186/s40468-024-00281-5>
- Yi, Y.-S. (2021). On the usefulness of CDA-based score reporting: Implications for self-regulated learning. *Language Testing in Asia*, 11, Article 13. <https://doi.org/10.1186/s40468-021-00127-4>
- Yigiter, M. S., & Dogan, N. (2023). Computerized multistage testing: Principles, designs and practices with R. *Measurement: Interdisciplinary Research and Perspectives*, 21(4), 254–277, <https://doi.org/10.1080/15366367.2022.2158017>
- Yilmaz, Y., Arroyo, D., Carver, C., Choi, J., & Dibartolomeo, M. (2023). The effects of distributed versus massed corrective feedback on the acquisition of Spanish differential object marking. *Modern Language Journal*, 107, 756–781. <https://doi.org/10.1111/modl.12862>
- Yin, S., Fan, J., Jin, Y., & Stapleton, P. (2024). Towards a framework of critical thinking for assessing EAP speaking. *Journal of English for Academic Purposes*, 1-14. <https://doi.org/10.1016/j.jeap.2024.101426>
- Yin, S., Fan, J., Jin, Y., & Stapleton, P. (2024). Towards a framework of critical thinking for assessing EAP speaking. *Journal of English for Academic Purposes*, 71, Article No.

101426. <https://doi.org/10.1016/j.jeap.2024.101426>
- Yin, S., Guo, K., & Li, D. (2024). To assign a topic or not: Examining critical thinking in English as a foreign language speaking. *RELC Journal*, 1-14. <https://doi.org/10.1177/00336882231222651>
- Yin, Y., Yu, G., & Huang, D. (2024). Implementing the ICAO language proficiency requirements: The localization of aviation English testing. *Applied Linguistics*, 45(4), 699-717. <https://doi.org/10.1093/applin/amad057>
- Yongqi GU, P. & LAM, R. (2023). Developing Assessment Literacy for Classroom-Based Formative Assessment. *Chinese Journal of Applied Linguistics*, 46(2), 155-161. <https://doi.org/10.1515/CJAL-2023-0201>
- Yu-Ting, K. (2023). Empowering preservice English teachers with language assessment literacy concepts and practices: Application of Vygotskian concept-based language instruction. *Modern Language Journal*, 107, 68-94. <https://doi.org/10.1111/modl.12818>
- Yu, C. (2023). Integrating information literacy and academic writing: Developing a self-assessment scale of information-based academic writing. *The Journal of Academic Librarianship*, 47(6), 102804. <https://doi.org/10.1016/j.acalib.2023.102804>
- Yu, C. (2023). Testing issues in writing. In H. Mohebbi, & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 56-70). Castledown. <https://doi.org/10.29140/9781914291159-5>
- Yu, C., & Xu, W. (2024). Writing assessment literacy and its impact on the learning of writing: A netnography focusing on Duolingo English Test examinees. *Language Testing in Asia*, 14, Article 24. <https://doi.org/10.1186/s40468-024-00297-x>
- Yu, G. & Xu, J. (Eds.). (2024). *Language test validation in a digital age*. Studies in Language Testing Volume 52. Cambridge University Press & Assessment.
- Yu, G., & Green, A. (2021). Preparing for admissions tests in English. *Assessment in Education: Principles, Policy & Practice*, 28(1), 1-12. <https://doi.org/10.1080/0969594X.2021.1880120>
- Yu, G., & Xu, J. (Eds.). (2024). *Language test validation in a digital age*. Cambridge University Press & Assessment.
- Yu, L., & Poehner, M. E. (2023). Developing L2 writing abilities through an instructional enrichment program informed by dynamic assessment. *Journal of Second Language Writing*, 60, 101013. <https://doi.org/10.1016/j.jslw.2023.101013>
- Yu, S. (2021). Giving genre-based peer feedback in academic writing: sources of knowledge and skills, difficulties and challenges. *Assessment and Evaluation in Higher Education*, 46(1), 36-53. <https://doi.org/10.1080/02602938.2020.1742872>
- Yu, S., & Jiang, L. (2023). L2 university students' motivational self system in English writing: a sociocultural inquiry. *Applied Linguistics Review*, 14(3), 553-578. <https://doi.org/10.1515/applirev-2020-0110>
- Yu, S., & Lee, I. (2024). Writing assessment and feedback literacy: Where do we stand and where can we go? *Assessing Writing*, 60, 100829. <https://doi.org/10.1016/j.asw.2024.100829>
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48, Article 100525. <https://doi.org/10.1016/j.asw.2021.100525>
- Yu, S., Geng, F., Liu, C., & Zheng, Y. (2021). What works may hurt: The negative side of feedback in second language writing. *Journal of Second Language Writing*, 54,

- Article 100850. <https://doi.org/10.1016/j.jslw.2021.100850>
- Yu, S., Zhang, E. D., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing*, 53, Article 100643. <https://doi.org/10.1016/j.asw.2022.100643>
- Yu, S., Zheng, Y., Jiang, L., Liu, C., & Xu, Y. (2021). "I even feel annoyed and angry": Teacher emotional experiences in giving feedback on student writing. *Assessing Writing*, 48, Article 100528. <https://doi.org/10.1016/j.asw.2021.100528>
- Yu, W., & Iwashita, N. (2021). Comparison of test performance on paper-based testing. (PBT). and computer-based testing. (CBT). by English-majored undergraduate students in China. *Language Testing in Asia*, 11, Article 32. <https://doi.org/10.1186/s40468-021-00147-0>
- Yuan, Y., Engelhard, G., Raczynski, K., & Wang, J. (2022). Using unfolding models to identify targeted feedback strategies for student writing. *Assessing Writing*, 54, Article 100670. <https://doi.org/10.1016/j.asw.2022.100670>
- Yuksel, D., Soruç, A., Altay, M., & Curle, S. (2023). A longitudinal study at an English medium instruction university in Turkey: the interplay between English language improvement and academic success. *Applied Linguistics Review*, 14(3), 533-552. <https://doi.org/10.1515/applirev-2020-0097>
- Yumsek, M. (2023). Educational L2 constructs and diagnostic measurement. *Language Testing in Asia*, 13, Article 3 (2023). <https://doi.org/10.1186/s40468-022-00214-0>
- Yung, K. W.-H. (2023). Engaging exam-oriented students in communicative language teaching by 'packaging' learning English through songs as exam practice. *RELC Journal*, 54(1), 280-290. <https://doi.org/10.1177/0033688220978542>

Z

- Zabala Delgado, J., & Rodriguez, C. (2022). Language assessment literacy of language teachers in the context of adult education in Spain. *Studies in Language Assessment*, 11(1), 64–91. https://arts.unimelb.edu.au/__data/assets/pdf_file/0009/4282560/SiLA-11.1-Delgado_and_Rodriguez.pdf
- Zabihi, R., Ghominejad, S., & Javad Ahmadian, M. (2024). Can willingness to communicate, communication in English anxiety, behavioural inhibition and behavioural action predict perceived L2 fluency? *Language Teaching Research*, 28(6), 2214-2233. <https://doi.org/10.1177/13621688211044071>
- Zadkhast, M., Rezvani, E., & Lotfi, A. R. (2023). Effects of concurrent and cumulative group dynamic assessments on EFL learners' development of reading comprehension micro-skills. *Language Testing in Asia*, 13, Article 29 (2023). <https://doi.org/10.1186/s40468-023-00240-6>
- Zakaria, A., & Aryadoust, V. (2024). From colonialism to self-determination: A historical study of language testing and institutional policy. *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amae057>
- Zarrinabadi, N., Mahmoudi-Gahrouei, V., & Mohammadabadi, A. M. (2023). Your writing could have been better: Examining the effects of upward and downward counterfactual communication on the motivational aspects of L2 writing. *Assessing Writing*, 56, Article 100714. <https://doi.org/10.1016/j.asw.2023.100714>
- Zeng, X., Liu, J., & Bowen, N. E. J. A. (2024). A comparative study of voice in Chinese

- English-major undergraduates' timed and untimed argument writing. *Assessing Writing*, 62, 100896. <https://doi.org/10.1016/j.asw.2024.100896>
- Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47, Article 100505. <https://doi.org/10.1016/j.asw.2020.100505>
- Zesch, T., Horbach, A., & Zehner, F. (2023). To score or not to score: Factors influencing performance and feasibility of automatic content scoring of text responses. *Educational Measurement, Issues and Practice*, 42(1), 44–58. <https://doi.org/10.1111/emip.12544>
- Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, 35(9), 2817-2842. <https://doi.org/10.1080/09588221.2021.1897019>
- Zhai, X., & Aryadoust, V. (2022). The metacognitive and neurocognitive signature of test methods in academic listening assessment. *Frontiers in Psychology*, 13. Article 930075. <https://doi.org/10.3389/fpsyg.2022.930075>
- Zhan, Y., Lee, J. C. K., & Wan, Z. H. (2022). Who perceives more value of teacher feedback? Exploring the roles of college students' possible second language selves and language learning strategies. *Language Testing in Asia*, 12, Article 62. <https://doi.org/10.1186/s40468-022-00212-2>
- Zhang-Wu, Q. (2023). Exploring multilingual students' feedback literacy in an asynchronous online writing course. *Assessing Writing*, 56, Article 100718. <https://doi.org/10.1016/j.asw.2023.100718>
- Zhang, C., Xu, X., & Zhang, Y. (2023). A communicative validation study on an English listening test in Korea. *Language Testing in Asia*, 13, Article 26 (2023). <https://doi.org/10.1186/s40468-023-00238-0>
- Zhang, E. D., & Yu, S. (2024). Understanding L2 student writers' self-assessment in digital multimodal composing: A process-oriented approach. *System*, 121, 103219. <https://doi.org/10.1016/j.system.2024.103219>
- Zhang, F., Min, H., He, P., Chen, S., & Ren, S. (2023). Understanding EFL students' feedback literacy development in academic writing: A longitudinal case study. *Assessing writing*, 58, Article 100770. <https://doi.org/10.1016/j.asw.2023.100770>
- Zhang, H. (2024). The washback of the National Matriculation English Test on senior high school English learning outcomes: Do test takers from different provinces think alike? *Language Testing in Asia*, 14, Article 14. <https://doi.org/10.1186/s40468-024-00286-0>
- Zhang, H., Sun, P., Bianglae, Y., & Widiawati, W. (2024). The development of a Chinese vocabulary proficiency test (CVPT) for learners of Chinese as a second/foreign language. *Language Testing*, 41(2), 412–442. <https://doi.org/10.1177/02655322231219998>
- Zhang, J. (2021). A moderated mediation analysis of the relationship between a high-stakes English test and test takers' extracurricular English learning activities. *Language Testing in Asia*, 11, Article 5. <https://doi.org/10.1186/s40468-021-00120-x>
- Zhang, L. (2022). College English Test–Spoken English Test (CET–SET). *Studies in Language Assessment*, 11(2), 164-180. https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4396265/SiLA_11.2_Test-review.pdf
- Zhang, L. J., & Zhang, J. (2024). EFL students' syntactic complexity development in argumentative writing: A latent class growth analysis (LCGA) approach. *Assessing*

- Writing*, 61, 100877. <https://doi.org/10.1016/j.asw.2024.100877>
- Zhang, L., & Jin, Y. (2021). Assessing interactional competence in the computer-based CET-SET: An investigation of the use of communication strategies. *Assessment in Education: Principles, Policy & Practice*, 28(4), 389–410. <https://doi.org/10.1080/0969594X.2021.1976107>
- Zhang, L., & Zhang, L. (2023). Development and validation of the questionnaire on EFL students' perceptions of authorial stance in academic writing. *Applied Linguistics Review*, 14(5), 1219–1251. <https://doi.org/10.1515/applirev-2021-0003>
- Zhang, L., Zheng, Y., O'Sullivan, B., & Stanley, G. (2023). Teacher perceptions of learner engagement with a new online Learning environment. In H. Işıl Işık, E. Kantarcıoğlu, A. Üner & C. Thomas (Eds.), *Re-vision and re-form in English for academic purposes contexts after the Pandemic* (pp. 124 –132). Cambridge Scholars.
- Zhang, M., & Chen, W. (2022). Assessing collaborative writing in the digital age: An exploratory study. *Journal of Second Language Writing*, 57, Article 100868. <https://doi.org/10.1016/j.jslw.2022.100868>
- Zhang, M., & Sinharay, S. (2022). Investigating the writing performance of educationally at-risk examinees using technology. *International Journal of Testing*, 22(3-4), 312-347. <https://doi.org/10.1080/15305058.2022.2050734>
- Zhang, P., Fan, J., & Jia, W. (2021). A bibliographic analysis of hotspots and trends in international language testing research. (2008-2018). *Foreign Language Teaching and Research*, 53(4), 618–627.
- Zhang, P., Fan, J., & Jia, W. (2022). Assessing L2 integrated writing self-efficacy: Development and validation of a scale. *Assessing Writing*, 54, Article 100665. <https://doi.org/10.1016/j.asw.2022.100665>
- Zhang, S., & Isaacs, T. (2023). Can interactions happen across the screens? The use of videoconferencing technology in assessing second language pragmatic competence. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts* (pp. 196–211). Routledge.
- Zhang, T., & Mao, Z. (2023). Exploring the development of student feedback literacy in the second language writing classroom. *Assessing Writing*, 55, Article 100697. <https://doi.org/10.1016/j.asw.2023.100697>
- Zhang, T., & Weng, Z. (2024). Implementing rubric co-construction in ESL writing teaching. *ELT Journal*, 78(1), 32-41. <https://doi.org/10.1093/elt/ccad033>
- Zhang, W., & Zhang, L. J. (2022). Understanding assessment tasks: Learners' and teachers' perceptions of cognitive load of integrated speaking tasks for TBLT implementation. *System*, 111, Article 102951. <https://doi.org/10.1016/j.system.2022.102951>
- Zhang, W., Zhao, M., & Zhu, Y. (2022). Understanding individual differences in metacognitive strategy use, task demand, and performance in integrated L2 speaking assessment tasks. *Frontiers in Psychology*, 13, Article 876208. <https://doi.org/10.3389/fpsyg.2022.876208>
- Zhang, X. (2021). Assessing EFL students' writing development as they are exposed to the integrated use of drama-based pedagogy and SFL-based teaching. *Assessing Writing*, 50, Article 100569. <https://doi.org/10.1016/j.asw.2021.100569>
- Zhang, X. (2022). The relationship between lexical use and L2 writing quality: A case of two genres. *International Journal of Applied Linguistics*, 32(3), 371-396.

- <https://doi.org/10.1111/ijal.12420>
- Zhang, X., & Li, W. (2021). Effects of n-grams on the rated L2 writing quality of expository essays: A conceptual replication and extension. *System*, 97, Article 102437. <https://doi.org/10.1016/j.system.2020.102437>
- Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, 51, Article 100597. <https://doi.org/10.1016/j.asw.2021.100597>
- Zhang, X., Zhao, B., & Li, W. (2023). N-gram use in EFL learners' retelling and monologic tasks. *International Review of Applied Linguistics in Language Teaching*, 61(3), 939–965. <https://doi.org/10.1515/iral-2021-0080>
- Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. *Educational Science*, 12(10), Article 723. <https://doi.org/10.3390/educsci12100723>
- Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing*, 60, 100842. <https://doi.org/10.1016/j.asw.2024.100842>
- Zhang, Y., & MacWhinney, B. (2023). Using diagnostic feedback to enhance the development of phonetic knowledge of an L2: A CALL design based on the unified competition model and the implementation with the Pinyin Tutor. *Language Testing in Asia*, 13, Article 35 (2023). <https://doi.org/10.1186/s40468-023-00232-6>
- Zhang, Y., & Ouyang, J. (2023). Linguistic complexity as the predictor of EFL independent and integrated writing quality. *Assessing Writing*, 56, Article 100727. <https://doi.org/10.1016/j.asw.2023.100727>
- Zhang, Y., & Papi, M. (2024). Regulatory focus predicts individual differences in pragmatic versus grammatical awareness and sensitivity. *Studies in Second Language Learning and Teaching*, 14(2), 267–290. <https://doi.org/10.14746/ssllt.35073>
- Zhang, Y., & Wu, W. (2021). How effective are lexical richness measures for differentiations of vocabulary proficiency? A comprehensive examination with clustering analysis. *Language Testing in Asia*, 11, Article 15. <https://doi.org/10.1186/s40468-021-00133-6>
- Zhang, Y., & Xi, J. (2023). Fostering self-regulated young writers: Dynamic assessment of metacognitive competence in secondary school EFL class. *Language Assessment Quarterly*, 20(1), 88–107. <https://doi.org/10.1080/15434303.2022.2103702>
- Zhang, Z. V., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, 58, Article 100784. <https://doi.org/10.1016/j.asw.2023.100784>
- Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, Article 100586. <https://doi.org/10.1016/j.asw.2021.100586>
- Zhao, C. G., & Qi, Q. (2023). Implementing learning-oriented assessment (LOA) among limited-proficiency EFL students: Challenges, strategies, and students' reactions. *TESOL Quarterly*, 57(2), 566–594. <https://doi.org/10.1002/tesq.3167>
- Zhao, C. G., & Wu, J. (2022). Perceptions of authorial voice: Why discrepancies exist. *Assessing Writing*, 53, Article 100632. <https://doi.org/10.1016/j.asw.2022.100632>
- Zhao, H., & Aryadoust, V. (2024). A meta-analysis of the reliability of second language reading comprehension assessment tools. *Studies in Second Language Acquisition*, 1–29. doi:10.1017/S0272263124000627

- Zhao, H., & Zhao, B. (2023). Co-constructing the assessment criteria for EFL writing by instructors and students: A participative approach to constructively aligning the CEFR, curricula, teaching and learning. *Language Teaching Research*, 27(3), 765–793. <https://doi.org/10.1177/1362168820948458>
- Zhao, J.-H., Chen, Z.-W., & Yang, Q.-F. (2024). I do and I understand: A virtual reality-supported collaborative design-assessing activity for EFL students. *System*, 121, 103213. <https://doi.org/10.1016/j.system.2023.103213>
- Zhao, R., Coss, M. D., Ruan, H., Li, B., & Ma, J. (2023). Examining secondary–postsecondary articulation of Chinese language programs: A survey of US college placement procedures. *Foreign Language Annals*, 56(3), 690–719.
- Zhao, Y., & Aryadoust, V. (2024). An automatized semantic analysis of two large-scale listening tests: A corpus-based study. *Language Testing*, 0(0). <https://doi.org/10.1177/02655322241288598>
- Zheng, C., Chen, X., Zhang, H., & Chai, C. S. (2024). Automated versus peer assessment: Effects on learners' English public speaking. *Language Learning & Technology*, 28(2), 210–228. <https://hdl.handle.net/10125/73577>
- Zheng, C., Wang., & Chai, C. S. (2023). Self-assessment first of peer-assessment first: effects of video-based formative practice on learners' English public speaking anxiety and performance. *Computer Assisted Language Learning*, 36(4), 806–839. <https://doi.org/10.1080/09588221.2021.1946562>
- Zhi, M., & Huang, B. (2021). Investigating the authenticity of computer-and paper-based ESL writing tests. *Assessing Writing*, 50, Article 100548. <https://doi.org/10.1016/j.asw.2021.100548>
- Zhou, J., Deneen, C., Tai, J., & Dawson, P. (2023). Feedback seeking by first-year Chinese international students: Understanding practices and challenges. *Assessing Writing*, 57, Article 100757. <https://doi.org/10.1016/j.asw.2023.100757>
- Zhou, Q., Chen, L., & Hou, X. (2022). Exploring Chinese EFL undergraduates' writing from sources: Self-efficacy and performance. *Assessing Writing*, 54, Article 100663. <https://doi.org/10.1016/j.asw.2022.100663>
- Zhou, S., & Thompson, G. (2023). A longitudinal study on students' self-regulated listening during transition to an English-medium transnational University in China. *Studies in Second Language Learning and Teaching*, 13(2), 427–450. <https://doi.org/10.14746/ssllt.38281>
- Zhu, M., Zhang, M., & Gu, L. (2023). Insights into editing and revising in writing process using keystroke logs. *Language Assessment Quarterly*, 20(4-5), 445–468. <https://doi.org/10.1080/15434303.2023.2291478>
- Zhu, X., & Aryadoust, V. (2022). An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test. *Computer Assisted Language Learning*, 35(3), 412–436. <https://doi.org/10.1080/09588221.2019.1704788>
- Zhu, X., Guan, Y., & Yao, Y. (2022). Transfer of ideal L1 and L2 writing selves and their impacts on L2 writing enjoyment and integrated writing performance. *Assessing Writing*, 54, Article 100674. <https://doi.org/10.1016/j.asw.2022.100674>
- Zhu, X., Li, G. Y., Cheong, C. M., & Wen, H. (2021). Effects of L1 single-text and multiple-text comprehension on L2 integrated writing. *Assessing Writing*, 49, Article 100546. <https://doi.org/10.1016/j.asw.2021.100546>

- Zimmermann, S., Schwarz, L., Peters, A. & Depner, G. (2024). Enhancing teachers and test takers' assessment literacy? Insights from test preparation for the digital TestDaF. In: B. Baker & L. Taylor (Eds.), *Language Assessment Literacy and Competence Volume 1: Research and Reflections from the Field* (pp. 110–128). Cambridge: Cambridge University Press.
- Zou, S., Yan, X., & Fan, J. (2024). Establishing analytic score profiles for large-scale L2 writing assessment: The case of the CET-4 writing test. *Assessing Writing*, 60, 100826. <https://doi.org/10.1016/j.asw.2024.100826>

Bibliography by Category

Assessing Specific Competences and Populations

Assessing Reading

- Akgün, H. & Ünalı, A. (2022). Investigating cognitive processes in different item formats in reading tests through eye-tracking and verbal protocols. *Studies in Language Assessment*, 11(2), 98–129. <https://doi.org/10.58379/UIXK2220>
- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Amjadi, A., & Talebi, S. H. (2024). Extending the efficacy of Collaborative Strategic Reading (CSR) in teaching reading strategies: A socio-cultural emotional perspective. *Language Teaching Research*, 28(2), 714–748. <https://doi.org/10.1177/13621688211011501>
- An, N., Wright, C., & Wang, J. (2024). How does background knowledge affect second language reading? An eye movement study. *International Journal of Applied Linguistics*. 34(4), 1767–1789. <https://doi.org/10.1111/ijal.12601>
- Arai, Y. (2023). Comparability between graded readers and the Common Test in Japan in terms of text difficulty perceived by learners. *Language Assessment Quarterly*, 20(3), 277-295. <https://doi.org/10.1080/15434303.2023.2237497>
- Ardasheva, Y., Crosson, A. C., Carbonneau, K. J., & French, B. F. (2021). Unpacking contributions of morphosyntactic awareness and vocabulary to science reading comprehension among linguistically diverse students. *TESOL Quarterly*, 55(3), 931-965. <https://doi.org/10.1002/tesq.3039>
- Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners' reading comprehension: Exploring the effects of working memory, grade level, and reading task. *Language Testing*, 38(3), 356–377. <https://doi.org/10.1177/0265532221991480>
- Bulut, H. C., Bulut, O., & Arikan, S. (2023). Evaluating group differences in online reading comprehension: The impact of item properties. *International Journal of Testing*, 23(1), 10-33. <https://doi.org/10.1080/15305058.2022.2044821>
- Cai, Y., & Chen, H. (2022). The fluctuating effect of thinking on language performance: New evidence for the Island Ridge Curve. *Language Assessment Quarterly*, 19(5), 465–479. <https://doi.org/10.1080/15434303.2022.2080553>
- Chen, H., Cai, Y., & de la Torre, J. (2023). Investigating second language (L2) reading subskill associations: A cognitive diagnosis approach. *Language Assessment Quarterly*, 20(2), 166-189. <https://doi.org/10.1080/15434303.2022.2140050>
- Chinpakdee, M., & Gu, P. Y. (2024). The impact of explicit strategy instruction on EFL secondary school learners' reading. *Language Teaching Research*, 28(1), 296–319. <https://doi.org/10.1177/1362168821994157>
- DE JONG, J. H. A. L. (2023). Reading: A precondition for Writing. In R. Horowitz (Ed.), *The Routledge international handbook of research on writing* (2nd. ed., pp. 556–562). Routledge.

- Elosua, P. (2024). A Three-Step DIF Analysis of a Reading Comprehension Test Across Regional Dialects to Improve Test Score Validity. *Language Assessment Quarterly*, 21(2), 141–158. <https://doi.org/10.1080/15434303.2024.2307621>
- Estrada-Araoz, E. G., Sayed, B. T., Niyazova, G. G., & Lami, D. (2023). Comparing the effects of computerized formative assessment vs. computerized dynamic assessment on developing EFL learners' reading motivation, reading self-concept, autonomy, and self-regulation. *Language Testing in Asia*, 13, Article 39 (2023). <https://doi.org/10.1186/s40468-023-00253-1>
- Fjørtoft, H., Lai, M. K., & Li, M. (2024). How curriculum and assessment policies affect the role of reading in an assessment culture: a New Zealand case study. *Assessment in Education: Principles, Policy & Practice*, 31(2), 1-31. <https://doi.org/10.1080/0969594X.2024.2365645>
- George, A. C., & Robitzsch, A. (2021). Validating theoretical assumptions about reading with cognitive diagnosis models. *International Journal of Testing*, 21(2), 105–129. <https://doi.org/10.1080/15305058.2021.1931238>
- Geramipour, M. (2021). Rasch testlet model and bifactor analysis: How do they assess the dimensionality of large-scale Iranian EFL reading comprehension tests? *Language Testing in Asia*, 11, Article 3. <https://doi.org/10.1186/s40468-021-00118-5>
- Gesuato, S., Pagliarini, E., & Sanfelici, E. (2023). YOUNG ADULTS' INFERENTIAL READING SKILLS - RECOGNISING IMPLICIT AND UNSTATED CONTENT IN ITALIAN L1 AND ENGLISH L2. *Italiano LinguaDue*, 15(2), 20–20. <https://doi.org/10.54103/2037-3597/22011>
- Gui, M., Cao, X., & Zhao, J. (2023). EAP reading ability development during the process of generating research questions: A CDST perspective. *English for Academic Purposes*, 62, Article 101249. <https://doi.org/10.1016/j.jeap.2023.101249>
- Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11, Article 20. <https://doi.org/10.1186/s40468-021-00131-8>
- Haghi, I., & Tuersley, I. (2022). An investigation into how various textual features of academic journals can interact with students' engagement with these texts: A case of novice and non-traditional learners in STEM subjects (Report 3/2022). British Council and MetaMetrics. https://www.britishcouncil.org/sites/default/files/rirg_tuersley_and_haghi_layout_22.pdf
- Hamada, A., & Takaki, S. (2021). Approximate replication of Matsuda and Gobel (2004) for psychometric validation of the Foreign Language Reading Anxiety Scale. *Language Teaching*, 54(4), 535–551. <http://doi.org/10.1017/S0261444819000296>
- He, L., Xiong, L., & Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: Performance stability and variability across modalities and performance levels. *System*, 106, Article 102787. <https://doi.org/10.1016/j.system.2022.102787>
- Heeren, J., Speelman, D., & Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458–1473. <https://doi.org/10.1080/13670050.2019.1709411>
- Heydarnejad, T., Tagavipour, F., Patra, I., & Farid Khafaga, A. (2022). The impacts of performance-based assessment on reading comprehension achievement, academic

- motivation, foreign language anxiety, and students' self-efficacy. *Language Testing in Asia*, 12, Article 51. <https://doi.org/10.1186/s40468-022-00202-4>
- Ho, P. J. (2022). Assessing the range of cognitive processes in the Hong Kong Diploma of Secondary Education Examination (HKDSE)'s English language reading literacy test. *Language Testing in Asia*, 12, Article 18. <https://doi.org/10.1186/s40468-022-00167-4>
- Ho, S.K., & Gan, Z. (2023). Instructional practices and students' reading performance: a comparative study of 10 top performing regions in PISA 2018. *Language Testing in Asia*, 13, Article 48 (2023). <https://doi.org/10.1186/s40468-023-00261-1>
- Holzknicht, F., Guggenbichler, E., Zehentner, M., Yoder, M., Konrad, E., & Kremmel, B. (2022). Comparing EMI university reading materials with students' reading proficiency utilizing Lexile® measures, Aptis test results, student questionnaire responses, and interviews with lecturers (Report 2/2022). British Council and MetaMetrics. https://www.britishcouncil.org/sites/default/files/rirg_holzknicht_et_al_layout.pdf
- Hsieh, C.-N. (2023). *Evaluating the use and interpretations of the TOEIC Listening and Reading test score report: Perspectives of test takers in Japan*. ETS Research Report Series (RR-23-02). <https://doi.org/10.1002/ets2.12364>
- Hsieh, C.-N. (2023). *Interpretation and use of a workplace English language proficiency test score report: Perspectives of test takers and score users in Taiwan*. ETS Research Report Series (RR-23-10). <https://doi.org/10.1002/ets2.12373>
- In'nami, Y., Hijikata, Y., & Koizumi, R. (2022). Working memory capacity and L2 reading: A meta-analysis. *Studies in Second Language Acquisition*, 44(2), 381-406. <https://doi.org/10.1017/S0272263121000267>
- Ito, S. (2021). The effects of retelling on reading comprehension: Focusing on different levels of comprehension and non-textual information in retelling protocols. *Japanese Language Testing Association Journal*, 24, 23-42. https://doi.org/10.20622/jltajournal.24.0_23
- Kahn-Horwitz, J., & Goldstein, Z. (2024). English foreign language reading and spelling diagnostic assessments informing teaching and learning of young learners. *Language Testing*, 41(1), 60–88. <https://doi.org/10.1177/02655322231162838>
- Kao, Y.-T. (2022). Effects of group dynamic assessment on L2 Chinese learners' literacy development: learners' responsiveness to interactive mediation. *Applied Linguistics Review*, 13(5), 843–871. <https://doi.org/10.1515/applirev-2019-0077>
- Kara, Y., Kamata, A., Qiao, X., Potgieter, C. J., & Nese, J. F. (2024). Equating oral reading fluency scores: A model-based approach. *Educational and Psychological Measurement*, 84(1), 190-209. <https://doi.org/10.1177/00131644221148122>
- Kaya, E., O'Grady, S., & Kalender, İ. (2022). IRT-based classification analysis of an English language reading proficiency subtest. *Language Testing*, 39(4), 541–566. <https://doi.org/10.1177/02655322211068847>
- Kim, A. A., Tywoniw, R. L., & Chapman, M. (2022). Technology-enhanced items in Grades 1–12 English language proficiency assessments. *Language Assessment Quarterly*, 19(4), 343–367. <https://doi.org/10.1080/15434303.2022.2039659>
- Kim, H. S., Lee, J. H., & Lee, H. (2024). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, 28(1), 7-28. <https://doi.org/10.1177/1362168820981394>
- Kim, M., Crossley, S. A., & Kim, B.-K. (2022). Second language reading and writing in relation to first language, vocabulary knowledge, and learning backgrounds.

- International Journal of Bilingual Education and Bilingualism, 25(6), 1992–2005.
<https://doi.org/10.1080/13670050.2020.1838434>
- Latimer, N., & Chan, S. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: Benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Latimer, N., & Chan, S. H. C. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Lee, M., & Jung, J. (2024). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*, 28(4), 1552-1571.
<https://doi.org/10.1177/13621688211034640>
- Li, L., & Zhang, D. (Eds.). (2022). *Reading in Chinese as an additional language: learners' development, instruction, and assessment*. Routledge.
<https://doi.org/10.4324/9781003029038>
- Liao, R. J. (2023). The use of generalizability theory in investigating the score dependability of classroom-based L2 reading assessment. *Language Testing*, 40(1), 86-106.
<https://doi.org/10.1177/02655322211070840>
- Lim, J., Whitehead, G. E. K., & Choi, Y. (2021). Interactive e-book reading vs. paper-based reading: Comparing the effects of different mediums on middle school students' reading comprehension. *System*, 97, Article 102434.
<https://doi.org/10.1016/j.system.2020.102434>
- Lin, J. (2023). ACTFL Chinese reading proficiency guidelines: Verifying the difficulty hierarchy. *Foreign Language Annals*, 56(3), 667–689.
<https://doi.org/10.1111/flan.12680>
- Lin, J. (2023). Assessing the dimensionality of Chinese as a second language reading: a confirmatory factor analysis approach. *Chinese as a Second Language Research*, 12(2), 173–204. <https://doi.org/10.1515/caslar-2023-2001>
- Liu, X., & Read, J. (2021). Investigating the skills involved in reading test tasks through expert judgement and verbal protocol analysis: Convergence and divergence between the two methods. *Language Assessment Quarterly*, 18(4), 357–381.
<https://doi.org/10.1080/15434303.2021.1881964>
- Lundgren, E., & Eklöf, H. (2023). Questionnaire-taking motivation: Using response times to assess motivation to optimize on the PISA 2018 student questionnaire. *International Journal of Testing*, 23:4, 231-256. <https://doi.org/10.1080/15305058.2023.2214647>
- Maplethorpe, L., Kim, H., Hunte, M. R., Vincett, M., & Jang, E. E. (2022). Student-generated questions in literacy education and assessment. *Journal of Literacy Research*, 54(1), 74-97. <https://doi.org/10.1177/1086296X221076436>
- Mikami, Y. (2024). Classification of multiple text reading test formats by analyzing English language proficiency tests. *Language Testing in Asia*, 14, Article 12.
<https://doi.org/10.1186/s40468-024-00285-1>
- Min, S., Bishop, K., & Gary Cook, H. (2022). Reading is a multidimensional construct at child-L2-English-literacy onset, but comprises fewer dimensions over time: Evidence from multidimensional IRT analysis. *Language Testing*, 39(2), 265–288.
<https://doi.org/10.1177/02655322211045296>
- Mohammadi, M., Abbasian, G. R., & Siyyari, M. (2022). Adaptation and validation of a critical thinking scale to measure the 3D critical thinking ability of EFL readers. *Language*

- Testing in Asia*, 12, Article 24. <https://doi.org/10.1186/s40468-022-00173-6>
- Molokopeeva, T., & Simard, D. (2024). Interaction between levels of text representation and working memory during L2 reading comprehension: What about it?. *International Journal of Applied Linguistics*, 34(2), 568–585. <https://doi.org/10.1111/ijal.12516>
- Moradi, E., Ghabanchi, Z., & Pishghadam, R. (2022). Reading comprehension test fairness across gender and mode of learning: Insights from IRT-based differential item functioning analysis. *Language Testing in Asia*, 12, Article 39. <https://doi.org/10.1186/s40468-022-00192-3>
- Murphy Odo, D. (2022). The Effect of Automatic Text Simplification on L2 Readers' Text Comprehension. *Applied Linguistics*, 44(6), 1030–1046. <https://doi.org/10.1093/applin/amac057>
- Nahatame, S. (2023). Predicting processing effort during L1 and L2 reading: The relationship between text linguistic features and eye movements. *Bilingualism: Language and Cognition*, 26(4), 724–737. <http://doi.org.10.1017/S136672892200089X>
- Nawas, A., Darmawan, I.N., & Maadad, N. (2023). Indonesian secular vs. Madrasah schools: assessing the discrepancy in English reading and listening tests. *Language Testing in Asia*, 13, Article 52 (2023). <https://doi.org/10.1186/s40468-023-00266-w>
- Owen, N., Shrestha, P. N., & Hultgren, A. K. (2021). Researching academic reading in two contrasting English as a medium of instruction contexts at a university level. *ETS Research Report Series*, 2021(1), 1–28. <https://doi.org/10.1002/ets2.12317>
- Prichard, C., & Atkins, A. (2021). Evaluating the vocabulary coping strategies of L2 readers: An eye tracking study. *TESOL Quarterly*, 55(2), 593–620. <https://doi.org/10.1002/tesq.3005>
- Ritonga, M., Tazik, K., Omar, A., & Saberi Dehkordi, E. (2022). Assessment and language improvement: the effect of peer assessment (PA) on reading comprehension, reading motivation, and vocabulary learning among EFL learners. *Language Testing in Asia*, 12, Article 36. <https://doi.org/10.1186/s40468-022-00188-z>
- Sarac, M., & Loken, E. (2023). Examining patterns of omitted responses in a large-scale English language proficiency test. *International Journal of Testing*, 23(1), 56–72. <https://doi.org/10.1080/15305058.2022.2070756>
- Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to Generate Reading Comprehension Items. *Educational Measurement: Issues and Practice*, 43(1), 5–18. <https://doi.org/10.1111/emip.12590>
- Shahmirzadi, N. (2024). Simulation study: Seeking fairness in a high-stakes placement test. *Language Testing in Focus: An International Journal*. 10, 18–28, (e-ISSN: 2717-9087). <https://doi.org/10.32038/ltf.2024.10.02>
- Shahmirzadi, N., & Marashi, H. (2023). Cognitive diagnostic assessment of reading comprehension for high-stakes tests: Using GDINA model. *Language Testing in Focus: An International Journal*, (8), 1–16. <https://doi.org/10.32038/ltf.2023.08.01>
- Shahmirzadi, N., Siyyari, M., Marashi, H., & Gramipour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA. *Journal of Foreign Languages Research*, 10(1), 152–165, (ISC), (ISSN: 2588–4123, Online ISSN: 2588–7521). DOI: 10.22059/JFLR.2019.282328.634
- Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts? *Language Learning & Technology*, 27(3), 27–40. <https://hdl.handle.net/10125/73530>

- Shin, J., & Gierl, M. J. (2022). Generating reading comprehension items using automated processes. *International Journal of Testing*, 22(3-4), 289-311. <https://doi.org/10.1080/15305058.2022.2070755>
- Sok, S., Shin, H. W., & Do, J. (2021). Exploring which test-taker characteristics predict young L2 learners' performance on listening and reading comprehension tests. *Language Testing*, 38(3), 378-400. <https://doi.org/10.1177/0265532221991134>
- Sparks, R. L. (2021). Identification and characteristics of strong, average, and weak foreign language readers: The simple view of reading model. *The Modern Language Journal*, 105(2), 507-525. <https://doi.org/10.1111/modl.12711>
- Tarlani-Aliabadi, H., Tazik, K., & Azizi, Z. (2022). Exploring the role of language knowledge and background knowledge in reading comprehension of specific-purpose tests in higher education. *Language Testing in Asia*, 12, Article 48. <https://doi.org/10.1186/s40468-022-00198-x>
- Toprak, T. E., & Cakir, A. (2021). Examining the L2 reading comprehension ability of adult ELLs: Developing a diagnostic test within the cognitive diagnostic assessment framework. *Language Testing*, 38(1), 106-131. <https://doi.org/10.1177/0265532220941470>
- Wang, W., Kingston, N. M., Davis, M. H., Tiemann, G. C., Tonks, S., & Hock, M. (2021). Applying evidence-centered design in the development of a multidimensional adaptive reading motivation measure. *Educational Measurement: Issues and Practice*, 40(4), 91-100. <https://doi.org/10.1111/emip.12468>
- Watzinger-Tharp, J., Tharp, D. S., & Rubio, F. (2021). Sustaining dual language immersion: Partner language outcomes in a statewide program. *The Modern Language Journal*, 105(1), 194-217. <https://doi.org/10.1111/modl.12694>
- Weideman, A. & Van Dyk, T. (2023). Achieving technical economy: A modification of cloze procedure. *Language Teaching Research Quarterly*, 37, 144-160. <https://doi.org/10.32038/ltrq.2023.37.07>
- Weir, C. & Chan, S. (2019). *Research and practice in assessing academic English: The case of IELTS*. Cambridge University Press.
- Wolf, M. K., & Lopez, A. (2022). Developing a technology-based classroom assessment of academic reading skills for English language learners and teachers: validity evidence for formative use. *Languages*, 7(2), Article 71. <https://doi.org/10.3390/languages7020071>
- Yan, Z., King, R. B., & Haw, J. Y. (2021). Formative assessment, growth mindset, and achievement: Examining their relations in the East and the West. *Assessment in Education: Principles, Policy & Practice*, 28(5-6), 676-702. <https://doi.org/10.1080/0969594X.2021.1988510>
- Yang, S. (2021). Diagnosing reading problems for low-level Chinese as second language learners. *System*, 97, Article 102433. <https://doi.org/10.1016/j.system.2020.102433>
- Yang, Y., & Qian, D. D. (2023). Enhancing EFL learners' reading proficiency through dynamic assessment. *Language Assessment Quarterly*, 20(1), 20-43. <https://doi.org/10.1080/15434303.2022.2132160>
- Yeom, S., & Llosa, L. (2024). Comparability of reading tasks in high-stakes English proficiency tests and university courses in Korea. *Language Testing in Asia*, 14, Article 8. <https://doi.org/10.1186/s40468-024-00281-5>
- Zadkhast, M., Rezvani, E., & Lotfi, A. R. (2023). Effects of concurrent and cumulative group dynamic assessments on EFL learners' development of reading comprehension

- micro-skills. *Language Testing in Asia*, 13, Article 29 (2023).
<https://doi.org/10.1186/s40468-023-00240-6>
- Zhao, H., & Aryadoust, V. (2024). A meta-analysis of the reliability of second language reading comprehension assessment tools. *Studies in Second Language Acquisition*, 1–29. doi:10.1017/S0272263124000627
- Zhu, X., & Aryadoust, V. (2022). An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test. *Computer Assisted Language Learning*, 35(3), 412-436.
<https://doi.org/10.1080/09588221.2019.1704788>

Assessing Listening

- Abdellatif, M. S., Alshehri, M. A., Alshehri, H. A., Hafez, W. E., Gafar, M. G., & Lamouchi, A. (2024). I am all ears: Listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement. *Language Testing in Asia*, 14, Article 54.
<https://doi.org/10.1186/s40468-024-00329-6>
- Abdolrezapour, P., & Ghanbari, N. (2021). Enhancing learning potential score in EFL listening comprehension and self-regulation through self-regulated dynamic assessment procedures. *Language Testing in Asia*, 11, Article 10.
<https://doi.org/10.1186/s40468-021-00126-5>
- Ahmed Abdel-Al Ibrahim, K., Karimi, A. R., Abdelrasheed, N.S.G., & Shatalebi, V. (2023). Group dynamic assessment vs. computerized dynamic assessment: Impacts on L2 listening development, L2 students' perfectionism, foreign language anxiety, and intrinsic motivation. *Language Testing in Asia*, 13, Article 32 (2023).
<https://doi.org/10.1186/s40468-023-00245-1>
- Antoniou, K., & Milaki, E. (2021). Irony comprehension in bidialectal speakers. *The Modern Language Journal*, 105(3), 697–719. <https://doi.org/10.1111/modl.12724>
- Aryadoust, V. (2024). Topic and accent coverage in a commercialized L2 listening test: Implications for test-takers' identity. *Applied Linguistics*, 45(5), 765-785.
<https://doi.org/10.1093/applin/amad062>
- Aryadoust, V. (2025). Generative Artificial Intelligence in assessment of listening. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from
<https://meteor.springer.com/container/contribute.jsf?id=263244>
- Aryadoust, V., & Luo, L. (2023). The typology of second language listening constructs: A systematic review. *Language Testing*, 40(2), 375–409.
<https://doi.org/10.1177/02655322221126604>
- Aryadoust, V., Foo, S., & Ng, L. Y. (2022). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Language Testing*, 39(1), 56–89.
<https://doi.org/10.1177/02655322211026876>
- Aryadoust, V., Min, S., & Chen, X. (2024). Investigating differential item functioning across interaction variables in listening comprehension assessment. *Studies in Educational Evaluation*, 80, 101322. <https://doi.org/10.1016/j.steduc.2024.101322>
- Aryadoust, V., Ng, L. Y., Foo, S., & Esposito, G. (2022). A neurocognitive investigation of test methods and gender effects in listening assessment. *Computer Assisted Language*

- Learning*, 35(4), 743–763. <https://doi.org/10.1080/09588221.2020.1744667>
- Aryadoust, V., Zakaria, A., & Jia, Y. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence*. Advance online publication. <https://doi.org/10.1016/j.caeai.2024.100204>
- Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 38(4), 511–535. <https://doi.org/10.1177/0265532220951504>
- Batty, A. O. (2021). Shūjukudo to seibetsu ga bideo L2 chōkairyoku tesuto ni okeru shikaku no kōka ni ataeru eikyō [The impact of proficiency and gender on the effect of visuals in video L2 listening comprehension tests]. *Journal of Foreign Language Education*, 17, 1–18.
- Bourdeaud'Hui, H., Aesaert, K., & van Braak, J. (2021). Exploring the validity of a comprehensive listening test to identify differences in primary school students' listening skills. *Language Assessment Quarterly*, 18(3), 228–252. <https://doi.org/10.1080/15434303.2020.1860059>
- Brunfaut, T., & Harding, L. (2024). Diagnostic approaches in teaching and assessing listening. In E. Wagner, A. O. Batty, & E. Galaczi (Eds.), *The Routledge Handbook of Second Language Acquisition and Listening* (pp. 255–267). (Routledge Handbooks in Second Language Acquisition). Routledge. <https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Listening/Wagner-Batty-Galaczi/p/book/9781032113647>
- Carney, N. (2021). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*, 55(2), 536–567. <https://doi.org/10.1002/tesq.3000>
- Choi, I. & Zu, J. (2022). The impact of using synthetically generated listening stimuli on test-taker performance: A case study with multiple-choice, single-selection items. *ETS Research Report Series*, 2022(1), 1–14. <https://doi.org/10.1002/ets2.12347>
- Chong, J. J. Q., & Aryadoust, V. (2023). Investigating the effect of multimodality and sentiments on speaking assessments: A facial emotional analysis. *Education and Information Technologies*, 28, 7413–7436. <https://doi.org/10.1007/s10639-022-11478-7>
- Dao, P., Nguyen, M. X. N. C., & Nguyen, N. B. C. (2021). Effect of pronunciation instruction on L2 learners' listening comprehension. *Journal of Second Language Pronunciation*, 7(1), 10–37. <https://doi.org/10.1075/jslp.19012.dao>
- Du, G., & Man, D. (2022). Person factors and strategic processing in L2 listening comprehension: Examining the role of vocabulary size, metacognitive knowledge, self-efficacy, and strategy use. *System*, 107, 102801. <https://doi.org/10.1016/j.system.2022.102801>
- Ebadi, S., Karimi, E., & Vakili, S. (2023). An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment. *Language Testing in Asia*, 13, Article 5 (2023). <https://doi.org/10.1186/s40468-023-00221-9>
- Eberharter, K., Kormos, J., Guggenbichler, E., Ebner, V. S., Suzuki, S., Moser-Frötscher, D., Konrad, E., & Kremmel, B. (2023). Investigating the impact of self-pacing on the L2 listening performance of young learner candidates with differing L1 literacy skills. *Language Testing*, 40(4), 960–983. <https://doi.org/10.1177/02655322221149642>
- Ghaemi, H. (2022). Application of nonparametric item response theory in determining the one-dimensionality and adaptability of TOEFL iBT listening test. *Language Testing in Asia*, 12, Article 22. <https://doi.org/10.1186/s40468-022-00171-8>

- Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11, Article 20. <https://doi.org/10.1186/s40468-021-00131-8>
- Hamada, Y., & Suzuki, S. (2021). Listening to Global Englishes: Script-assisted shadowing. *International Journal of Applied Linguistics*, 31(1), 31–47. <https://doi.org/10.1111/ijal.12318>
- Han, D., & Qian, D. D. (2024). Evaluating the roles of breadth and depth of aural vocabulary knowledge in listening comprehension of EFL learners: An investigation applying auditory measures. *System*, 120, 103207. <https://doi.org/10.1016/j.system.2023.103207>
- He, L., Xiong, L., & Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: Performance stability and variability across modalities and performance levels. *System*, 106, Article 102787. <https://doi.org/10.1016/j.system.2022.102787>
- Holzknicht, F., & Harding, L. (2024). Repeating the listening text: effects on listener performance, metacognitive strategy use, and anxiety. *TESOL Quarterly*, 58(1), 451–478. <https://doi.org/10.1002/tesq.3249>
- Holzknicht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiby, R., & Dunlea, J. (2021). The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test. *Language Testing*, 38(1), 41–61. <https://doi.org/10.1177/0265532220917316>
- Hsieh, C.-N. (2023). *Evaluating the use and interpretations of the TOEIC Listening and Reading test score report: Perspectives of test takers in Japan*. ETS Research Report Series (RR-23-02). <https://doi.org/10.1002/ets2.12364>
- Hsieh, C.-N. (2023). *Interpretation and use of a workplace English language proficiency test score report: Perspectives of test takers and score users in Taiwan*. ETS Research Report Series (RR-23-10). <https://doi.org/10.1002/ets2.12373>
- Hui, B., Wong, S. S. Y., & Au, R. K. (2022). Reading aloud listening test items to young learners: Attention, item understanding, and test performance. *System*, 108, Article 102831. <https://doi.org/10.1016/j.system.2022.102831>
- In'nami, Y., & Koizumi, R. (2022). The relationship between L2 listening and metacognitive awareness across listening tests and learner samples. *International Journal of Listening*, 36(2), 100–117. <http://dx.doi.org/10.1080/10904018.2021.1955683>
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2023). Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach. *Language Learning*, 73(3), 759–798. <https://doi.org/10.1111/lang.12548>
- In'nami, Y., Koizumi, R., Jeon, E. H., & Arai, Y. (2022). L2 listening and its correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 235–283). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Kang, O., Yan, X., Kostromitina, M., Thomson, R., & Isaacs, T. (2024). Fairness of using different English accents: The effect of shared L1s in listening tasks of the Duolingo English test. *Language Testing*, 41(2), 263–289. <https://doi.org/10.1177/02655322231179134>
- Kho, S. Q. E., Aryadoust, V., & Foo, S. (2023). An eye-tracking investigation of the keyword-matching strategy in listening assessment. *Education and Information Technologies*, 28, 3739–3763.

- <https://link.springer.com/article/10.1007/s10639-022-11322-y>
- Kim, M., Nam, Y., & Crossley, S. A. (2022). Roles of working memory, syllogistic inferencing ability, and linguistic knowledge on second language listening comprehension for passages of different lengths. *Language Testing*, 39(4), 593–617. <https://doi.org/10.1177/02655322211060076>
- Koizumi, R., Agawa, T., Asano, K., & In'nami, Y. (2022). Skill profiles of Japanese English learners and reasons for uneven patterns. *Language Testing in Asia*, 12, Article 53. <https://doi.org/10.1186/s40468-022-00203-3>
- Kormos, J., Eberharter, K., Guggenbichler, E., Baumgartinger, S., Ebner, V., & Kremmel, B. (2024). First Language Literacy Skills, Listening Strategy Use, and Anxiety in Self-Paced versus Single-Play L2 Listening Tests. *Language Assessment Quarterly*, 21(4–5), 303–323. <https://doi.org/10.1080/15434303.2024.2435811>
- Kwon, S. K., & Yu, G. (2023). Investigating differences in test-takers' use of cognitive and metacognitive strategies in audio-only and video-based listening comprehension test. *System*, 114, Article 103017. <https://doi.org/10.1016/j.system.2023.103017>
- Kwon, S. K., & Yu, G. (2024). The effect of viewing visual cues in a listening comprehension test on second language learners' test-taking process and performance: An eye-tracking study. *Language Testing*, 41(3), 649–680. <https://doi.org/10.1177/02655322241239356>
- Lange, K., & Matthews, J. (2023). Analyzing trends in the aural decoding errors of Japanese EFL learners. *International Review of Applied Linguistics in Language Teaching*, 61(2), 545–570. <https://doi.org/10.1515/iral-2020-0050>
- Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*, 55(3), 1011–1023. <https://doi.org/10.1002/tesq.3055>
- Liu, T., & Aryadoust, V. (2024). Does modality matter? A meta-analysis of the effect of video input in L2 listening assessment. *System*, 120, 103191. <https://doi.org/10.1016/j.system.2023.103191>
- Liu, T., Aryadoust, V., & Foo, S. (2022). Examining the factor structure and its replicability across multiple listening test forms: Validity evidence for the Michigan English test. *Language Testing*, 39(1), 142–171. <https://doi.org/10.1177/02655322211018139>
- Low, A. R. L., & Aryadoust, V. (2023). Investigating test-taking strategies in listening assessment: A comparative study of eye-tracking and self-report questionnaires. *International Journal of Listening*, 37(2), 93–112. DOI: 10.1080/10904018.2021.1883433
- Mahalingappa, L., Zong, J., & Polat, N. (2024). The impact of captioning and playback speed on listening comprehension of multilingual English learners at varying proficiency levels. *System*, 120, 103192. <https://doi.org/10.1016/j.system.2023.103192>
- Meng, Y., & Fu, H. (2023). Modeling mediation in the dynamic assessment of listening ability from the cognitive diagnostic perspective. *Modern Language Journal*, 107, 137–160. <https://doi.org/10.1111/modl.12820>
- Meng, Y., Fu, H., & Wang, C. (2024). The effectiveness of computerized listening dynamic assessment: Attribute-based mediation model. *Language Learning & Technology*, 28(1), 1–28. <https://hdl.handle.net/10125/73580>
- Miao, Y. (2023). The relationship among accent familiarity, shared L1, and comprehensibility: A path analysis perspective. *Language Testing*, 40(3), 723–747. <https://doi.org/10.1177/02655322231156105>

- Milliner, B., & Dimoski, B. (2024). The effects of a metacognitive intervention on lower-proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research*, 28(2), 679-713. <https://doi.org/10.1177/13621688211004646>
- Min, S., & He, L. (2022). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, 39(1), 90-116. <https://doi.org/10.1177/0265532221995475>
- Min, S., Cai, H., & He, L. (2022). Application of bi-factor MIRT and higher-order CDM models to an in-house EFL listening test for diagnostic purposes. *Language Assessment Quarterly*, 19(2), 189-213. <https://doi.org/10.1080/15434303.2021.1980571>
- Min, S., Zhang, J., Li, Y., & He, L. (2022). Bridging local needs and national standards: Use of standards-based individualized feedback of an in-house EFL listening test in China. *Language Testing*, 39(3), 425-452. <https://doi.org/10.1177/02655322211070990>
- Nawas, A., Darmawan, I.N., & Maadad, N. (2023). Indonesian secular vs. Madrasah schools: assessing the discrepancy in English reading and listening tests. *Language Testing in Asia*, 13, Article 52 (2023). <https://doi.org/10.1186/s40468-023-00266-w>
- Nishizawa, H. (2023). Construct validity and fairness of an operational listening test with World Englishes. *Language Testing*, 40(3), 493-520. <https://doi.org/10.1177/02655322221137869>
- Nishizawa, H. (2024). Authenticity of academic lecture passages in high-stakes tests: A temporal fluency perspective. *Language Testing*, 41(4), 792-816. <https://doi.org/10.1177/02655322241262453>
- O'Grady, S. (2023). Adapting multiple-choice comprehension question formats in a test of second language listening comprehension. *Language Teaching Research*, 27(6), 1431-1455. <https://doi.org/10.1177/1362168820985367>
- O'Sullivan, B. Saville, N. (2024). Foreword. In: Field, J. *Insights into Assessing Academic Listening: The Case of IELTS*. SiLT 53, Cambridge: Cambridge University Press and Assessment. <https://www.cambridge.org/cf/cambridgeenglish/teacher-development/insights-assessing-academic-listening-case-ielts/insights-assessing-academic-listening-case-ielts-paperback>
- Ockey, G. J., & Chukharev-Hudilainen, E. (2021). Human versus computer partner in the paired oral discussion test. *Applied Linguistics*, 42(5), 924-944. <https://doi.org/10.1093/applin/amaa067>
- Panahi, A., & Mohebbi, H. (2022). Cognitive diagnostic assessment of IELTS listening: Providing feedback from its internal structure. *Language Teaching Research Quarterly*, 29, 147-160. <https://doi.org/10.32038/ltrq.2022.29.10>
- Park, Y., Lee, S., & Shin, S.-Y. (2022). Developing a local academic English listening test using authentic unscripted audio-visual texts. *Language Testing*, 39(3), 401-424. <https://doi.org/10.1177/02655322221076024>
- Qiu, Y., & Aryadoust, V. (2024). The predictive value of gaze behavior and mouse-clicking in testing listening proficiency: A sensor technology study. *System*, 126, 103440. <https://doi.org/10.1016/j.system.2024.103440>
- Rossi, O., & Brunfaut, T. (2021). Text authenticity in listening assessment: Can item writers be trained to produce authentic-sounding texts? *Language Assessment Quarterly*,

- 18(4), 398–418. <https://doi.org/10.1080/15434303.2021.1895162>
- Rukthong, A. (2021). MC listening questions vs. integrated listening-to-summarize tasks: What listening abilities do they assess? *System*, 97, Article 102439. <https://doi.org/10.1016/j.system.2020.102439>
- Shahmirzadi, N. (2023). Validation of a language center placement test: Differential item functioning. *International Journal of Language Testing*, 13(1), 1-17. <https://doi.org/10.22034/IJLT.2022.336779.1151>
- Shang, Y., Aryadoust, V., & Hou, Z. (2024). A meta-analysis of the reliability of second language listening tests (1991–2022). *Brain Sciences*, 14(8), 746. <https://doi.org/10.3390/brainsci14080746>
- Shin, S. Y., Lee, S., & Lidster, R. (2021). Examining the effects of different English speech varieties on an L2 academic listening comprehension test at the item level. *Language Testing*, 38(4), 580–601. <https://doi.org/10.1177/0265532220985432>
- Shin, S., & Lee, S. (2024). Effects of the type of audio-visual input and listenability on young L2 learners' listening comprehension. *TESOL Quarterly*, 58(4), 1580–1608. <https://doi.org/10.1002/tesq.3298>
- Sok, S., Shin, H. W., & Do, J. (2021). Exploring which test-taker characteristics predict young L2 learners' performance on listening and reading comprehension tests. *Language Testing*, 38(3), 378–400. <https://doi.org/10.1177/0265532221991134>
- Suvorov, R., & He, S. (2022). Visuals in the assessment and testing of second language listening: A methodological synthesis. *International Journal of Listening*, 36(2), 80-99. <https://doi.org/10.1080/10904018.2021.1941028>
- Suvorov, R., & Li, Z. (2023). *Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS*. British Council, Cambridge Assessment English, and IDP: IELTS Australia. <https://search.informit.org/doi/epdf/10.3316/informit.262825338814353>
- Suzuki, R., & Yamane, S. (2024). Japanese EFL learners' comprehension and perceptions of three different accents in listening assessment. *International Journal of Applied Linguistics*, 34(4), 1497-1513. <https://doi.org/10.1111/ijal.12584>
- Tao, X., & Aryadoust, V. (2024). A multidimensional analysis of a high-stakes English listening test: A corpus-Based Approach. *Education Sciences*, 14(2), 137. <https://doi.org/10.3390/educsci14020137>
- Wagner, E. (2022). Assessing listening. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing* (2nd Ed.) (pp. 223–235). Routledge.
- Wagner, E. (2022). Assessing second language listening. In H. Mohebbi & C. Coombe (Eds.), *Research questions in language education and applied linguistics* (pp. 299–304). Springer. https://doi.org/10.1007/978-3-030-79143-8_53
- Wagner, E., Batty, A. O. & Galaczi, E. (Eds.). *The Routledge Handbook of Second Language Acquisition and Listening* (1st ed.). Routledge.
- Wagner, E., Liao, Y. F., & Wagner, S. (2021). Authenticated spoken texts for L2 listening tests. *Language Assessment Quarterly*, 18(3), 205–227. <https://doi.org/10.1080/15434303.2020.1860057>
- Wallace, M.P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72(1), 5-44. <https://doi.org/10.1111/lang.12424>
- Westbrook, C. (2023). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61,

- Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2023). High-level listening comprehension in advanced English as a second language: Effects of the first language and inhibitory control. *Bilingualism: Language and Cognition*, 26(5), 865-879.
<https://doi.org/10.1017/S1366728923000135>
- Yeager, R. R., Park, G., & Liao, R. J. T. (2024). Notetaking as Validity Evidence: A Mixed-Methods Investigation of Question Preview in Listening Assessment. *Journal of English for Academic Purposes*, 68(3), 101346.
<https://doi.org/10.1016/j.jeap.2024.101346>
- Yeager, R., Park, G., & Liao, R. J. (2024). Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment. *Journal of English for Academic Purposes*, 68, Article 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Yeldham, M., & Chen, R. T.-H. (2024). Conducting L2 listener verbal reports: Common missteps by novice researchers and ways to address them. *Research Methods in Applied Linguistics*, 3(2), Article 100113. <https://doi.org/10.1016/j.rmal.2024.100113>
- Yenkimaleki, M., van Heuven, V. J., & Hosseini, M. (2024). The effect of providing feedback and feedforward in prosody instruction for developing listening comprehension skills by interpreter trainees. *Foreign Language Annals*, 57(1), 184-200.
<https://doi.org/10.1111/flan.12725>
- Zhai, X., & Aryadoust, V. (2022). The metacognitive and neurocognitive signature of test methods in academic listening assessment. *Frontiers in Psychology*, 13, Article 930075. <https://doi.org/10.3389/fpsyg.2022.930075>
- Zhang, C., Xu, X., & Zhang, Y. (2023). A communicative validation study on an English listening test in Korea. *Language Testing in Asia*, 13, Article 26 (2023).
<https://doi.org/10.1186/s40468-023-00238-0>
- Zhao, Y., & Aryadoust, V. (2024). An automatized semantic analysis of two large-scale listening tests: A corpus-based study. *Language Testing*, 0(0).
<https://doi.org/10.1177/02655322241288598>
- Zhou, S., & Thompson, G. (2023). A longitudinal study on students' self-regulated listening during transition to an English-medium transnational University in China. *Studies in Second Language Learning and Teaching*, 13(2), 427–450.
<https://doi.org/10.14746/sslt.38281>

Assessing Speaking

- Abdulaal, M. A. A. D., Alenazi, M. H., Tajuddin, A. J. A., & Hamidi, B. (2022). Dynamic vs. diagnostic assessment: Impacts on EFL learners' speaking fluency and accuracy, learning anxiety, and cognitive load. *Language Testing in Asia*, 12, Article 32.
<https://doi.org/10.1186/s40468-022-00179-0>
- Ahmadi, A. (2024). Patterns of interaction in a paired speaking test: Comparing L1 and L2 interactions. *Language Testing in Asia*, 14, Article 31.
<https://doi.org/10.1186/s40468-024-00299-9>
- Al-Ghezi, R., Voskoboinik, K., Getman, Y., Von Zansen, A., Kallio, H., Kurimo, M., Huhta, A., & Hildén, R. (2023). Automatic speaking assessment of spontaneous L2 Finnish and Swedish. *Language Assessment Quarterly*, 20(4-5), 421–444.
<https://doi.org/10.1080/15434303.2023.2292265>

- Ali, M. M. E. (2023). The foreign-accentedness, comprehensibility, and intelligibility of L2 Arabic speech. *Language Teaching Research*, 0(0).
<https://doi-org.ezproxy2.library.arizona.edu/10.1177/13621688231158787>
- Aryadoust, V., & Zakaria, A. (2025). ChatGPT in assessment of speaking. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from
<https://meteor.springer.com/container/contribute.jsf?id=263242>
- Bao, G. (2023). Factor structures of speed and breakdown fluency in EFL learners' story retelling performances. *International Review of Applied Linguistics in Language Teaching*, 61(2), 631-654. <https://doi.org/10.1515/iral-2022-0125>
- Barrows, J., & Cox, T. (2021). The effect of prompt accent on elicited imitation assessments in English as a second language. *Papers in Language Testing and Assessment*, 10(2), 1–29.
http://www.altanz.org/uploads/5/9/0/8/5908292/2._plta_10_2__barrows_cox.pdf
- Bell, T. R., & Borden, R. S. (2022). Examining target language proficiency among in-service K-12 language teachers. *Foreign Language Annals*, 55 (3), 815–831.
<https://doi.org/10.1111/flan.12637>
- Bijani, H., Hashempour, B., Ibrahim, K. A. A. A., Orabah, S. S. B., & Heydarnejad, T. (2022). Investigating the effect of classroom-based feedback on speaking assessment: A multifaceted Rasch analysis. *Language Testing in Asia*, 12, Article 26.
<https://doi.org/10.1186/s40468-022-00176-3>
- Burton, J. D. (2021). The face of communication breakdown: Multimodal repair in L2 oral proficiency interviews. *Papers in Language Testing and Assessment*, 10(2), 30–61.
http://www.altanz.org/uploads/5/9/0/8/5908292/3._plta_10_2__burton.pdf
- Burton, J. D. (2023). Gazing into cognition: Eye behavior in online L2 speaking tests. *Language Assessment Quarterly*, 20(2), 190-214.
<https://doi.org/10.1080/15434303.2022.2143680>
- Burton, J. D. (2024). Evaluating the impact of nonverbal behavior on language ability ratings. *Language Testing*, 41(4), 729–758.
<https://doi.org/10.1177/02655322241255709>
- Canals, E. (2023). Modified output and metalanguage during conversational interaction: A qualitative look at interactional feedback. *Studies in Second Language Learning and Teaching*, 13(3), 601-626. <https://doi.org/10.14746/ssllt.31128>
- Chan, J. Y. H. (2021). Bridging the gap between ELF and L2 learners' use of communication strategies: Rethinking current L2 assessment and teaching practices. *System*, 101, Article 102609.
<https://doi.org/10.1016/j.system.2021.102609>
- Chapelle, C. A. & Lee, H. (2022). Validation of second language spoken assessments. In T. Haug, W. Mann & U. Knoch (Eds.), *The handbook of language assessment across modalities* (pp. 273–284). Oxford University Press.
<https://doi.org/10.1093/oso/9780190885052.003.0023>
- Chapelle, C. A., Hauser, P. C., Lee, H., Rathmann, C., & Schönström, K. (2022). Discussion of validation issues in signed and spoken assessments for adult L2 learners. In T. Haug, W. Mann & U. Knoch (Eds.), *The handbook of language assessment across modalities* (pp. 295–298). Oxford University Press.
<https://doi.org/10.1093/oso/9780190885052.003.0025>

- Chen Hsieh, J. (2021). Digital storytelling outcomes and emotional experience among middle school EFL Learners: Robot-assisted versus PowerPoint-assisted mode. *TESOL Quarterly*, 55(3), 994–1010. <https://doi.org/10.1002/tesq.3043>
- Chen, T.-H. (2024). The effects of task complexity on L2 English rapport-building language use and its relationship with paired speaking test task performance. *Applied Linguistics Review*, 15(2), 737–769. <https://doi.org/10.1515/applirev-2021-0199>
- Choi, J. S., & Loewen, S. (2022). Exploring young learners' strategic behaviors in a speaking test. *TESOL Quarterly*, 56(4), 1384–1396. <https://doi.org/10.1002/tesq.313>
- Choi, S., & Kang, O. (2023). The roles of suprasegmental features in assessing paired speaking tasks in high-stakes language assessment. *System*, 119, Article 103183. <https://doi.org/10.1016/j.system.2023.103183>
- Chong, J. J. Q., & Aryadoust, V. (2023). Investigating the effect of multimodality and sentiments on speaking assessments: A facial emotional analysis. *Education and Information Technologies*, 28, 7413–7436. <https://doi.org/10.1007/s10639-022-11478-7>
- Chou, M. H. (2021). An integrated approach to developing and assessing EFL students' speaking ability and strategy use. *Language Education & Assessment*, 4(1), 19–37. <https://doi.org/10.29140/lea.v4n1.428>
- Chukharev-Hudilainen, E., & Ockey, G. J. (2021). The development and evaluation of interactional competence elicitor for oral language assessments. *ETS Research Report Series*, 2021(1), 1–20. <https://doi.org/10.1002/ets2.12319>
- Cox, T. L., Brown, A. V., & Thompson, G. L. (2023). Temporal fluency and floor/ceiling scoring of intermediate and advanced speech on the ACTFL Spanish Oral Proficiency Interview–computer. *Language Testing*, 40(2), 325–351. <https://doi.org/10.1177/02655322221114614>
- Crowther, D., Isbell, D. R., & Nishizawa, H. (2023). Second language speech comprehensibility and acceptability in academic settings: Listener perceptions and speech stream influences. *Applied Psycholinguistics*, 44(5), 858–888. <https://doi.org/10.1017/S0142716423000346>
- Davis, L., & Norris, J. (2021). Developing an innovative elicited imitation task for efficient English proficiency assessment. *ETS Research Report Series*, 2021(1), 1–30. <https://doi.org/10.1002/ets2.12338>
- Davis, L., & Papageorgiou, S. (2021). Complementary strengths? Evaluation of a hybrid human-machine scoring approach for a test of oral academic English. *Assessment in Education: Principles, Policy & Practice*, 28(4), 437–455. <https://doi.org/10.1080/0969594X.2021.1979466>
- De John, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9(1), 541–560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>
- De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476. <https://doi.org/10.1080/0969594X.2021.1951162>
- De Malsche, F., Tobback, E., & Vandenbroucke, M. (2024). Small talk as relational practice during assessment. *Journal of Applied Linguistics and Professional Practice*, (0), 1–24.

- Ding, W., & Yu, G. (2023). Using causal explanation speaking tasks to assess young EFL learners' speaking ability: The effects of age, cognitive, and L2 linguistic development. *Language Assessment Quarterly*, 20(3), 251-276. <https://doi.org/10.1080/15434303.2023.2206811>
- Ducasse, A.M. (2022) Oral reflection tasks: Advanced Spanish L2 learner insights on emergency remote teaching assessment practices in a higher education context. *Languages*, 7(1), Article 26. <https://doi.org/10.3390/languages7010026>
- Duong, P. T., Perez, M. M., Nguyen, L. Q., Desmet, P., & Peters, E. (2023). The impact of input, input repetition, and task repetition on L2 lexical use and fluency in speaking. *Studies in second language learning and teaching*, 13(1), 101-124. <https://doi.org/10.14746/ssllt.29727>
- Eckes, T., & Jin, K. Y. (2021). Examining severity and centrality effects in TestDaF writing and speaking assessments: An extended Bayesian many-facet Rasch analysis. *International Journal of Testing*, 21(3-4), 131-153. <https://doi.org/10.1080/15305058.2021.1963260>
- el Majidi, A., de Graaff, R., & Janssen, D. (2024). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*, 28(6), 2431-2452. <https://doi.org/10.1177/13621688211050619>
- François, J., & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. *Language Learning & Technology*, 25(2), 26-41. <http://hdl.handle.net/10125/73429>
- Frisén, L. B. (2024). Eyes on the ball: Teachers' rating processes when assessing a national L2 speaking test through the lens of their scoring document. *Studies in Language Assessment (SILA)*, 13(1), 1-36. <https://doi.org/10.58379/DAGH1483>
- Frost, K., Wigglesworth, G., & Clothier, J. (2021). Relationships between comprehension, strategic behaviours and content-related aspects of test performances in integrated speaking tasks. *Language Assessment Quarterly*, 18(2), 133-153. <https://doi.org/10.1080/15434303.2020.1835918>
- Gao, J., & Sun, P. P. (2024). How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics*, 34(1), 276-291. <https://doi.org/10.1111/ijal.12493>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- Garcia-Ponce, E. E., & Tavakoli, P. (2022). Effects of task type and language proficiency on dialogic performance and task engagement. *System*, 105, Article 102734. <https://doi.org/10.1016/j.system.2022.102734>
- Getman, E. P. (2024). Four task characteristics and their influence on speaking performance and engagement. *English Teaching & Learning*, 48(2), 265-290. <https://doi.org/10.1007/s42321-024-00176-x>
- Getman, E. P. (2024). Speaking test tasks for young language learners: Development and validation. In L. Davis & J. M. Norris (Eds.), *Challenges and Innovations in Speaking Assessment*. Routledge. <https://doi.org/10.4324/9781315201504-7>
- Ghaffarvand-Mokari, P. (2024). Effects of listeners' dialectal variation on comprehensibility and accentedness judgements of second language speech. *Journal of Second Language Pronunciation*, 10(1), 35-58. <https://doi.org/10.1075/jslp.23018.gha>

- Gilbert, S., & May, L. (2023). Video-call mode speaking tests: an examiner verbal protocol study. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 71–86). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Glasson, N. (2022). Is the devil you know better? Testwiseness and eliciting evidence of interactional competence in familiar versus unfamiliar triadic speaking tasks. *Studies in Language Assessment*, 11(2), 58–97.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0004/4396261/SiLA_11.2_Glasson.pdf
- Glasson, N., & Devine, A. (2023). The eye of the stakeholder—perceptions of remote speaking. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 53–70). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Glasson, N., & Halley, K. (2024). Less talk, more action? Exploring proficiency scores and embodied resources in online L2 interactions. *Classroom Discourse*, 1–22.
<https://doi.org/10.1080/19463014.2024.2408421>
- Gokturk, N., & Chukharev-Hudilainen, E. (2023). Strategy use in a spoken dialog system–delivered paired discussion task: A stimulated recall study. *Language Testing*, 40(3), 630–657. <https://doi.org/10.1177/02655322231152620>
- Gokturk, N., & Chukharev, E. (2024). Exploring the Potential of a Spoken Dialog System-Delivered Paired Discussion Task for Assessing Interactional Competence. *Language Assessment Quarterly*, 21(1), 60–99.
<https://doi.org/10.1080/15434303.2023.2289173>
- Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58–76.
<https://doi.org/10.1080/0969594X.2020.1735995>
- Guo, J., & East, M. (2024). The Role of Vocabulary in Speaking: Voices from Learners and Raters. *Language Assessment Quarterly*, 21(4–5), 452–472.
<https://doi.org/10.1080/15434303.2024.2440887>
- Handley, Z. L., & Wang, H. (2023). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency? *Language Assessment Quarterly*, 21(1), 3–32. <https://doi.org/10.1080/15434303.2023.2283839>
- Handley, Z. L., & Wang, H. (2024). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency?. *Language Assessment Quarterly*, 21(1), 3–32.
- Hannah, L., Kim, H., & Jang, E. E. (2022). Investigating the effects of task type and linguistic background on accuracy in automated speech recognition systems: Implications for use in language assessment of young learners. *Language Assessment Quarterly*, 19(3), 289–313.
<https://doi.org/10.1080/15434303.2022.2038172>
- Hanzawa, K. (2024). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, 28(3), 816–838.
<https://doi.org/10.1177/13621688211008693>
- Hayashi, Y., Kondo, Y., & Ishii, Y. (2023). Automated speech scoring of dialogue response by Japanese learners of English as a foreign language. *Innovation in Language Learning and Teaching*, 18(1), 32–46.

- <https://doi.org/10.1080/17501229.2023.2217181>
- Heidari Vinchek, M., Mirzaei, A., & Roohani, A. (2024). A cognitive diagnostic approach to IELTS speaking test: Unveiling the subskills and test-takers' perceptions. *Language Testing in Asia*, 14, Article 42. <https://doi.org/10.1186/s40468-024-00311-2>
- Hougham, D., Clenton, J., & Uchiyara, T. (2024). Disentangling the contributions of shorter vs. Longer lexical bundles to L2 oral fluency. *System*, 121, 103243. <https://doi.org/10.1016/j.system.2024.103243>
- Huang, B. H., Bailey, A. L., Sass, D. A., & Shawn Chang, Y. (2021). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401–428. <https://doi.org/10.1177/0265532220925731>
- Hunte, M. R., McCormick, S., Shah, M., Lau, C., & Jang, E. E. (2021). Investigating the potential of NLP-driven linguistic and acoustic features for predicting human scores of children's oral language proficiency. *Assessment in Education: Principles, Policy & Practice*, 28(4), 477–505. <https://doi.org/10.1080/0969594X.2021.1999209>
- Inoue, C., & Lam, D. M. K. (2021). *The effects of extended planning time on candidates' performance, processes, and strategy use in the lecture listening-into-speaking tasks of the TOEFL iBT Test*. (TOEFL Research Report, TOEFL-RR-93, ETS Research Report No. RR-21-09). <https://doi.org/10.1002/ets2.12322>
- Inoue, C., & Nakatsuhara, F. (2022). Validation of a large-scale task-based test: Functional progression in dialogic speaking performance. In N. P. Sudharshana & L. Mukhopadhyay (eds.), *Task-based language teaching and assessment: Contemporary reflections from across the world* (pp. 217 - 247). Springer Nature.
- Inoue, C., Khabbazzashi, N., Lam, D. M. K., & Nakatsuhara, F. (2021). Towards new avenues for the IELTS Speaking Test: Insights from examiners' voices. *IELTS Research Reports Online Series*, 2. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/online-series-2021-2.ashx>
- Isaacs, T. (2025). Assessing pronunciation. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 216-229). Wiley.
- Isaacs, T. & Thomson, R. I. (2022). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *The evolution of pronunciation teaching and research: 25 years of intelligibility, comprehensibility, and accentedness* (pp. 125-151). John Benjamins.
- Isbell, D. R. (2021). Can the test support student learning? Validating the use of a second language pronunciation diagnostic. *Language Assessment Quarterly*, 18(4), 331–356. <https://doi.org/10.1080/15434303.2021.1874382>
- Isbell, D. R., & Son, Y. A. (2022). Measurement properties of a standardized elicited imitation test: An integrative data analysis. *Studies in Second Language Acquisition*, 44(3), 859-885. <https://doi.org/10.1017/S0272263121000383>
- Isbell, D. R., Crowther, D., & Nishizawa, H. (2024). Speaking performances, stakeholder perceptions, and test scores: Extrapolating from the Duolingo English test to the university. *Language Testing*, 41(2), 233–262. <https://doi.org/10.1177/02655322231165984>
- Isbell, D. R., Kim, K. M., & Chen, X. (2023). Exploring the potential of automated speech recognition for scoring the Korean Elicited Imitation Test. *Research Methods in Applied Linguistics*, 2(3), Article 100076. <https://doi.org/10.1016/j.rmal.2023.100076>

- Isbell, D. R., Lee, J., & Jang, J. (2024). Linguistic influences on comprehensibility and accentedness in second language Korean speech. *International Journal of Applied Linguistics*, 34(4), 1575-1591. <https://doi.org/10.1111/ijal.12580>
- Isbell, D., & Sakai, M. (2022). Pronunciation assessment in classroom contexts. In T. Derwing, J. Levis, & S. Sonsaat (Eds.), *Pronunciation in Second Language Learning and Teaching: Innovations and Developments in Research and Teaching* (pp. 194-214). Routledge.
- Isbell, D.R., & Lee, J. (2022). Self-Assessment of comprehensibility and accentedness in second language Korean. *Language Learning*, 72(3), 806-852. <https://doi.org/10.1111/lang.12497>
- Iwashita, N., Sasaki, M., Stell, S., & Yucel, M. (2021). Japanese stakeholders' perceptions of IELTS writing and speaking tests and their impact on communication and achievement. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/teaching-and-research/research-reports>
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, Article 102521. <https://doi.org/10.1016/j.system.2021.102521>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). Discussion, limitations, and future research. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 369-385). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). L2 speaking and its external correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 339-367). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Johnson, C., Cardoso, W., Zuercher, B., Brannen, K., & Springer, S. (2024). Assessing pronunciation using dictation tools: The use of Google Voice Typing to score a pronunciation placement test. *Journal of Second Language Pronunciation*, 10(1), 10–34. <https://doi.org/10.1075/jslp.23033.joh>
- Jones, J., & Isaacs, T. (2022). Assessing second language pronunciation. In H. Moehebbi, & C. Coombe (Eds.), *Research questions in language education: A reference guide for teachers* (pp. 305-310). Springer.
- Jung Youn, S. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing*, 40(1), 54–60. <https://doi.org/10.1177/02655322221126606>
- Kang, O., & Johnson, D. (2021). *Linguistic features and automatic scoring of Aptis speaking performance*. (AR-G/2021/3). British Council. https://www.britishcouncil.org/sites/default/files/kang_and_johnson_layout.pdf
- Kang, O., Ahn, H., Yaw, K., & Chung, S-Y. (2021). Investigation of relationships between learner background, linguistic progression, and score gain on IELTS. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia. https://www.ielts.org/-/media/research-reports/ielts-rr_2021-1_kang-et-al.ashx
- Kang, O., Egbert, J., & Miao, Y. (2023). *Speaking and writing features: Distinguishing IELTS proficiency levels and progression over time* (IELTS Research Report Series No. 1/23). British Council, Cambridge Assessment English, and IDP: IELTS Australia.

- <https://search.informit.org/doi/epdf/10.3316/informit.262788072871836>
- Kang, O., Kermad, A., & Taguchi, N. (2021). The interplay of proficiency and study abroad experience on the prosody of L2 speech acts. *Journal of Second Language Pronunciation*, 7(3), 343–369. <https://doi.org/10.1075/jslp.20024.kan>
- Karatay, Y. (2023). Using spoken dialog systems to assess L2 learners' oral skills in a local language testing context. In Y. Xun, D. Slobodanka & J. G. April (Eds.), *Local language testing: Practice across contexts* (pp. 231–252). Springer. https://link.springer.com/chapter/10.1007/978-3-031-33541-9_12
- Karyolemou, M. (2022). Teaching endangered languages of oral tradition: how and what to assess?. *Applied Linguistics*, 43(2), 389–411. <https://doi.org/10.1093/applin/amab045>
- Khabbazzbashi, N. (2021). *On topic validity of speaking test*. Cambridge University Press.
- Khabbazzbashi, N., Nakatsuhara, F., Inoue, C., Kaplan, G., & Green, A. (2022). The design and validation of an online speaking test for young learners in Uruguay: challenges and innovations. *International Journal of TESOL Studies*, 4(1), 141–168. <https://doi.org/10.46451/ijts.2022.01.10>
- Khabbazzbashi, N., Xu, J., & Galaczi, E. (2021). Opening the black box: Exploring automated speaking assessment. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Issues in language testing around the world* (pp. 333–344). Springer. https://doi.org/10.1007/978-981-33-4232-3_25
- Khau, A. H., & Huynh, V. T. M. (2022). An investigation into oral fluency perceived by teachers and students—in a Vietnamese context of English education. *Language Testing in Asia*, 12(1), Article 27. <https://doi.org/10.1186/s40468-022-00174-5>
- Khodi, A., Khezerlou, H., & Sahraei, H. (2024). Dependability and utility of using e-portfolios in assessing EFL learners' speaking proficiency. *Computer Assisted Language Learning*, 37(7), 1579–1601. <https://doi.org/10.1080/09588221.2022.2093379>
- Kim, H., & Ro, E. (2023). Assessment of sentence sophistication in L2 spoken production: Expansion of verbs and argument structure constructions. *System*, 119, Article 103175. <https://doi.org/10.1016/j.system.2023.103175>
- Kim, H., Song, C., Kim, J., Jeong, H., & Park, J. (2024). Validation of an elicited Imitation test as a measure of Korean language proficiency. *Language Testing in Asia*, 14, Article 17. <https://doi.org/10.1186/s40468-024-00287-z>
- Kirkpatrick, R., Vafadar, H., & Mohebbi, H. (2024). A scoping review of willingness to communicate in language education: premises to doubt, lessons to learn, and future research questions to examine. *Language Testing in Asia*, 14, 9. <https://doi.org/10.1186/s40468-024-00284-2>
- Knoch, U., Fairbairn, J., & Jin, Y. (2021). *Scoring second language spoken and written performance: Issues, options and directions*. University of Toronto Press.
- Koizumi, R. (2021). The Telephone Standard Speaking Test: An outside evaluator's investigation of a rebuttal to the generalization inference. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of argument-based validation research* (pp. 154–175). Cambridge University Press. <https://doi.org/10.1017/9781108669849.010>
- Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142–161. <https://doi.org/10.1080/15434303.2021.2023542>

- Koizumi, R., & In'nami, Y. (2024). Predicting functional adequacy from complexity, accuracy, and fluency of second-language picture-prompted speaking. *System*, 120, 103208. <https://doi.org/10.1016/j.system.2023.103208>
- Koizumi, R., & Watanabe, A. (2021). Rater reliability in classroom speaking assessment in a Japanese senior high school. *Annual Review of English Language Education in Japan*, 32, 129-144. https://doi.org/10.20581/arele.32.0_129
- Koizumi, R., Agawa, T., Asano, K., & In'nami, Y. (2022). Skill profiles of Japanese English learners and reasons for uneven patterns. *Language Testing in Asia*, 12, Article 53. <https://doi.org/10.1186/s40468-022-00203-3>
- Koizumi, R., Fukazawa, M., & Inoue, C. (2024). Usability of a Speaking Assessment Portal for Japanese teachers of English. *SHIKEN: A Journal of Language Testing and Evaluation in Japan*, 28(1), 19-37. <https://doi.org/10.37546/JALTSIG.TEVAL28.1-2>
- Koizumi, R., In'nami, Y., & Jeon, E. H. (2022). L2 speaking and its internal correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 307-338). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self vs. peer assessment activities in EFL-speaking classes: impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia*, 13, Article 36(2023). <https://doi.org/10.1186/s40468-023-00251-3>
- Kyle, K., & Eguchi, M. (2023). Assessing spoken lexical and lexicogrammatical proficiency using features of word, bigram, and dependency bigram use. *Modern Language Journal*, 107, 531–564. <https://doi.org/10.1111/modl.12845>
- Kyle, K., & Eguchi, M. (2024). Evaluating NLP models with written and spoken L2 samples. *Research Methods in Applied Linguistics*, 3(2), Article 100120. <https://doi.org/10.1016/j.rmal.2024.100120>
- Lam, D. M. K. (2021). Don't turn a deaf ear: A case for assessing interactive listening. *Applied Linguistics*, 42(4), 740–764. <https://doi.org/10.1093/applin/amaa064>
- Lam, D. M. K. (2021). Feedback as a learning-oriented assessment practice: Principles, opportunities, and challenges. In A. Gebriel (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 86–107). Routledge. <https://doi.org/10.4324/9781003014102-7>
- Lam, M. K. D., Galaczi E. D., Nakatsuhara, F., & May, L. (2023). Assessing interactional competence: Exploring ratable challenges. *Applied Pragmatics*, 5(2), 208–233. <https://doi.org/10.1075/ap.00014.lam>
- Lambert, C., Aubrey, S., & Leeming, P. (2021). Task preparation and second language speech production. *TESOL Quarterly*, 55(2), 331–365. <https://doi.org/10.1002/tesq.598>
- Lampropoulou, L. (2023). The use and impact of pre-task planning time in the monologic task of LanguageCert speaking tests. *Language Education & Assessment*, 6(1), 1–18. <https://doi.org/10.29140/lea.v6n1.1180>
- Le Bouthillier, J., & Kristmanson, P. (2023). Becoming a French second language teacher: Supporting confidence and competence. *Second Language Teacher Education*, 2(1), 21–42. <https://doi.org/10.1558/slte.24011>
- Lee, H-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring Interaction in Video-call Paired Speaking Tests: A Look at Scores, Language, and Perceptions, *Applied*

- Linguistics, 45(3), 533–552. <https://doi.org/10.1093/applin/amad046>
- Lee, H. (2023). Looking into an innovative test mode in paired speaking from the perspective of scores. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 13–32). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Lee, H.-W., & Galaczi, E. (Eds.). (2023). *Exploring video-call paired speaking tests*. Research Notes, 86. Retrieved from: <https://www.cambridgeenglish.org/english-research-group/published-research/research-notes/>
- Lee, H.-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring interaction in video-call paired speaking tests: A look at scores, language, and perceptions. *Applied Linguistics*, 45(3), 533–552. <https://doi.org/10.1093/applin/amad046>
- Lee, H., & Galaczi, E. (2023). Insights into video-call paired speaking tests: Research background and design. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 5–12). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Lee, H., Patel, M., Lynch, J., & Galaczi, E. D. (2021). Development of the IELTS Video Call Speaking Test: Phase 4 operational research trial and overall summary of a four-phase test development cycle. *IELTS Partnership Research Papers, 2021/1*. IELTS Partners: British Council, Cambridge Assessment English and IDP: IELTS Australia. www.ielts.org/teaching-and-research/research-reports
- Leonard, K., & Bonilla, N. (2022). Teachers' oral proficiency: what happens after they graduate? *Foreign Language Annals*, 55(4), 1152–1168. <https://doi.org/10.1111/flan.12657>
- Li, S., Wen, T., Li, X., Feng, Y., & Lin, C. (2023). Comparing holistic and analytic marking methods in assessing speech act production in L2 Chinese. *Language Testing*, 40(2), 249–275. <https://doi.org/10.1177/02655322221113917>
- Liu, G. (2022). Schema association training: Empowering students for speaking assessments. *RELC Journal*, 53(1), 216–224. <https://doi.org/10.1177/0033688220927099>
- Liu, T. & Aryadoust, V. (2022). The effect of in-class and one-on-one video feedback on EFL learners' English public speaking competency and anxiety. *Studies in Language Assessment*, 11(2), 25–57. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4396264/SiLA_11.2_Liu_and_Aryadoust.pdf
- Lv, Z. (2024). Prompt effects on assessment of interactional competence in a group oral test. *International Journal of Applied Linguistics*, 34(3), 1089–1107. <https://doi.org/10.1111/ijal.12554>
- Ma, W. (2022). What the analytic versus holistic scoring of international teaching assistants can reveal: Lexical grammar matters. *Language Testing*, 39(2), 239–264. <https://doi.org/10.1177/02655322211040020>
- Ma, W., & Winke, P. (2022). An investigation of the impact of jagged profile on L2 speaking test ratings: Evidence from rating and eye-tracking data. *Language Assessment Quarterly*, 19(4), 394–421. <https://doi.org/10.1080/15434303.2022.2078720>
- Mabrouk, F. (2023). Tunisian EFL learners' perceptions of the use of oral presentations in assessing speaking skills: An exploratory practice study. *Studies in Applied Linguistics and TESOL*, 23(3), 18–40. <https://doi.org/10.52214/salt.v23i2.12083>

- Martin, J. L., & Wright, K. E. (2023). Bias in automatic speech recognition: The case of African American language. *Applied Linguistics*, 44(4), 613–630. <https://doi.org/10.1093/applin/amac066>
- McDonough, K., & Uludag, P. (2021). Individual and shared assessment of ESL students' paired oral test performance: Examining rater judgments and lexico-grammatical features. In W. Crawford (Ed.), *Multiple perspectives on learner interaction: The corpus of collaborative oral tasks* (pp. 69–90). Mouton De Gruyter. <https://doi.org/10.1515/9781501511370-005>
- Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking anxiety and task complexity effects on second language speech. *International Journal of Applied Linguistics*, 34(1), 292–315. <https://doi.org/10.1111/ijal.12494>
- Mostafa, T., Crossley, S. A., & Kim, Y. (2021). Predictors of English as second language learners' oral proficiency development in a classroom context. *International Journal of Applied Linguistics*, 31(3), 526–548. <https://doi.org/10.1111/ijal.12358>
- Mullooly, A. (2023). Operational considerations for video-call delivery of paired format speaking tests. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 87–96). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Mullooly, A., & Glasson, N. (2023). Investigating language functions elicited by in-person and online paired speaking tests. In E. Galaczi, & H. Lee (Eds.), *Exploring video-call paired speaking tests* (pp. 33–52). Cambridge. <https://www.cambridgeenglish.org/Images/702820-research-notes-86.pdf>
- Murphy, D., & Sedivy, S. K. (2024). The speaking proficiency outcomes of face-to-face and online intensive summer LCTL programs. *Foreign Language Annals*, 57(4), 872–899. <https://doi.org/10.1111/flan.12758>
- Nagle, C. (2022). Rethinking pronunciation posttesting. *Journal of Second Language Pronunciation*, 8(2), 161–167. <https://doi.org/10.1075/jslp.22019.nag>
- Nagle, C. L., Huensch, A., & Zárate-Sánchez, G. (2023). Exploring phonetic predictors of intelligibility, comprehensibility, and foreign accent in L2 Spanish speech. *Modern Language Journal*, 107, 202–221. <https://doi.org/10.1111/modl.12827>
- Nagle, C., & Rehman, I. (2021). Doing L2 speech research online: Why and how to collect online ratings data. *Studies in Second Language Acquisition*, 43(4), 916–939. <https://doi.org/10.1017/S0272263121000292>
- Nakatsuhara, F., Inoue, C., & Khabbazzbashi, N. (2021). Measuring L2 speaking. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.285–294). Routledge. <https://doi.org/10.4324/9781351034784-31>
- Nakatsuhara, F., Inoue, C., & Taylor, L. (2021). Comparing rating modes: Analysing live, audio and video ratings of IELTS Speaking Test performance. *Language Assessment Quarterly*, 18(2), 83–106. <https://doi.org/10.1080/15434303.2020.1799222>
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2021). Video-conferencing speaking tests: Do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice*, 28(4), 369–388. <https://doi.org/10.1080/0969594X.2021.1951163>
- Nakatsuhara, F., May, L., & Khabbazzbashi, N. (2023). *Exploring the speaking construct in academic settings in a digital age*. British Council, Cambridge Assessment English,

- and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262806705843094>
- Nakatsuhara, F., May, L., Inoue, C., Willcox-Ficzere, E., Westbrook, C., & Spiby, R. (2021). *Exploring the potential for assessing interactional and pragmatic competence in semi-direct speaking tests* (Report No. VS/2021/001). British Council.
https://www.britishcouncil.org/sites/default/files/nakatsuhara_et_al_2021_-_assessing_interactional_and_pragmatic_competence_in_semi-direct_speaking_tests_v2.pdf
- Neittaanmäki, R., & Lamprianou, I. (2024). Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, 41(3), 584–605. <https://doi.org/10.1177/02655322241239363>
- Nguyen, T. H. H., Nguyen, B. T. T., Hoang, G. T. L., Pham, N. T. H., & Dang, T. T. C. (2024). Computer-delivered vs. face-to-face score comparability and test takers' perceptions: The case of the two English speaking proficiency tests for Vietnamese EFL learners. *Language Testing in Asia*, 14, Article 6.
<https://doi.org/10.1186/s40468-024-00277-1>
- O'Sullivan, B. (2024). Assessing Speaking. In Antony Kunnan (ed), *The Concise Companion to Language Assessment*, (pp.143-153). Wiley. DOI: 10.1002/9781118411360.wbcla084
- Ockey, G. J., & Chukharev-Hudilainen, E. (2021). Human versus computer partner in the paired oral discussion test. *Applied Linguistics*, 42(5), 924–944.
<https://doi.org/10.1093/applin/amaa067>
- Ockey, G. J., & Neiriz, R. (2021). Evaluating technology-mediated second language oral communication assessment delivery models. *Assessment in Education: Principles, Policy & Practice*, 28(4), 350–368. <https://doi.org/10.1080/0969594X.2021.1976106>
- Ockey, G. J., Muhammad, A. A., Prasetyo, A. H., Elnegahy, S., Kochem, T., Neiriz, R., Kim, H., & Beck, J. (2021). Iowa State University's English placement test of oral communication in times of COVID-19. *Language Assessment Quarterly*, 18(1), 26–35. <https://doi.org/10.1080/15434303.2020.1862122>
- Ockey, G. J., Vo, S., & Baghestani, S. (2021). Establishing appropriate cut scores of a standardized test for a local placement context. *TESOL Quarterly*, 55(3), 719–742.
<https://doi.org/10.1002/tesq.3007>
- Palmour, L. (2023). Assessing oral presentations: An analysis of score-reaching dialogue between EAP practitioners. *Journal of English for Academic Purposes*, 62, Article 101210. <https://doi.org/10.1016/j.jeap.2022.101210>
- Palmour, L. (2024). Assessing speaking through multimodal oral presentations: The case of construct underrepresentation in EAP contexts. *Language Testing*, 41(1), 9–34.
<https://doi.org/10.1177/02655322231183077>
- Panahi, A., Mohebbi, H., & Oxford, R. L. (2024). Rebecca Oxford's 50 years of contributions to language education and related fields: A systematic review. *Language Teaching Research Quarterly*, 41, 169–230.
- Pang, F., & Skehan, P. (2021). Performance profiles on second language speaking tasks. *The Modern Language Journal*, 105(1), 371–390. <https://doi.org/10.1111/modl.12699>
- Pérez Castillejo, S. (2023). Prior processing, foreign language classroom anxiety, and L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 61(2), 519-544. <https://doi.org/10.1515/iral-2021-0091>
- Perry, S. (2024). Podcasting as a form of assessment: Increasing student motivation in academic English-speaking assessment. *Language Testing in Asia*, 14, Article 15.

- <https://doi.org/10.1186/s40468-024-00288-y>
- Pesau, H. G., Immanuel, A. S., Sulastri, A., & van Lijstelaar, G. (2023). The role of daily spoken language on the performance of language tests: The Indonesian experience. *Bilingualism: Language and Cognition*, 26(3), 538-549. <https://doi.org/10.1017/S136672892200075X>
- Purpura, J.E (2021). A rationale for using a scenario-based assessment to measure competency-based, situated second and foreign language proficiency. In M. Masperi, C Cervini, Y Bardière (Eds.), *Évaluation des acquisitions langagières: du formatif au certificatif* (pp. 54-96). MediAzioni. <http://www.mediazioni.sitlec.unibo.it>.
- Quaid, E., & Barrett, A. (2021). Interpretations of spoken utterance fluency in simulated and face-to-face oral proficiency interviews. *Language Education & Assessment*, 4(1), 1-18. <https://doi.org/10.29140/lea.v4n1.385>
- Rafi, F., & Pourdana, N. (2023). E-diagnostic assessment of collaborative and individual oral tiered task performance in differentiated second language instruction framework. *Language Testing in Asia*, 13, Article 6 (2023). <https://doi.org/10.1186/s40468-023-00223-7>
- Rassaei, E. (2023). Implementing mobile-mediated dynamic assessment for teaching request forms to EFL learners. *Computer Assisted Language Learning*, 36(3), 257–287. <https://doi.org/10.1080/09588221.2021.1912105>
- Rezadoust Siah Khaleh Sar, H., Khalili Sabet, M., Mahdavi Zafarghandi, A., & Hassaskhah, J. (2024). Iranian teachers' and learners' attitudes towards the use of e-portfolio in second language speaking assessment. *Language Testing in Asia*, 14, Article 61. <https://doi.org/10.1186/s40468-024-00335-8>
- Ritonga, M., Farhangi, F., Ajanil, B., & Farid Khafaga, A. (2022). Interventionist vs. interactionist models of dynamic assessment (DA) in the EFL classroom: impacts on speaking accuracy and fluency (SAF), foreign language classroom anxiety (FLCA), and foreign language learning motivation (FLLM). *Language Testing in Asia*, 12, Article 43. <https://doi.org/10.1186/s40468-022-00195-0>
- Roever, C., & Ikeda, N. (2022). What scores from monologic speaking tests can(not) tell us about interactional competence. *Language Testing*, 39(1), 7–29. <https://doi.org/10.1177/02655322211003332>
- Roever, C., & Ikeda, N. (2024). The relationship between L2 interactional competence and proficiency. *Applied Linguistics*, 45(4), 676-698. <https://doi.org/10.1093/applin/amad053>
- Rousse-Malpat, A., Koote, L., Steinkrauss, R., & Verspoor, M. (2024). Parlez-vous francais? Effects of structure-based versus dynamic-usage-based approaches on oral proficiency. *Language Teaching Research*, 28(5), 1893-1917. <https://doi.org/10.1177/13621688211040298>
- Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, 55(3), 866–900. <https://doi.org/10.1002/tesq.3027>
- Saito, K., & Liu, Y. (2022). Roles of collocation in L2 oral proficiency revisited: Different tasks, L1 vs. L2 raters, and cross-sectional vs. longitudinal analyses. *Second Language Research*, 38(3), 531–554. <https://doi.org/10.1177/0267658320988055>
- Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A

- test-retest reliability study. *Studies in Second Language Acquisition*, 46(4), 1206–1230. <https://doi.org/10.1017/S027226312200047X>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14, Article 49. <https://doi.org/10.1186/s40468-024-00321-0>
- Sayyadi, A., & Rezvani, R. (2021). Questioning in TOEFL iBT speaking test: A case of washback and construct underrepresentation. *Language Testing in Asia*, 11, Article 21. <https://doi.org/10.1186/s40468-021-00137-2>
- Schmidgall, J., & Powers, D. E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC® Speaking test scores from linguistic laypersons. *Language Testing*, 38(2), 302–325. <https://doi.org/10.1177/0265532220941803>
- Seedhouse, P., & Satar, M. (2021). VEO IELTS Project Report: Which specific features of candidate talk do examiners orient to when taking scoring decisions? *IELTS Research Reports Online Series*, 5. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/for-researchers/research-reports/online-series-2021-5>
- Shehata, A. (2024). Arabic pronunciation instruction: Learner cognition and speech comprehensibility. *Journal of Second Language Pronunciation*, 10(2), 179–203. <https://doi.org/10.1075/jslp.23015.she>
- Shi, X., Wang, X., & Zhang, W. (2024). Exploring the relationships between ASS indices and CAF and the impact on Chinese college students' oral English performance. *Language Testing in Asia*, 14, Article 30. <https://doi.org/10.1186/s40468-024-00306-z>
- Shin, J. (2022). Investigating and optimizing score dependability of a local ITA speaking test across language groups: A generalizability theory approach. *Language Testing*, 39(2), 313–337. <https://doi.org/10.1177/02655322211052680>
- Skehan, P., Bui, G., Wang, Z., & Shum, S. (2024). Re-examining accuracy measures in second language task-based spoken performance. *Research Methods in Applied Linguistics*, 3(1), Article 100098. <https://doi.org/10.1016/j.rmal.2024.100098>
- Spino, L. L., Echevarría, M. M., & Wu, Y. (2022). Assisted self-assessment to optimize the OPIc test experience. *Foreign Language Annals*, 55(3), 853–876. <https://doi.org/10.1111/flan.12645>
- Stansfield, C.W., Gao, J., & Rivers, W.P. (2010). A concurrent validity study of self-assessments and the federal interagency roundtable oral proficiency interview. *Russian Language Journal / Русский Язык*, 60, 299–315. <http://www.jstor.org/stable/43669189>
- Stephenson, M. & Hall, G. (2021). Organizing talk in group speaking tests: Learning from high-scoring students. *ELT Journal*, 75(1), 42–54. <https://doi.org/10.1093/elt/ccaa052>
- Sundqvist, P., & Uztosun, M. S. (2024). Extramural English in Scandinavia and Asia: scale development, learner engagement, and perceived speaking ability. *TESOL Quarterly*, 58(4), 1638–1665. <https://doi.org/10.1002/tesq.3296>
- Sureeyatanapas, P., Sureeyatanapas, P., Panitanarak, U., Kraisriwattana, J., Sarootyanapat, P., & O'Connell, D. (2024). The analysis of marking reliability

- through the approach of gauge repeatability and reproducibility (GR&R) study: A case of English-speaking test. *Language Testing in Asia*, 14, Article 1. <https://doi.org/10.1186/s40468-023-00271-z>
- Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, 105(2), 435–463. <https://doi.org/10.1111/modl.12706>
- Suzumura, N. (2022). Content analysis of test taker responses on an AP Japanese computer-simulated conversation test: A mixed methods approach for a validity argument. *Language Assessment Quarterly*, 19(5), 480-502. <https://doi.org/10.1080/15434303.2022.2130326>
- Taguchi, N. (2022). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, 56(1), 308-335. <https://doi.org/10.1002/tesq.3070>
- Tanaka, M., & Ross, S. J. (2023). Impact of self-construal on rater severity in peer assessments of oral presentations. *Assessment in Education: Principles, Policy & Practice*, 30(2), 203-220. <https://doi.org/10.1080/0969594X.2023.2189564>
- Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s. *System*, 100, Article 102567. <https://doi.org/10.1016/j.system.2021.102567>
- Thai, T., & Sheehan, S. (2022). The processes of rating L2 speaking performance using an analytic rating scale – A qualitative exploration. *Language Education and Assessment*, 5(1), 34–51. <https://doi.org/10.29140/lea.v5n1.777>
- Thomson, R. I., & Isaacs, T. (2022). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research* (pp. 24-44). Springer.
- Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2024). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*, 37(1-2), 149-178. <https://doi.org/10.1080/09588221.2022.2032184>
- Torregrossa, J., Listanti, A., Bongartz, C., & Narinis T. (2024). Adding discourse to sentence repetition tasks: Under which conditions does bilingual children's performance improve?. *Research Methods in Applied Linguistics*, 3(2), Article 100107. <https://doi.org/10.1016/j.rmal.2024.100107>
- Trofimovich, P., Isaacs, T., Kennedy, S., & Tsunemoto, A. (2022). Speech comprehensibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and L2 speaking* (pp. 174-187). Routledge.
- Trofimovich, P., Tekin, O., & McDonough, K. (2021). Task engagement and comprehensibility in interaction: Moving from what second language speakers say to what they do. *Journal of Second Language Pronunciation*, 7(3), 435–461. <https://doi.org/10.1075/jslp.21006.tro>
- Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Language Annals*, 55(1), 135–154. <https://doi.org/10.1111/flan.12571>
- Tsunemoto, A., Uludag, P., McDonough, K., & Isaacs, T. (2023). Research forum: The relationship between Japanese EFL learners' perceived fluency and temporal speech measures in a read-aloud task. *JALT Journal*, 45, 107–121.

- <https://doi.org/10.37546/JALTJJ45.1-5>
- Turner, C. E. (2023). Assessment and evaluation: mixed methods research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Wiley Online Library. <https://doi.org/10.1002/9781405198431.wbeal0040.pub2>
- Uludag, P., McDonough, K., & Trofimovich, P. (2022). Exploring shared and individual assessment of paired oral interactions. *Studies in Language Assessment*, 11(2), 1–24. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4258357/2022-Early-View_Uludag-et-al.pdf
- Vandeweerd, N., Housen, A., & Paquot, M. (2023). Proficiency at the lexis–grammar interface: Comparing oral versus written French exam tasks. *Language Testing*, 40(3), 658–683. <https://doi.org/10.1177/02655322231153543>
- Viera, C., & Arispe, K. (2022). Can Spanish programs evaluate oral proficiency gains for all learners? An empirical analysis of traditional and nontraditional Spanish majors' oral proficiency development. *Foreign Language Annals*, 55(3), 832– 852. <https://doi.org/10.1111/flan.12635>
- Wang, Z., & Zhang, J. (2023). Mediation and learner reciprocity: Applying dynamic assessment to the oral proficiency interview framework. *Language and Sociocultural Theory*, 10(1), 82–105. <https://doi.org/10.1558/lst.22181>
- Watson, S. (2022). *Formative speaking assessment: Beginning with story* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Watzinger-Tharp, J., Tharp, D. S., & Rubio, F. (2021). Sustaining dual language immersion: Partner language outcomes in a statewide program. *The Modern Language Journal*, 105(1), 194–217. <https://doi.org/10.1111/modl.12694>
- Weisi, H., & Zandi, M. (2024). A mixed-method concurrent transformative study of metadiscourse markers employed by L2 speakers: Does proficiency level matter? *Language Testing in Asia*, 14, Article 60. <https://doi.org/10.1186/s40468-024-00330-z>
- Winke, P., Zhang, X., & Pierce, S. (2023). A closer look at a marginalized test method: Self-assessment as a measure of speaking proficiency. *Studies in Second Language Acquisition*, 45(2), 416–441. doi:10.1017/S0272263122000079
- Wu, C., & Gao, F. (2023). Comparing the Intelligibility of Different Varieties of English Through Predicting Learners Comprehension: A Phonetic Experimental Approach. *Chinese Journal of Applied Linguistics*, 46(1), 19–30. <https://doi.org/10.1515/CJAL-2023-0102>
- Xi, X., & Norris, J. M. (Eds.). (2021). *Assessing English for higher education admissions*. Routledge. <https://doi.org/10.4324/9781351142403>
- Xu, J., & Fan, Y. (2024). Task complexity, L2 proficiency and EFL learners' L1 use in task-based peer interaction. *Language Teaching Research*, 28(2), 346–365. <https://doi.org/10.1177/13621688211004633>
- Xu, J., & Qiu, X. (2024). Engaging L2 Learners in Information-gap Tasks: How Task Type and Topic Familiarity Affect Learner Engagement. *RELC Journal*, 55(1), 29–45. <https://doi.org/10.1177/00336882211061628>
- Xu, J., Jones, E., Laxton, V., & Galaczi, E. (2021). Assessing L2 English speaking using automated scoring technology: Examining automarker reliability. *Assessment in Education: Principles, Policy & Practice*, 28(4), 411–436.

- <https://doi.org/10.1080/0969594X.2021.1979467>
- Yan, X., & Staples, S. (2023). *Investigating the cognitive and social aspects of IELTS speaking performances across proficiency levels: Comparing the CAF-based and register-linguistic analyses*. British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262843971785611>
- Yan, X., Kim, H. R., & Kim, J. Y. (2021). Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency. (CAF). features of speaking performances on the Aptis test. *Language Testing*, 38(4), 485–510.
<https://doi.org/10.1177/0265532220951508>
- Yin, S., Fan, J., Jin, Y., & Stapleton, P. (2024). Towards a framework of critical thinking for assessing EAP speaking. *Journal of English for Academic Purposes*, 1-14.
<https://doi.org/10.1016/j.jeap.2024.101426>
- Yin, S., Fan, J., Jin, Y., & Stapleton, P. (2024). Towards a framework of critical thinking for assessing EAP speaking. *Journal of English for Academic Purposes*, 71, Article No. 101426. <https://doi.org/10.1016/j.jeap.2024.101426>
- Yin, S., Guo, K., & Li, D. (2024). To assign a topic or not: Examining critical thinking in English as a foreign language speaking. *RELC Journal*, 1-14.
<https://doi.org/10.1177/00336882231222651>
- Zabihi, R., Ghominejad, S., & Javad Ahmadian, M. (2024). Can willingness to communicate, communication in English anxiety, behavioural inhibition and behavioural action predict perceived L2 fluency? *Language Teaching Research*, 28(6), 2214-2233.
<https://doi.org/10.1177/13621688211044071>
- Zhang, S., & Isaacs, T. (2023). Can interactions happen across the screens? The use of videoconferencing technology in assessing second language pragmatic competence. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts* (pp. 196–211). Routledge.
- Zhang, W., & Zhang, L. J. (2022). Understanding assessment tasks: Learners' and teachers' perceptions of cognitive load of integrated speaking tasks for TBLT implementation. *System*, 111, Article 102951. <https://doi.org/10.1016/j.system.2022.102951>
- Zhang, W., Zhao, M., & Zhu, Y. (2022). Understanding individual differences in metacognitive strategy use, task demand, and performance in integrated L2 speaking assessment tasks. *Frontiers in Psychology*, 13, Article 876208.
<https://doi.org/10.3389/fpsyg.2022.876208>
- Zhang, X., Zhao, B., & Li, W. (2023). N-gram use in EFL learners' retelling and monologic tasks. *International Review of Applied Linguistics in Language Teaching*, 61(3), 939–965. <https://doi.org/10.1515/iral-2021-0080>
- Zheng, C., Wang., & Chai, C. S. (2023). Self-assessment first of peer-assessment first: effects of video-based formative practice on learners' English public speaking anxiety and performance. *Computer Assisted Language Learning*, 36(4), 806–839.
<https://doi.org/10.1080/09588221.2021.1946562>

Assessing Writing

- Abdi Tabari, M., & Wang, Y. (2022). Assessing linguistic complexity features in L2 writing: Understanding effects of topic familiarity and strategic planning within the realm of task readiness. *Assessing Writing*, 52, Article 100605.
<https://doi.org/10.1016/j.asw.2022.100605>

- Acar, A. S. (2023). Genre pedagogy: A writing pedagogy to help L2 writing instructors enact their classroom writing assessment literacy and feedback literacy. *Assessing Writing*, 56, Article 100717. <https://doi.org/10.1016/j.asw.2023.100717>
- Agrawal, S., Oza, P., Kakkar, R., Tanwar, S., Jetani, V., Undhad, J., & Singh, A. (2024). Analysis and recommendation system-based on PRISMA checklist to write systematic review. *Assessing Writing*, 61, Article 100866. <https://doi.org/10.1016/j.asw.2024.100866>
- Akbarian, I. H., & Farrokhi, J. (2021). Lexical availability and writing ability of EFL learners. *International Journal of Applied Linguistics*, 31(1), 48–64. <https://doi.org/10.1111/ijal.12319>
- Al Sharoufi, H. (2022). Towards a unified English technology-based writing curriculum in the Arabian Gulf countries: the case of Oman. *Language Testing in Asia*, 12, Article 33. <https://doi.org/10.1186/s40468-022-00178-1>
- Al-Hawamdeh, B. O. S., Hussien, N., & Abdelrasheed, N. S. G. (2023). Portfolio vs. summative assessment: Impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13, Article 12. <https://doi.org/10.1186/s40468-023-00225-5>
- Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing*, 52, Article 100610. <https://doi.org/10.1016/j.asw.2022.100610>
- Allal, L. (2021). Involving primary school students in the co-construction of formative assessment in support of writing. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 584–601. <https://doi.org/10.1080/0969594X.2021.1951164>
- Alves-Wold, A., Walgermo, B. R., & Foldnes, N. (2024). Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses? *Assessing Writing*, 60, Article 100833. <https://doi.org/10.1016/j.asw.2024.100833>
- An, H., & Li, S. (2024). Task-specific writing anxiety and self-efficacy are separate from general L2 writing anxiety and self-efficacy and they have differential associations with the effects of written corrective feedback in pre-task and within-task planning. *System*, 126, Article 103480. <https://doi.org/10.1016/j.system.2024.103480>
- Anderson, R. C., Chaparro, E. A., Smolkowski, K., & Cameron, R. (2023). Visual thinking and argumentative writing: A social-cognitive pairing for student writing development. *Assessing Writing*, 55, Article 100694. <https://doi.org/10.1016/j.asw.2023.100694>
- Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, 110, Article 102899. <https://doi.org/10.1016/j.system.2022.102899>
- Appel, R., Arias, A., Baker, B., & Loignon, G. (2023). Insights from lexical and syntactic analyses of a French for academic purposes assessment. *Assessing Writing*, 58, Article 100789. <https://doi.org/10.1016/j.asw.2023.100789>
- Archila, P. A., Molina, J., Danies, G., Truscott de Mejía, A.-M., & Restrepo, S. (2022). Using formal formative assessment (FFA) to promote undergraduates' bilingual written scientific argumentation (BWSA). *International Journal of Bilingual Education and Bilingualism*, 25(9), 3137–3150. <https://doi.org/10.1080/13670050.2021.2021386>
- Arias-Hermoso, R., Agirre, A. I., & Larrañaga, E. G. (2024). A comparison between input modalities and languages in source-based multilingual argumentative writing. *Assessing Writing*, 60, 100813. <https://doi.org/10.1016/j.asw.2024.100813>
- Arques, A. C., & Ferrero, C. L. (2023). Peer-feedback of an occluded genre in the Spanish language classroom: A case study. *Assessing Writing*, 57, Article 100756. <https://doi.org/10.1016/j.asw.2023.100756>
- Aryadoust, V., & Yichen, J. (2025). ChatGPT in assessment of writing. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from <https://meteor.springer.com/container/contribute.jsf?id=263243>

- Atak, N., & Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. *Assessing Writing*, 47, Article 100506. <https://doi.org/10.1016/j.asw.2020.100506>
- Aull, L. (2021). Directed Self-Placement: Subconstructs and group differences at a US university. *Assessing Writing*, 48, Article 100522. <https://doi.org/10.1016/j.asw.2021.100522>
- Baker, B. & Gebril, A. (2024). Editorial: The assessment of writing in languages other than English. *Assessing Writing*, 60, 1-2. <https://doi.org/10.1016/j.asw.2024.100840>.
- Baker, B., & Gebril, A. (2024). The assessment of writing in languages other than English (LOTE). *Assessing Writing*, 60, 100840. <https://doi.org/10.1016/j.asw.2024.100840>
- Baker, B., Polikar, S., & Homayounzadeh, M. (2021). "I was not thoughtful enough before." Exploiting the learning-oriented-assessment potential of a test taker-oriented rubric for summative assessment. In A. Gebril. (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 162–181). Routledge. <https://doi.org/10.4324/9781003014102-12>
- Barkaoui, K. (2021). *Evaluating tests of second language development*. Peter Lang. <https://doi.org/10.3726/b18059>
- Barkaoui, K. (2023). Exploring task effects on register variation in second language learners' writing. *TASK: Journal on Task-Based Language Teaching and Learning*, 3, 109-139. <https://doi.org/10.1075/task.22009.bar>
- Barkaoui, K. (2024). Examining performance on an integrated writing task from a Canadian English language proficiency test. *Canadian Modern Language Review*, 80(2), 77-115. <https://doi.org/10.3138/cmlr-2023-0022>
- Barkaoui, K. (2024). Exploring the effects of task difficulty and learner variables on performance on picture description writing tasks. *Assessing Writing*, 60, 100827. <https://doi.org/10.1016/j.asw.2024.100827>
- Barkaoui, K. & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26-63. 2._sila_12_1__barkaoui_and_woodworth.pdf (altaanz.org)
- Barkaoui, K., & Hadidi, A. (2021). *Assessing changes in second Language writing performance*. Routledge. <https://doi.org/10.4324/9781003092346>
- Barkaoui, K., & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26–63. <https://doi.org/10.58379/QCFS2805>
- Barkaoui, K., Shi, L., & Riazzi, M. (2021). "Follow your interests because those will motivate you to excel": An interview with Alister Cumming. *Language Assessment Quarterly*, 18(1), 56–75. <https://doi.org/10.1080/15434303.2020.1846190>
- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing*, 47, Article 100510. <https://doi.org/10.1016/j.asw.2020.100510>
- Bastola, M. N., & Hu, G. (2024). Engagement with supervisory feedback on master's theses: Do supervisors and students see eye to eye? *Assessing Writing*, 60, 100841. <https://doi.org/10.1016/j.asw.2024.100841>
- Baxter, G., & Neumann, H. (2023). Source-based writing of the high- and low-proficiency adolescent writer in the high-school L2 classroom. *Journal of Second Language Writing*, 62, Article 101064. <https://doi.org/10.1016/j.jslw.2023.101064>
- Beltrán-Palanques, V. (2024). Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP. *Assessing Writing*, 60, 100809. <https://doi.org/10.1016/j.asw.2024.100809>

- Bennett, R. E., Zhang, M., Sinharay, S., Guo, H., & Deane, P. (2022). Are there distinctive profiles in examinee essay-writing processes? *Educational Measurement: Issues and Practice*, 41(2), 55–69. <https://doi.org/10.1111/emip.12469>
- Berberovic, V. (2022). *Implementing peer feedback for writing tasks* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Boggs, J. A., & Manchón, R. M. (2023). Feedback literacy in writing research and teaching: Advancing L2 WCF research agendas. *Assessing Writing*, 58, Article 100786. <https://doi.org/10.1016/j.asw.2023.100786>
- Bondar, N. A., Day, S. X., Mountain, L., Turchi, L. B., & Hutchison, L. F. (2023). Composition Organization and Development Analysis (CODA) Scale: Equipping high school students to evaluate argumentative essays. *Assessing Writing*, 57, Article 100724. <https://doi.org/10.1016/j.asw.2023.100724>
- Bouwer, R., Koster, M., & Van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*, 30(3-4), 302-319. <https://doi.org/10.1080/0969594X.2023.2241656>
- Britton, E. R. (2023). Developing teacher feedback literacy through self-study: Exploring written commentary in a critical language writing curriculum. *Assessing Writing*, 56, Article 100709. <https://doi.org/10.1016/j.asw.2023.100709>
- Cai, H., & Yan, X. (2024). Examining the direct and indirect impacts of verbatim source use on linguistic complexity in integrated argumentative writing assessment. *Assessing Writing*, 61, 100868. <https://doi.org/10.1016/j.asw.2024.100868>
- Cai, H., & Yan, X. (2024). Triangulating natural language processing (NLP)-based analysis of rater comments and many-facet Rasch measurement (MFRM): An innovative approach to investigating raters' application of rating scales in writing assessment. *Language Testing*, 41(2), 384–411. <https://doi.org/10.1177/02655322231210231>
- Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, Article 102464. <https://doi.org/10.1016/j.system.2021.102464>
- Carr, N. (2023). Feedback on writing through the lens of activity theory: An exploration of changes to peer-to-peer interactions. *Assessing Writing*, 56, Article 100720. <https://doi.org/10.1016/j.asw.2023.100720>
- Carrió-Pastor, M. L. (2021). The assessment of metadiscourse devices in English as a foreign language. *Assessing Writing*, 50, Article 100560. <https://doi.org/10.1016/j.asw.2021.100560>
- Castillo, P. A. P., Soto, C., Asún, R. A., & Gutiérrez, F. (2023). Profiling support in literacy development: Use of natural language processing to identify learning needs in higher education. *Assessing Writing*, 58, Article 100787. <https://doi.org/10.1016/j.asw.2023.100787>
- Cen, Y., & Zheng, Y. (2024). The motivational aspect of feedback: A meta-analysis on the effect of different feedback practices on L2 learners' writing motivation. *Assessing Writing*, 59, 100802. <https://doi.org/10.1016/j.asw.2023.100802>
- Chan, S., & Yamashita, J. (2022). Integrated writing and its correlates: A meta-analysis. *Assessing Writing*, 54, Article 100662. <https://doi.org/10.1016/j.asw.2022.100662>
- Chang, T. (2023). A move analysis of Chinese L2 student essays from the sociocognitive perspective: Genres, languages, and writing quality. *Assessing Writing*, 57, Article 100750. <https://doi.org/10.1016/j.asw.2023.100750>
- Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*, 57, Article 100915. <https://doi.org/10.1016/j.jslw.2022.100915>
- Cheong, C. M., Liu, Y., & Mu, R. (2024). Characteristics of students' task representation and its association with argumentative integrated writing performance. *Assessing Writing*, 60, 100845. <https://doi.org/10.1016/j.asw.2024.100845>

- Cheong, C. M., Yao, Y., & Zhang, J. (2023). Growth mindset and emotions in tandem: Their effects on L2 writing performance based on writers' proficiency levels. *Assessing Writing*, 58, Article 100785. <https://doi.org/10.1016/j.asw.2023.100785>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Choi, I., & Deane, P. (2021). Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly*, 18(2), 107–132. <https://doi.org/10.1080/15434303.2020.1804913>
- Choi, I., Hao, J., Deane, P., & Zhang, M. (2021). Benchmark keystroke biometrics accuracy from high-stakes writing tasks. *ETS Research Report Series*, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12326>
- Choi, Y. (2021). What interpretations can we make from scores on graphic-prompt writing (GPW) tasks? An argument-based approach to test validation. *Assessing Writing*, 48, Article 100523. <https://doi.org/10.1016/j.asw.2021.100523>
- Choi, Y. D. (2022). Validity of score interpretations on an online English placement writing test. *Language Testing in Asia*, 12, Article 42. <https://doi.org/10.1186/s40468-022-00187-0>
- Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, 60, Article 101003. <https://doi.org/10.1016/j.jslw.2023.101003>
- Chuang, P., & Yan, X. (2022). An investigation of the relationship between argument structure and essay quality in assessed writing. *Journal of Second Language Writing*, 56, Article 100892. <https://doi.org/10.1016/j.jslw.2022.100892>
- Clark, T., & Yu, G. (2021). Beyond the IELTS test: Chinese and Japanese postgraduate UK experiences. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1512–1530. <https://doi.org/10.1080/13670050.2020.1829538>
- Clark, T., & Yu, G. (2024). The pedagogical remit of test preparation: the case of writing acquisition on an IELTS course. *Applied Linguistics Review*, 15(3), 1009–1045. <https://doi.org/10.1515/applirev-2021-0164>
- Consolini, C. H., & Kyle, K. (2024). The relationship between L2 Spanish proficiency and features of written lexical and lexicogrammatical use. *Applied Linguistics*, 45(6), 992–1009. <https://doi.org/10.1093/applin/amae032>
- Cook, B., Luke, J., Valoe, A., & Barkaoui, K. (2021). Institutional language policy and ESL teachers' conceptions of L2 writing and classroom assessment practices. *Canadian Modern Language Review*, 77(2), 93–109. <https://doi.org/10.3138/cmlr-2020-0013>
- Crossley, S. A., Baffour, P., Tian, Y., Franklin, A., Benner, M., & Boser, U. (2023). A large-scale corpus for assessing written argumentation: PERSUADE 2.0. *Assessing Writing*, 61, 100865. <https://doi.org/10.1016/j.asw.2024.100865>
- Crusan, D. (2023). Shaping teachers' writing assessment practices: The effect of teacher beliefs. In R. Kohls & C. P. Casanave (Eds.), *Perspectives on good writing in Applied Linguistics and TESOL* (pp. w60-74). University of Michigan Press. <http://doi.org/10.3998/mpub.11930285>
- Crusan, D. (2024). Ungrading: Revolution or evolution. *Journal of Second Language Writing*, 66, 1-7. <https://doi.org/10.1016/j.jslw.2024.101149>
- Crusan, D., Panahi, A., & Mohebbi, H. (2023). Writing teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 45-55). Castledown. <https://castledown.online/reference/9781914291159/>
- Cui, Y., Jin, H., & Gao, Y. (2023). Developing EFL teachers' feedback literacy for research and publication purposes through intra- and inter-disciplinary collaborations: A

- multiple-case study. *Assessing Writing*, 57, Article 100751.
<https://doi.org/10.1016/j.asw.2023.100751>
- DE JONG, J. H. A. L. (2023). Reading: A precondition for Writing. In R. Horowitz (Ed.), *The Routledge international handbook of research on writing* (2nd. ed., pp. 556–562). Routledge.
- Deane, P., Yan, D., Castellano, K., Attali, Y., Lamar, M., Zhang, M., Blood, I., Bruno, J. V., Li, C., Cui, W., Ruan, C., Appel, C., James, K., Long, R., & Qureshi, F. (2024). Modeling writing traits in a formative essay corpus. *ETS Research Report Series*, 2024(1), 1–62. <https://doi.org/10.1002/ets2.12377>
- deBoer, M., Leontjev, D., & Friederich, L. (2023). From language to function: developing self-and peer-assessment tools. *ELT Journal*, 77(1), 94–104.
<https://doi.org/10.1093/elt/ccac014>
- del Mar Sánchez-Pérez, M. (2023). The impact of EMI on student English writing proficiency in a Spanish undergraduate engineering context. *Studies in Second Language Learning and Teaching*, 13(2), 373–397.
<https://doi.org/10.14746/ssllt.38279>
- Deng, Z., Uccelli, P., & Snow, C. (2022). Diversity of Advanced Sentence Structures (DASS) in writing predicts argumentative writing quality and receptive academic language skills of fifth-to-eighth grade students. *Assessing Writing*, 53, Article 100649.
<https://doi.org/10.1016/j.asw.2022.100649>
- Derakhshan, A., & Fathi, J. (2024). Growth mindset, self-efficacy, and self-regulation: A symphony of success in L2 speaking. *System*, 123, 103320.
<https://doi.org/10.1016/j.system.2024.103320>
- DeVore, S., & Kyle, K. (2023). Assessing syntactic and lexicogrammatical use in second language Mandarin writing samples. *Journal of Second Language Writing*, 60, Article 101014. <https://doi.org/10.1016/j.jslw.2023.101014>
- Di Zhang, E., & Yu, S. (2023). Investigating the relationship between linguistic changes in L2 writers' paraphrasing, paraphrasing performance and L2 proficiency. *Applied Linguistics Review*, 14(5), 1451–1473. <https://doi.org/10.1515/applirev-2021-0014>
- Dobrić, N. (2024). Effects of errors on ratings of writing performances—Evidence from a high-stakes exam. *Assessing Writing*, 59, 100806.
<https://doi.org/10.1016/j.asw.2023.100806>
- Dobrić, N., Sigott, G., Ilc, G., Lazović, V., Cesnik, H., & Stopar, A. (2021). Errors as indicators of writing task difficulty at the Slovene general matura in English. *International Journal of Applied Linguistics*, 31(3), 475–491.
<https://doi.org/10.1111/ijal.12345>
- Doeniyas, C., Gül, Z. T., & Alıcı, B. (2023). A non-Western adaptation of the Situated Academic Writing Self-Efficacy Scale (SAWSES). *Assessing Writing*, 57, Article 100763. <https://doi.org/10.1016/j.asw.2023.100763>
- Dong, J., Zhao, Y., & Buckingham, L. (2024). Thirty years of writing assessment: A bibliometric analysis of research trends and future directions. *Assessing Writing*, 61, 100862. <https://doi.org/10.1016/j.asw.2024.100862>
- Durrant, P., & Brenchley, M. (2021). The development of academic collocations in children's writing. In P. Szudarski & S. Barclay (Eds.), *Vocabulary theory, patterning and teaching* (pp. 99–120). Multilingual Matters.
- Durrant, P., & Durrant, A. (2022). Appropriateness as an aspect of lexical richness: What do quantitative measures tell us about children's writing? *Assessing Writing*, 51, Article 100596. <https://doi.org/10.1016/j.asw.2021.100596>
- East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–273.
<https://doi.org/10.1017/s0261444823000034>
- Eckes, T., & Jin, K. Y. (2021). Examining severity and centrality effects in TestDaF writing and speaking assessments: An extended Bayesian many-facet Rasch analysis.

- International Journal of Testing*, 21(3–4), 131–153.
<https://doi.org/10.1080/15305058.2021.1963260>
- Effatpanah, F., Baghaei, P., & Karimi, M. N. (2024). A mixed Rasch model analysis of multiple profiles in L2 writing. *Assessing Writing*, 59, 100803.
<https://doi.org/10.1016/j.asw.2023.100803>
- El Ebyary, K., Shabara, R., & Boraie, D. (2024). The differential role of AI-operated WCF in L2 students' noticing of errors and its impact on writing scores. *Language Testing in Asia*, 14, Article 59. <https://doi.org/10.1186/s40468-024-00312-1>
- Elosua, P. (2022). Validity evidences for scoring procedures of a writing assessment task. A case study on consistency, reliability, unidimensionality and prediction accuracy. *Assessing Writing*, 54, Article 100669. <https://doi.org/10.1016/j.asw.2022.100669>
- Emmanouil, K., Zoe, G., Christos, A., & Markos, A. (2024). Linguistic factors affecting L1 language evaluation in argumentative essays of students aged 16 to 18 attending secondary education in Greece. *Assessing Writing*, 60, 100844.
<https://doi.org/10.1016/j.asw.2024.100844>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14, Article 51.
<https://doi.org/10.1186/s40468-024-00326-9>
- Eskin, D. (2023). Writing task performance and first language background on an ESL placement exam: A many-facets rasch analysis of facet main effects and differential facet functioning. *Studies in Applied Linguistics & TESOL*, 23(1), 37-65.
<https://doi.org/10.52214/salt.v23i1.11805>
- Estaji, M., & Hashemi, M. (2022). Phraseological competence in IELTS academic writing task 2: Phraseological units and test-takers' perceptions and use. *Language Testing in Asia*, 12, Article 34.
<https://doi-org.proxy.bib.uottawa.ca/10.1186/s40468-022-00180-7>
- Estaji, M., & Safari, F. (2023) Learning-oriented assessment and its effects on the perceptions and argumentative writing performance of impulsive vs. reflective learners. *Language Testing in Asia*, 13, Article 31 (2023).
<https://doi.org/10.1186/s40468-023-00248-y>
- Fan, J., Knoch, U., & Chen, I. (2021). *Linking the GEPT writing subtest (Part 1) to the Common European Framework of Reference (CEFR)*. Language Testing Research Centre.
https://www.lttc.ntu.edu.tw/lttc-grants/doc/108report/GEPT_linking_study_Final_report.pdf
- Feng, H.-H., & Chukharev-Hudilainen, E. (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology*, 26(2), 58–77. <https://doi.org/10.1257/73479>
- Finch, M., & Willis, J. (2021). Learning to be expert writers: Feedback in secondary English. *Assessment in Education: Principles, Policy & Practice*, 28(2), 101–117.
<https://doi.org/10.1080/0969594X.2021.1908226>
- Finkelstein, S., & Netz, H. (2023). Challenging folk-linguistics: Grammatical and spelling variation in students' writing in hebrew on WhatsApp and in essays. *Applied Linguistics*, 44(3), 555–575. <https://doi-org/10.1093/applin/amac072>
- Finkenstaedt-Quinn, S. A., Watts, F. M., & Shultz, G. V. (2024). Reading, receiving, revising: A case study on the relationship between peer review and revision in writing-to-learn. *Assessing Writing*, 59, 100808.
<https://doi.org/10.1016/j.asw.2024.100808>
- Firoozi, T., Mohammadi, H., & Gierl, M. J. (2023). Using active learning methods to strategically select essays for automated scoring. *Educational Measurement, Issues and Practice*, 42(1), 34–43. <https://doi.org/10.1111/emip.12537>

- Fladung, I. T., Gruhn, S., Österbauer, V., & Jost, J. (2023). Assessing writing in fourth grade: Rhetorical specification effects on text quality. *Assessing Writing*, 57, Article 100764. <https://doi.org/10.1016/j.asw.2023.100764>
- Flores, J. A. A. (2021). The semiotics of writing: How adult L2 learners with emergent literacy make meaning in assessment texts through writing. *Journal of Second Language Writing*, 51, Article 100793. <https://doi.org/10.1016/j.jslw.2021.100793>
- Fong, C. J., Schallert, D. L., Williamson, Z. H., Lin, S., Williams, K. M., & Kim, Y. W. (2023). Are self-compassionate writers more feedback literate? Exploring undergraduates' perceptions of feedback constructiveness. *Assessing Writing*, 57, Article 100761. <https://doi.org/10.1016/j.asw.2023.100761>
- Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(1-2), 179-221. <https://doi.org/10.1080/09588221.2022.2033787>
- Fukunaga, T. (2023). L2 writing development through two types of writing task repetition. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1109-1138. <https://doi.org/10.1515/iral-2021-0144>
- Gaillat, T., Lafontaine, A., & Knefati, A. (2023). Visualizing linguistic complexity and proficiency in learner English writings. *CALICO Journal*, 40(2), 178-197. <https://doi.org/10.1558/cj.19487>
- Gao, S., & Pan, M. (2024). Implementing dynamic assessment in writing in Chinese as a second language: Towards an adjusted cumulative group dynamic assessment. *System*, 123, 103327. <https://doi.org/10.1016/j.system.2024.103327>
- Ghaneiarani, S., Alavi, S. M., & Kaivanpanah, S. (2024). Enhancing writing ability of Iranian EFL learners through learning-oriented assessment: Peer and teacher feedback implementation. *Language Testing in Asia*, 14, Article 39. <https://doi.org/10.1186/s40468-024-00298-w>
- Golparvar, S. E., & Abolhasani, H. (2022). Unpacking the contribution of linguistic features to graph writing quality: An analytic scoring approach. *Assessing Writing*, 53, Article 100644. <https://doi.org/10.1016/j.asw.2022.100644>
- Golparvar, S. E., & Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assessing Writing*, 47, Article 100504. <https://doi.org/10.1016/j.asw.2020.100504>
- Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple-text source-based writing. *System*, 99, Article 102524. <https://doi.org/10.1016/j.system.2021.102524>
- Gomez Laich, M., & Taguchi, N. (2023). Effects of task complexity on the learning of genre specific rhetorical moves and linguistic forms: the case of contrast and argumentative essays. *International Review of Applied Linguistics in Language Teaching*, 61(2), 571-599. <https://doi.org/10.1515/iral-2021-0111>
- Gong, T., Zhang, M., & Li, C. (2022). Association of keyboarding fluency and writing performance in online-delivered assessment. *Assessing Writing*, 51, Article 100575. <https://doi.org/10.1016/j.asw.2021.100575>
- Goodwin, R., & Kirkpatrick, R. (2023). Using rubrics to improve writing skills: a study in Kuwait. *Language Testing in Asia*, 13, Article 17 (2023). <https://doi.org/10.1186/s40468-023-00224-6>
- Graham, K. M. (2024). Examining the criterion validity of curriculum-based measurement of writing (CBM-W) with young EFL learners. *Research Methods in Applied Linguistics*, 3(2), Article 100108. <https://doi.org/10.1016/j.rmal.2024.100108>
- Graham, S., Harbaugh-Schattenkirk, A. G., Aitken, A., Harris, K. R., Ng, C., Ray, A., Wilson, J. M., & Wdowin, J. (2022). Writing motivation questionnaire: Validation and application as a formative assessment. *Assessment in Education: Principles, Policy & Practice*, 29(2), 238-261. <https://doi.org/10.1080/0969594X.2022.2080178>

- Graves, B. C. (2023). Specifications grading to promote student engagement, motivation and learning: Possibilities and cautions. *Assessing Writing*, 57, Article 100754. <https://doi.org/10.1016/j.asw.2023.100754>
- Guapacha Chamorro, M. E. (2022). Cognitive validity evidence of computer- and paper-based writing tests and differences in the impact on EFL test-takers in classroom assessment. *Assessing Writing*, 51, Article 100594. <https://doi.org/10.1016/j.asw.2021.100594>
- Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing*, 61, 100864. <https://doi.org/10.1016/j.asw.2024.100864>
- Hadidi, A. (2023). Comparing summative and dynamic assessments of L2 written argumentative discourse: Microgenetic validity evidence. *Assessing Writing*, 55, Article 100691. <https://doi.org/10.1016/j.asw.2022.100691>
- Hanjani, A. M. (2021). Insights into the cognitive processes of trained vs untrained EFL peer reviewers on writing: An exploratory study. *Assessing Writing*, 49, Article 100530. <https://doi.org/10.1016/j.asw.2021.100530>
- Hanjani, A. M. (2024). Comparing trained EFL peer reviewers' feedback: From claim to reality. *Assessing Writing*, 60, 100836. <https://doi.org/10.1016/j.asw.2024.100836>
- Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity arguments for automated essay scoring of young students' writing traits. *Language Assessment Quarterly*, 20(4-5), 399–420. <https://doi.org/10.1080/15434303.2023.2288253>
- Hao, Y., Wang, X., Bin, S., Yang, Q., & Liu, H. (2024). How syntactic complexity indices predict Chinese L2 writing quality: An analysis of unified dependency syntactically-annotated corpus. *Assessing Writing*, 61, 100847. <https://doi.org/10.1016/j.asw.2024.100847>
- Harrasi, K. T. S. AL. (2023). Reexamining feedback in the context of different rhetorical patterns of writing. *Language Testing in Asia*, 13, Article 4 (2023). <https://doi.org/10.1186/s40468-023-00216-6>
- Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing*, 62, 100894. <https://doi.org/10.1016/j.asw.2024.100894>
- Hartwell, K., & Aull, L. (2024). Navigating innovation and equity in writing assessment. *Assessing Writing*, 61, 100873. <https://doi.org/10.1016/j.asw.2024.100873>
- He, L., Jiang, Z., & Min, S. (2021). Diagnosing writing ability using China's Standards of English Language Ability: Application of cognitive diagnosis models. *Assessing Writing*, 50, Article 100565. <https://doi.org/10.1016/j.asw.2021.100565>
- Henry, N., Mandel, H., & VanderHeijden, V. (2024). Dialogue journals and traditional writing assignments: A comparison of writing in L2 German classrooms. *Foreign Language Annals*, 57(2), 339–361. <https://doi.org/10.1111/flan.12752>
- Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101, Article 102569. <https://doi.org/10.1016/j.system.2021.102569>
- Hodges, T. S., Wright, K. L., & McTigue, E. M. (2021). The Preservice Teacher Self-Efficacy for Writing Inventory. (PTSWI): A tool for measuring beliefs about writing. *Assessing Writing*, 49, Article 100545. <https://doi.org/10.1016/j.asw.2021.100545>
- Honko, M., Neittaanmäki, R., Jarvis, S., & Huhta, A. (2023). Beyond literacy and competency – The effects of raters' perceived uncertainty on assessment of writing. *Assessing Writing*, 57, Article 100768. <https://doi.org/10.1016/j.asw.2023.100768>
- Hoomanfar, M. (2024). L2 master's and doctoral students' preferences for supervisor written feedback on their theses/dissertations. *Assessing Writing*, 62, 100891. <https://doi.org/10.1016/j.asw.2024.100891>

- Hosseinpur, R. M., & Kazemi, Z. (2022). Composing strategies employed by high-and low-performing Iranian EFL students in essay writing classes. *Assessing Writing*, 51, Article 100601. <https://doi.org/10.1016/j.asw.2021.100601>
- Hsieh, C. (2024). The role of task types and reading proficiency on Young English as a foreign language learners' writing performances. *TESOL Quarterly*, 58(2), 978–990. <https://doi.org/10.1002/tesq.3286>
- Huang, D., & Watzinger-Tharp, J. (2024). The use of cohesive devices as proficiency level discriminators in Chinese DLI learners' writing. *Foreign Language Annals*, 57(2), 472–498. <https://doi.org/10.1111/flan.12722>
- Huang, J., & Chen, G. (2022). Individualized feedback to raters in language assessment: Impacts on rater effects. *Assessing Writing*, 52, Article 100623. <https://doi.org/10.1016/j.asw.2022.100623>
- Huang, J., Shu, T., Dong, Y., & Zhu, D. (2023). Constructing and validating a self-assessment scale for Chinese college English-major students' feedback knowledge repertoire in EFL academic writing: Item response theory and factor analysis approaches. *Assessing Writing*, 56, Article 100716. <https://doi.org/10.1016/j.asw.2023.100716>
- Huang, J., Zhu, D., Xie, D., & Shu, T. (2023). Examining the reliability of an international Chinese proficiency standardized writing assessment: Implications for assessment policy makers. *Assessing Writing*, 55, Article 100693. <https://doi.org/10.1016/j.asw.2023.100693>
- Huang, T., Steinkrauss, R., & Verspoor, M. (2021). Variability as predictor in L2 writing proficiency. *Journal of Second Language Writing*, 52, Article 100787. <https://doi.org/10.1016/j.jslw.2020.100787>
- Huang, Y., Wu, J., Li, M., & Yang, Y. (2022). The trajectory of syntactic complexity development in L1 Chinese narrative writings of primary school children: A systematic 5-year longitudinal study. *Assessing Writing*, 52, Article 100622. <https://doi.org/10.1016/j.asw.2022.100622>
- Huhta, A., Harsch, C., Leontjev, D., & Niemegen, L. (2024). *Diagnosing Writing*. London: Routledge. DOI: 10.4324/9781315511979-5
- Huhta, A., Harsch, C., Leontjev, D., & Nieminen, L. (2024). *The diagnosis of writing in a second or foreign language*. Routledge.
- Hwang, H., & Kim, H. (2023). Automatic analysis of constructional diversity as a predictor of EFL students' writing proficiency. *Applied Linguistics*, 44(1), 127–147. <https://doi.org/10.1093/applin/amac046>
- Hyung-Jo, Y., & Mahmoud, A.T. (2023). Authorial voice in source-based and opinion-based argumentative writing: Patterns of voice across task types and proficiency levels. *English for Academic Purposes*, 62, Article 101228. <https://doi.org/10.1016/j.jeap.2023.101228>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The Terminator Versus the Machines". *Language Testing in Asia*, 13, Article 46 (2023). <https://doi.org/10.1186/s40468-023-00260-2>
- Ignacio, A., Jang, E. E., Heywood, A., Hunte, M. R., & Lai, H. (2024). Fostering young students' writing skill development and self-regulation through Dynamic Assessment feedback during pandemic disruptions. *Studies in Language Assessment (SILA)*, 13(2), 147–186. <https://doi.org/10.58379/asqh2857>
- Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, 55, Article 100676. <https://doi.org/10.1016/j.asw.2022.100676>
- Issa, B. I., Koronkiewicz, B., & Faretta-Stutenberg, M. (2022). Second-language writing in university-level basic language programs: A survey of student and instructor beliefs. *Foreign Language Annals*, 55(2), 383–407. <https://doi.org/10.1111/flan.12609>

- Iwashita, N., Sasaki, M., Stell, S., & Yucel, M. (2021). Japanese stakeholders' perceptions of IELTS writing and speaking tests and their impact on communication and achievement. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/teaching-and-research/research-reports>
- Jalali, H., & Rahimi, M. (2022). Incongruence between learning style and written corrective feedback type: Mediating effect of implicit theory of learning style. *Canadian Journal of Applied Linguistics*, 25(2), 1–22. <https://doi.org/10.37213/cjal.2022.32478>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). Discussion, limitations, and future research. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 369–385). John Benjamins.
<https://doi.org/10.1075/bpa.13>
- Jia, W., & Zhang, P. (2023). Rater cognitive processes in integrated writing tasks: from the perspective of problem-solving. *Language Testing in Asia*, 13, Article 50 (2023).
<https://doi.org/10.1186/s40468-023-00265-x>
- Jiang, J., Bi, P., Xie, N., & Liu, H. (2023). Phraseological complexity and low- and intermediate-level L2 learners' writing quality. *International Review of Applied Linguistics in Language Teaching*, 61(3), 765–790.
<https://doi.org/10.1515/iral-2019-0147>
- Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing*, 57, Article 100869.
<https://doi.org/10.1016/j.jslw.2022.100869>
- Joo, S. H., Seong, Y., Suh, J., Jung, J.-Y., & Purpura, J. E. (2023). Assessing Korean writing ability through a scenario-based assessment approach. *Assessing Writing*, 58, Article 100766. <https://doi.org/10.1016/j.asw.2023.100766>
- Jung, J., & Fu, X. (2023). The impact of pragmalinguistic support on video-conferenced collaborative suggestion-giving task. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1083–1107. <https://doi.org/10.1515/iral-2021-0138>
- Kang, O., Egbert, J., & Miao, Y. (2023). *Speaking and writing features: Distinguishing IELTS proficiency levels and progression over time* (IELTS Research Report Series No. 1/23). British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262788072871836>
- Kang, S., & Kim, Y. (2024). Examining the quality of mobile-assisted, video-making task outcomes: The role of proficiency, narrative ability, digital literacy, and motivation. *Language Teaching Research*, 28(6), 2326–2353.
<https://doi.org/10.1177/13621688211047984>
- Kao, Y.-T. (2022). Effects of group dynamic assessment on L2 Chinese learners' literacy development: learners' responsiveness to interactive mediation. *Applied Linguistics Review*, 13(5), 843–871. <https://doi.org/10.1515/applirev-2019-0077>
- Karaca, M., & Uysal, H. H. (2021). The development and validation of an inventory on English writing teacher beliefs. *Assessing Writing*, 47, Article 100507.
<https://doi.org/10.1016/j.asw.2020.100507>
- Karaca, M., & Uysal, H. H. An investigation into L2 writing teacher beliefs and their possible sources. *Assessing Writing*, 56, Article 100710.
<https://doi.org/10.1016/j.asw.2023.100710>
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, 103332.
<https://doi.org/10.1016/j.system.2024.103332>
- Karatay, Y., & Karatay, L. (2024). The impact of automated writing evaluation on language learning and teaching: A research synthesis. *System*, 123.
<https://doi.org/10.1016/j.system.2024.103332>

- Karimi, M. N., & Mousavi-Moghadam, T. (2022). Integrated writing across conflicting L2 texts: the effects of prior beliefs on perspectivism and propositional content. *Assessing Writing*, 53, Article 100648. <https://doi.org/10.1016/j.asw.2022.100648>
- Kennedy, E., & Shiel, G. (2022). Writing assessment for communities of writers: rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy & Practice*, 29(2), 127–149. <https://doi.org/10.1080/0969594X.2022.2047608>
- Kessler, M. (2024). Assessment. In M. Kessler (Ed.), *Digital multimodal composing: Connecting theory, research and practice in second language acquisition* (pp. 109-120). Multilingual Matters. <https://doi.org/10.2307/jj.22679723.11>
- Kessler, M., Ma, W., & Solheim, I. (2022). The effects of topic familiarity on text quality, complexity, accuracy, and fluency: a conceptual replication. *TESOL Quarterly*, 56(4), 1163-1190. <https://doi.org/10.1002/tesq.3096>
- Khodi, A. (2021). The affectability of writing assessment scores: A G-theory analysis of rater, task, and scoring method contribution. *Language Testing in Asia*, 11, Article 30. <https://doi.org/10.1186/s40468-021-00134-5>
- Khushik, G. A. (2024). Is the variation in syntactic complexity features observed in argumentative essays produced by B1 level EFL learners in Finland and Pakistan attributable exclusively to their L1? *Assessing Writing*, 60, 100839. <https://doi.org/10.1016/j.asw.2024.100839>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73–95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Kim, J. (2023). Memorizing and fabricating? Uncovering high-Stakes writing test preparation. *Language Education & Assessment*, 6(1), 19–37. <https://doi.org/10.29140/lea.v6n1.1175>
- Kim, M. K., Kim, J., & Heidari, A. (2024). Exploring the multi-dimensional human mind: Model-based and text-based approaches. *Assessing Writing*, 61, 100878. <https://doi.org/10.1016/j.asw.2024.100878>
- Kim, M., Crossley, S. A., & Kim, B.-K. (2022). Second language reading and writing in relation to first language, vocabulary knowledge, and learning backgrounds. *International Journal of Bilingual Education and Bilingualism*, 25(6), 1992–2005. <https://doi.org/10.1080/13670050.2020.1838434>
- Kim, P., Beltrán, J., & Joo, S. H. (2022). An interview with APPLE Lecture speaker Dr. Paul D. Deane. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University*, 22(1), 47–55. <https://doi.org/10.52214/salt.v22i1.9901>
- Kim, S. (2021). Generalizability of CEFR criterial grammatical features in a Korean EFL corpus across A1, A2, B1, and B2 levels. *Language Assessment Quarterly*, 18(3), 273–295. <https://doi.org/10.1080/15434303.2020.1855647>
- Kim, S. (2021). Prepping for the TOEFL iBT Writing test, Gangnam style. *Assessing Writing*, 49, Article 100544. <https://doi.org/10.1016/j.asw.2021.100544>
- Kim, S., & Kessler, M. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, Article 100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Kim, Y. J., & Hammill, M. J. (2024). Influence of prior educational contexts on directed self-placement of L2 writers. *Assessing Writing*, 61, 100870. <https://doi.org/10.1016/j.asw.2024.100870>
- Kim, Y., & Emeliyanova, L. (2021). The effects of written corrective feedback on the accuracy of L2 writing: Comparing collaborative and individual revision behavior. *Language Teaching Research*, 25(2), 234–255. <https://doi.org/10.1177/1362168819831406>

- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247. <https://doi.org/10.1016/j.jeap.2023.101247>
- Kirkwood, M. (2023). Using ipsative feedback for personal bests in grammatical accuracy in L2 writing. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 105–128). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Knoch, U., Fairbairn, J., & Jin, Y. (2021). *Scoring second language spoken and written performance: Issues, options and directions*. University of Toronto Press.
- Knoch, U., Fan, J., Elder, C., Zhao, K., & Pitman, A. (2021). *Investigating the discourse produced at score levels B2.2 to C2 on the Aptis advanced writing test* (AR-G/2021/5). British Council. https://www.britishcouncil.org/sites/default/files/knoch_et_al_report_layout.pdf
- Koizumi, R., Agawa, T., Asano, K., & In'nami, Y. (2022). Skill profiles of Japanese English learners and reasons for uneven patterns. *Language Testing in Asia*, 12, Article 53. <https://doi.org/10.1186/s40468-022-00203-3>
- Kushki, A., Nassaji, H., & Rahimi, M. (2022). Interventionist and interactionist dynamic assessment of argumentative writing in an EFL program. *System*, 107, Article 102800. <https://doi.org/10.1016/j.system.2022.102800>
- Kushki, A., Rahimi, M., & Davin, K. J. (2022). Dynamic assessment of argumentative writing: mediating task response. *Assessing Writing*, 52, Article 100606. <https://doi.org/10.1016/j.asw.2022.100606>
- Kyle, K., & Eguchi, M. (2024). Evaluating NLP models with written and spoken L2 samples. *Research Methods in Applied Linguistics*, 3(2), Article 100120. <https://doi.org/10.1016/j.rmal.2024.100120>
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781–812. <https://doi.org/10.1017/S0272263120000546>
- Lamprianou, I., Tsigari, D., & Kyriakou, N. (2021). The longitudinal stability of rating characteristics in an EFL examination: Methodological and substantive considerations. *Language Testing*, 38(2), 273–301. <https://doi.org/10.1177/0265532220940960>
- Lamprianou, I., Tsigari, D., & Kyriakou, N. (2023). Experienced but detached from reality: Theorizing and operationalizing the relationship between experience and rater effects. *Assessing Writing*, 56, Article 100713. <https://doi.org/10.1016/j.asw.2023.100713>
- Latif, M. M. A. (2021). Remodeling writers' composing processes: Implications for writing assessment. *Assessing Writing*, 50, Article 100547. <https://doi.org/10.1016/j.asw.2021.100547>
- Latif, M. M. A., Alsuhaibani, Z., & Alsahil, A. (2024). Matches and mismatches between Saudi university students' English writing feedback preferences and teachers' practices. *Assessing Writing*, 61, 100863. <https://doi.org/10.1016/j.asw.2024.100863>
- Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62–85. <https://doi.org/10.1177/0265532220929918>
- Lau, W. M. E., Aubrey, S., & Ellis, R. (2024). Error logs in the second language classroom: Exploring the relationship between learner engagement with written corrective feedback and improvements in writing accuracy. *System*, 123, 103329. <https://doi.org/10.1016/j.system.2024.103329>
- Laudenbach, M., Brown, D. W., Guo, Z., Ishizaki, S., Reinhart, A., & Weinberg, G. (2024). Visualizing formative feedback in statistics writing: An exploratory study of student motivation using DocuScope Write & Audit. *Assessing Writing*, 60, 100830. <https://doi.org/10.1016/j.asw.2024.100830>

- Lázaro-Ibarrola, A. (2023). Model texts in collaborative and individual writing among EFL children: noticing, incorporations, and draft quality. *International Review of Applied Linguistics in Language Teaching*, 61(2), 329-359. <https://doi.org/10.1515/iral-2020-0160>
- Lecouvet, M. (2023). Non-canonical word order as a measure of syntactic complexity in advanced L2 German. *International Review of Applied Linguistics in Language Teaching*, 61(3), 877–907. <https://doi.org/10.1515/iral-2021-0029>
- Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, 97, Article 102461. <https://doi.org/10.1016/j.system.2021.102461>
- Lee, E. (2022). Examining the social consequences of a locally-developed placement test using test takers' attitudes. *Assessing Writing*, 53, Article 100626. <https://doi.org/10.1016/j.asw.2022.100626>
- Lee, I. (2021). The development of feedback literacy for writing teachers. *TESOL Quarterly*, 55(3), 1048–1059. <https://doi.org/10.1002/tesq.3012>
- Lee, I., Karaca, M., & Inan, S. (2023). The development and validation of a scale on L2 writing teacher feedback literacy. *Assessing Writing*, 57, Article 100743. <https://doi.org/10.1016/j.asw.2023.100743>
- Lee, J. J., Tytko, T., & Larkin, R. (2021). (Un)attended this/these in undergraduate student writing: A corpus analysis of high- and low-rated L2 writers. *Journal of English for Academic Purposes*, 50, Article 100967. <https://doi.org/10.1016/j.jeap.2021.100967>
- Lee, S., Lim, G. S., & Basse, R. (2021). The effect of additional time on the quality of argumentation in L2 writing assessment: A mixed-methods study. *Language Assessment Quarterly*, 18(3), 253–272. <https://doi.org/10.1080/15434303.2021.1872080>
- Leontjev, D., & Pollari, P. (2023). Guiding and assessing development of L2 writing process: The role of peer collaboration. *Innovation in Language Learning and Teaching*, 17(2), 440–452. <https://doi.org/10.1080/17501229.2022.2058514>
- Li, A. W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57, Article 100746. <https://doi.org/10.1016/j.asw.2023.100746>
- Li, C., Li, W., & Lu, X. (2024). Task complexity and L2 writing performance of young learners: Contributions of cognitive and affective factors. *The Modern Language Journal*, (108)3, 741-770. <https://doi.org/10.1111/modl.12954>
- Li, J., & Huang, J. (2022). The impact of essay organization and overall quality on the holistic scoring of EFL writing: perspectives from classroom English teachers and national writing raters. *Assessing Writing*, 51, Article 100604. <https://doi.org/10.1016/j.asw.2021.100604>
- Li, M., Zhang, X., & Reynolds, B. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: an ETS corpus study. *Applied Linguistics Review*, 14(4), 847–873. <https://doi.org/10.1515/applirev-2020-0129>
- Li, S., & Yu, H. (2024). Effects of topic familiarity on L2 writing processes and behaviors. *International Journal of Applied Linguistics*, 34(1), 348-366. <https://doi.org/10.1111/ijal.12497>
- Li, Y. K., Lin, S., Liu, Y., & Lu, X. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, 56, Article 100707. <https://doi.org/10.1016/j.asw.2023.100707>
- Li, Y., & Yang, R. (2023). Assessing the writing quality of English research articles based on absolute and relative measures of syntactic complexity. *Assessing Writing*, 55, Article 100692. <https://doi.org/10.1016/j.asw.2022.100692>

- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, Article 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Liao, J. (2023). What skills are being assessed? Evaluating L2 Chinese essays written by hand and on a computer keyboard. *Assessing Writing*, 57, Article 100765. <https://doi.org/10.1016/j.asw.2023.100765>
- Ling, G., Elliot, N., Burstein, J. C., McCaffrey, D. F., MacArthur, C. A., & Holtzman, S. (2021). Writing motivation: A validation study of self-judgment and performance. *Assessing Writing*, 48, Article 100509. <https://doi.org/10.1016/j.asw.2020.100509>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, C., & Yu, S. (2022). Reconceptualizing the impact of feedback in second language writing: A multidimensional perspective. *Assessing Writing*, 53, Article 100630. <https://doi.org/10.1016/j.asw.2022.100630>
- Liu, C., Sun, L., He, Y., & Wu, N. (2023). The effects of task complexity and language aptitude on EFL learners' writing performance. *Assessing Writing*, 58, Article 100791. <https://doi.org/10.1016/j.asw.2023.100791>
- Liu, F., Du, J., Zhou, D. Q., & Huang, B. (2021). Exploiting the potential of peer feedback: The combined use of face-to-face feedback and e-feedback in doctoral writing groups. *Assessing Writing*, 47, Article 100482. <https://doi.org/10.1016/j.asw.2020.100482>
- Liu, S., & Lin, D. (2022). Developing and validating an analytic rating scale for a paraphrase task. *Assessing Writing*, 53, Article 100646. <https://doi.org/10.1016/j.asw.2022.100646>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78-105. <https://doi.org/10.1257/73480>
- Liu, Y., & Liu, J. (2023). Comparing computer-based and paper-based rating modes in an English writing test. *Assessing Writing*, 57, Article 100771. <https://doi.org/10.1016/j.asw.2023.100771>
- Liu, Y., Cheong, C. M., & Zhang, J. (2024). The Role of Achievement Goals and Strategy Use in Secondary School students' Integrated Writing Performance. *Language Assessment Quarterly*, 21(4-5), 324-349. <https://doi.org/10.1080/15434303.2024.2435818>
- Liu, Y., Storch, N., & Morton, J. (2022). It takes two to tango: Investigating teacher-student interactions related to written corrective feedback with activity theory. *Assessing Writing*, 53, Article 100647. <https://doi.org/10.1016/j.asw.2022.100647>
- Lombardi, A. (2021). More is more: Explicit intertextuality in university writing placement exam essays. *Journal of English for Academic Purposes*, 50, Article 100955. <https://doi.org/10.1016/j.jeap.2020.100955>
- Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing*, 56, Article 100725. <https://doi.org/10.1016/j.asw.2023.100725>
- Lu, Q., Zhu, X., Zhu, S., & Yao, Y. (2024). Effects of writing feedback literacies on feedback engagement and writing performance: A cross-linguistic perspective. *Assessing Writing*, 62, 100889. <https://doi.org/10.1016/j.asw.2024.100889>
- Lukácsi, Z. (2021). Developing a level-specific checklist for assessing EFL writing. *Language Testing*, 38(1), 86-105. <https://doi.org/10.1177/0265532220916703>
- Luquin, M., & García Mayo, M. D. P. (2024). A longitudinal study of the effects of model texts on EFL children's written production. *System*, 120, 103190. <https://doi.org/10.1016/j.system.2023.103190>

- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49, Article 100529. <https://doi.org/10.1016/j.asw.2021.100529>
- Lydster, C. (2024). Who says “Ps get Degrees”? Examining the Profile of Undergraduate Students Maintaining High Achievement at University. *Student Success*, 15(2), 23–33. <https://doi.org/10.5204/ssj.3402>
- Ma, H., Wang, J., & He, L. (2024). Linguistic features distinguishing students' writing ability aligned with CEFR levels. *Applied Linguistics*, 45(4), 637-657. <https://doi.org/10.1093/applin/amad054>
- Ma, M. (2023). Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives. *Language Testing in Asia*, 13, Article 33 (2023). <https://doi.org/10.1186/s40468-023-00249-x>
- Ma, M., & Bui, G. (2022). Implementing continuous assessment in an academic English writing course: An exploratory study. *Assessing Writing*, 53, Article 100629. <https://doi.org/10.1016/j.asw.2022.100629>
- Maamujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, Article 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Mao, Z., & Lee, I. (2022). Researching L2 student engagement with written feedback: Insights from sociocultural theory. *TESOL Quarterly*, 56(2), 788-798. <https://doi.org/10.1002/tesq.3071>
- Martin, C., & Dockrell, J. E. (2024). Writing productivity development in elementary school: A systematic review. *Assessing Writing*, 60, 100834. <https://doi.org/10.1016/j.asw.2024.100834>
- Matsumura, L. C., Wang, E. L., Correnti, R., & Litman, D. (2023). Tasks and feedback: An exploration of students' opportunity to develop adaptive expertise for analytic text-based writing. *Assessing Writing*, 55, Article 100689. <https://doi.org/10.1016/j.asw.2022.100689>
- Matta, M., Mercer, S. H., & Keller-Margulis, M. A. (2022). Evaluating validity and bias for hand-calculated and automated written expression curriculum-based measurement scores. *Assessment in Education: Principles, Policy & Practice*, 29(2), 200–218. <https://doi.org/10.1080/0969594X.2022.2043240>
- McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, 52, Article 100608. <https://doi.org/10.1016/j.asw.2022.100608>
- McDonough, K., Uludag, P., & Neumann, H. (2021). Instructor evaluation of business student writing: Does language play a role? *Business and Professional Communication Quarterly*, 84(2), 116–134. <https://doi.org/10.1177/23294906211012398>
- McNamara, D. S., & Kendeou, P. (2022). The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*, 29(2), 150–182. <https://doi.org/10.1080/0969594X.2022.2037509>
- Meletiadiou, E., & Tsigari, D. (2022). Exploring EFL teachers' perceptions of the use of peer assessment in external exam-dominated writing classes. *Languages*, 7(1), 16. <https://doi.org/10.3390/languages7010016>
- Mesina, V. G. (2023). The “Problem” of L2 writers in college composition placement: Identity, outcomes, and the future of directed self placement. *Studies in Applied Linguistics and TESOL*, 22(2), 23–36. <https://doi.org/10.52214/salt.v22i2.10682>
- Mitchell, K. M., McMillan, D. E., Lobchuk, M. M., Nickel, N. C., Rabbani, R., & Li, J. (2021). Development and validation of the situated academic writing self-efficacy scale (SAWSES). *Assessing Writing*, 48, Article 100524. <https://doi.org/10.1016/j.asw.2021.100524>

- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.
- Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), Article 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
- Mohebbi, H. (2021). 25 years on, the correction debate continues: An interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), Article 3. <https://doi.org/10.1186/s40862-021-00110-9>
- Mohsen, M. A. (2024). L1 versus L2 writing processes: What insight can we obtain from a keystroke logging program? *Language Teaching Research*, 28(6), 2251-2275. <https://doi.org/10.1177/13621688211041292>
- Morton, T., & Nashaat-Sobhy, N. (2024). Exploring Bases of achievement in content and Language Integrated assessment in a bilingual education program. *TESOL Quarterly*, 58(1), 5–31. <https://doi.org/10.1002/tesq.3207>
- Mujtaba, S. M., Reynolds, B. L., Parkash, R., & Singh, M. K. M. (2021). Individual and collaborative processing of written corrective feedback affects second language writing accuracy and revision. *Assessing Writing*, 50, Article 100566. <https://doi.org/10.1016/j.asw.2021.100566>
- Müller, A. & Han, W. (2022). *IELTS writing band scores 5.5–7.5: Grammatical error rates, stakeholder perceptions, and risk* (IELTS Research Reports Online Series, No. 1/22). International English Language Testing System. <https://www.ielts.org/-/media/research-reports/muller-and-han-2022.ashx>
- Mumford, S., & Atay, D. (2021). Teachers' perspectives on the causes of rater discrepancy in an English for Academic Purposes context. *Assessing Writing*, 48, Article 100527. <https://doi.org/10.1016/j.asw.2021.100527>
- Myint Maw, T. M., Clenton, J., & Higginbotham, G. (2022). Investigating whether a flemma count is a more distinctive measurement of lexical diversity. *Assessing Writing*, 53, Article 100640. <https://doi.org/10.1016/j.asw.2022.100640>
- Naismith, B., Attali, Y., & LaFlair, G. T. (2024). The impact of task duration on the scoring of independent writing responses of adult L2-English writers. *Assessing Writing*, 62, 100895. <https://doi.org/10.1016/j.asw.2024.100895>
- Namkung, Y., & Kim, Y. (2024). Learner engagement in collaborative writing: The effects of SCMC mode, interlocutor familiarity, L2 proficiency, and task repetition. *System*, 121, 103251. <https://doi.org/10.1016/j.system.2024.103251>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, Article 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Nguyen, B. T. T., Nguyen, T. T. B., Phan, Q. N. (2022). Vietnamese IELTS learners' perceptions of IELTS Writing Task 2 and their performance-based lexical resource. IELTS Research Reports Online Series, No. 5/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/nguyen-nguyen--phan-oct-2022.ashx>
- Nguyen, L. Q., & Le, H. V. (2022). Improving L2 learners' IELTS task 2 writing: The role of model essays and noticing hypothesis. *Language Testing in Asia*, 12, Article 58. <https://doi.org/10.1186/s40468-022-00206-0>
- Nguyen, L. Q., Nguyen, B. T. T., & Phuong, H. Y. (2024). Exploring the use of model texts as a feedback instrument in expository writing: EFL learners' noticing, incorporations, and text quality. *Assessing Writing*, 62, 100890. <https://doi.org/10.1016/j.asw.2024.100890>
- Nguyen, L., & Barrot, J. S. (2024). Detecting and assessing AI-generated and human-produced texts: The case of second language writing teachers. *Assessing Writing*, 62, 100899. <https://doi.org/10.1016/j.asw.2024.100899>

- Niu, R., Shan, P., & You, X. (2021). Complementation of multiple sources of feedback in EFL learners' writing. *Assessing Writing*, 49, Article 100549. <https://doi.org/10.1016/j.asw.2021.100549>
- Norris, J.M., Sasayama, S. and Kim, M. (2023), Simulating real-world context in an email writing task: Implications for task-based language assessment. *ETS Research Report Series*, 2023, 1-22. <https://doi.org/10.1002/ets2.12366>
- Oh, S. (2022). The use of spelling and reference tools in second language writing: Their impact on students' writing performance and process. *Journal of Second Language Writing*, 57, Article 100916. <https://doi.org/10.1016/j.jslw.2022.100916>
- Ortiz, M., and Díaz, C. (2022). *The effect of screencast feedback on the performance of the IELTS essay writing component among pre-service teachers of English: An intervention study in Chile*. IELTS Research Reports Online Series, No. 3/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/oritz-and-diaz-2022-june.ashx>
- Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51, Article 100603. <https://doi.org/10.1016/j.asw.2021.100603>
- Owen, N., Shrestha, P., & Bax, S. (2021). *Researching lexical thresholds and lexical profiles across the Common European Framework of References for Languages (CEFR) levels assessed in the Aptis test* (Report No. AR-G/2021/1). British Council. https://www.britishcouncil.org/sites/default/files/owen_shrestha_and_bax_layout.pdf
- Papi, M., Wolff, D., Nakatsukasa, K., & Bellwoar, E. (2021). Motivational factors underlying learner preferences for corrective feedback: Language mindsets and achievement goals. *Language Teaching Research*, 25(6), 858–877. <https://doi.org/10.1177/13621688211018808>
- Parkinson, J., Watterson, G., Sawczuc da Silva, A., & Kithulgoda, E. (2023). Rhetorical features of high-graded and low-graded engineering case study responses. *Journal of English for Academic Purposes*, 61, Article 101206. <https://doi.org/10.1016/j.jeap.2022.101206>
- Patekar, J. (2021). A look into the practices and challenges of assessing young EFL learners' writing in Croatia. *Language Testing*, 38(3), 456–479. <https://doi.org/10.1177/0265532221990657>
- Pearson, W. S. (2022). The mediating effects of student beliefs on engagement with written feedback in preparation for high-stakes English writing assessment. *Assessing Writing*, 52, Article 100611. <https://doi.org/10.1016/j.asw.2022.100611>
- Peng, C. X. (2024). Beyond accuracy gains: Investigating the impact of individual and collaborative feedback processing on L2 writing development. *Assessing Writing*, 61, 100876. <https://doi.org/10.1016/j.asw.2024.100876>
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang H. (2023). Predicting Chinese EFL Learners' Human-rated Writing Quality in Argumentative Writing Through Multidimensional Computational Indices of Lexical Complexity. *Assessing Writing*, 56, Article 100722. <https://doi.org/10.1016/j.asw.2023.100722>
- Peng, Y., Zheng, Y., Sun, J., Jiang, Y., Lin, J., & Zhang, H. (2024). Modeling relationships among large-grained, fine-grained absolute syntactic complexity and assessed L2 writing quality: An SEM approach. *Assessing Writing*, 61, 100875. <https://doi.org/10.1016/j.asw.2024.100875>
- Peterson, S. S., Altidor, A., & Kerwood, J. (2021). Young children's written and verbal responses in a dynamic assessment context. *Assessing Writing*, 49, Article 100543. <https://doi.org/10.1016/j.asw.2021.100543>
- Phakiti, A. (2024). Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study. IELTS Research Reports Online Series, No. 1/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at <https://ielts.org/researchers/our-research/research-reports>

- Phuoc, V. D., & Barrot, J. S. (2022). Complexity, accuracy, and fluency in L2 writing across proficiency levels: A matter of L1 background? *Assessing Writing*, 54, Article 100673. <https://doi.org/10.1016/j.asw.2022.100673>
- Phuong, H. Y., Phan, Q. T. & Le, T. T. (2023). The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: a Vietnamese contextual study. *Language Testing in Asia*, 13, Article 42 (2023). <https://doi.org/10.1186/s40468-023-00256-y>
- Poehner, M. E., & Yu, L. (2022). Dynamic assessment of L2 writing: Exploring the potential of rubrics as mediation in diagnosing learner emerging abilities. *TESOL Quarterly*, 56(4), 1191-1217. <https://doi.org/10.1002/tesq.3098>
- Pourdana, N., & Asghari, S. (2021). Different dimensions of teacher and peer assessment of EFL learners' writing: Descriptive and narrative genres in focus. *Language Testing in Asia*, 11, Article 6. <https://doi.org/10.1186/s40468-021-00122-9>
- Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, 12, Article 7. <https://doi.org/10.1186/s40468-022-00156-7>
- Pun, J., & Li, W. K. (2024). A structural equation investigation of linguistic features as indices of writing quality in assessed secondary-level EMI learners' scientific reports. *Assessing Writing*, 62, 100897. <https://doi.org/10.1016/j.asw.2024.100897>
- Pun, J., Tan, S., & Li, X. (2024). Discourse competence in Hong Kong secondary students' disciplinary research writing. *Assessing Writing*, 61, 100872. <https://doi.org/10.1016/j.asw.2024.100872>
- Pusey, K., & Butler, Y. G. (2023). Investigating the ecological validity of second language writing assessment tasks. *System*, 119, Article 103174. <https://doi.org/10.1016/j.system.2023.103174>
- Pusey, K., & Butler, Y. G. (2024). Amplifying test-taker voices in the validation of L2 writing assessment tasks. *Assessing Writing*, 59, 100790. <https://doi.org/10.1016/j.asw.2023.100790>
- Qayyum, R. (2024). A teacher's inquiry into diagnostic assessment in an EAP writing course. *Assessing Writing*, 61, 100848. <https://doi.org/10.1016/j.asw.2024.100848>
- Qian, L. (2023). Use of lexical features in high-stakes tests: Evidence from the perspectives of complexity, accuracy and fluency. *Assessing Writing*, 57, Article 100758. <https://doi.org/10.1016/j.asw.2023.100758>
- Rad, H. S., & Alipour, R. (2024). Unlocking writing success: Building assessment literacy for students and teachers through effective interventions. *Assessing Writing*, 59, 100804. <https://doi.org/10.1016/j.asw.2023.100804>
- Rahimi, M. (2021). A comparative study of the impact of focused vs. comprehensive corrective feedback and revision on ESL learners' writing accuracy and quality. *Language Teaching Research*, 25(5), 687–710. <https://doi.org/10.1177/1362168819879182>
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, Article 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Rastgou, A. (2022). How feedback conditions broaden or constrain knowledge and perceptions about improvement in L2 writing: A 12-week exploratory study. *Assessing Writing*, 53, Article 100633. <https://doi.org/10.1016/j.asw.2022.100633>
- Razavipour, K. (2023). Classroom writing assessment and feedback practices: A new materialist encounter. *Assessing Writing*, 57, Article 100760. <https://doi.org/10.1016/j.asw.2023.100760>
- Reed, D. K., Ma, J., & Gerde, H. K. (2023). Resiliency and vulnerability in early grades writing performance during the COVID-19 pandemic. *Assessing Writing*, 57, Article 100741. <https://doi.org/10.1016/j.asw.2023.100741>

- Reifman, J. B. (2024). Reading the reader through raciolinguistic ideologies: An investigation of the evidence students present in self-placement. *Assessing Writing*, 59, 100792. <https://doi.org/10.1016/j.asw.2023.100792>
- Revesz, A., Michel, M., Lu, X., Kourtali, Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, 58, Article 100927. <https://doi.org/10.1016/j.jslw.2022.100927>
- Reyes, N. Á., Carrasco, D., Escribano, R., Espinosa, M. J., Figueroa, J., & Castillo, C. (2024). Effects of a genre and topic knowledge activation device on a standardized writing test performance. *Assessing Writing*, 62, 100898. <https://doi.org/10.1016/j.asw.2024.100898>
- Rezaei, A., & Barkaoui, K. (2021). Peer and teacher assessment of L2 writing in high- and low-stakes conditions. *International Journal of Applied Linguistics*, 172(2), 199–228. <https://doi.org/10.1075/itl.20006.rez>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2021). Linguistic complexity in teachers' assessment of German essays in high stakes testing. *Assessing Writing*, 50, Article 100561. <https://doi.org/10.1016/j.asw.2021.100561>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2024). The interplay of task characteristics, linguistic complexity, and language proficiency in High-Stakes English as a Foreign language writing. *TESOL Quarterly*, 58(2), 775–801. <https://doi.org/10.1002/tesq.3254>
- Rock, K. (2022). Constructing a data-based analytic rubric for an academic blog post. *Assessing Writing*, 51, Article 100602. <https://doi.org/10.1016/j.asw.2021.100602>
- Ruegg, R., & Prommas, P. (2024). Comparing automated analysis and human analysis of hedging in academic texts. *Studies in Language Assessment (SILA)*, 13(1), 263–287. <https://doi.org/10.58379/SLVL8621>
- Ryan, M., Khosronejad, M., Barton, G., Myhill, D., & Kervin, L. (2022). Reflexive writing dialogues: Elementary students' perceptions and performances as writers during classroom experiences. *Assessing Writing*, 51, Article 100592. <https://doi.org/10.1016/j.asw.2021.100592>
- Ryu, J., Ae Kim, Y., Eum, S., Park, S., Chun S., & Yang, S. (2022). The assessment of memes as digital multimodal composition in L2 classrooms. *Journal of Second Language Writing*, 57, Article 100914. <https://doi.org/10.1016/j.jslw.2022.100914>
- Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes*, 61, Article 101208. <https://doi.org/10.1016/j.jeap.2022.101208>
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50–61. <https://doi.org/10.1016/j.asw.2017.09.003>
- Saeed, M. A., & Alharbi, M. A. (2023). Towards fostering Saudi EFL learners' collaborative engagement and feedback literacy in writing. *Assessing Writing*, 56, Article 100721. <https://doi.org/10.1016/j.asw.2023.100721>
- Saeed, M. A., AbuSa'aleek, A., & Alharbi, M. A. (2024). Examining teacher's evaluative language in written, audio and screencast feedback on EFL learners' writing from the appraisal framework: A linguistic perspective. *Assessing Writing*, 61, 100871. <https://doi.org/10.1016/j.asw.2024.100871>
- Saeli, H., & Rahmati, P. (2023). Developments in learners' affective engagement with written peer feedback: The affordances of in situ translanguaging. *Assessing Writing*, 58, Article 100788. <https://doi.org/10.1016/j.asw.2023.100788>
- Safari, F., & Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *Journal of Second Language Writing*, 60, Article 101007. <https://doi.org/10.1016/j.jslw.2023.101007>

- Şahin, A. (2021). The Rater Performance Categorization System (RPCS) for intensive English programs. *Shanlax International Journal of Education*, 9(3), 225–241. <https://doi.org/10.34293/education.v9i3.3986>
- Sánchez-Pérez, M. D. M. (2021). Predicting content proficiency through disciplinary-literacy variables in English-medium writing. *System*, 97, Article 102463. <https://doi.org/10.1016/j.system.2021.102463>
- Sarte, K. M., & Gnevsheva, K. (2022). Noun phrasal complexity in ESL written essays under a constructed-response task: Examining proficiency and topic effects. *Assessing Writing*, 51, Article 100595. <https://doi.org/10.1016/j.asw.2021.100595>
- Sasayama, S., Garcia Gomez, P., & Norris, J. M. (2021). Designing efficient L2 writing assessment tasks for low-proficiency learners of English. *ETS Research Report Series*, 2021(1), 1–31. <https://doi.org/10.1002/ets2.12341>
- Sato, T. (2022). Assessing critical thinking through L2 argumentative essays: an investigation of relevant and salient criteria from raters' perspectives. *Language Testing in Asia*, 12, Article 9. <https://doi.org/10.1186/s40468-022-00159-4>
- Sato, T. (2024). Assessing the content quality of essays in content and language integrated learning: Exploring the construct from subject specialists' perspectives. *Language Testing*, 41(2), 316–337. <https://doi.org/10.1177/02655322231190058>
- Sawaki, Y., Ishii, Y., & Oi, Y. (2024). The efficacy of automated feedback on summary content in an online formative assessment module for undergraduate L2 writing instruction in Japan. *Studies in Language Assessment (SILA)*, 13(2), 38–72. <https://doi.org/10.58379/mjfp3942>
- Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency, and task effects in Spanish Dual Language Immersion. *Language Learning & Technology*, 25(1), 53–72. <https://hdl.handle.net/10125/73425>
- Screti, F. (2023). Using google docs for guided academic writing assessments: Students' perspectives. *Language Learning in Higher Education*, 13(1), 153–174. <https://doi.org/10.1515/cercles-2023-2007>
- Shan, Z., Yang, H., & Xu, H. (2023). The mediating role of curriculum configuration on teacher's L2 writing assessment literacy and practices in embedded French writing. *Assessing Writing*, 57, Article 100742. <https://doi.org/10.1016/j.asw.2023.100742>
- Shen, B., & Bai, B. (2024). Enhancing Chinese university students' writing performance and self-regulated learning (Srl) writing strategy use through a strategy-based intervention. *System*, 122, 103249. <https://doi.org/10.1016/j.system.2024.103249>
- Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361–388.
- Shi, G., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*. Advance online publication. <https://doi.org/10.1017/S0958344023000265>
- Shi, H., & Aryadoust, V. (2023). A systematic review of automated writing evaluation within the argument-based validation framework. *Education and Information Technology*, 28, 771–795. <https://link.springer.com/article/10.1007/s10639-022-11200-7>
- Shi, X., Ma, X., Du, W., & Gao, X. (2024). Diagnosing Chinese EFL learners' writing ability using polytomous cognitive diagnostic models. *Language Testing*, 41(1), 109–134. <https://doi.org/10.1177/02655322231162840>
- Shi, Z., Liu, F., Lai, C., & Jin, T. (2022). Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback. *Language Learning & Technology*, 26(2), 106–128. <https://doi.org/10125/73481>
- Shin, D., Kwon, S. K., & Lee, Y. (2021). The effect of using online language-support resources on L2 writing performance. *Language Testing in Asia*, 11, Article 4. <https://doi.org/10.1186/s40468-021-00119-4>

- Shin, J., & Gierl, M. J. (2021). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 38(2), 247–272. <https://doi.org/10.1177/0265532220937830>
- Shrestha, P. N. (2022). Examining evaluative language used in assessment feedback on business students' academic writing. *Assessing Writing*, 54, Article 100664. <https://doi.org/10.1016/j.asw.2022.100664>
- Siekmann, L., Parr, J. M., & Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing*, 54, Article 100672. <https://doi.org/10.1016/j.asw.2022.100671>
- Sims, M. L. (2023). Shifting perceptions of socially just writing assessment: Labor-based contract grading and multilingual writing instruction. *Assessing Writing*, 57, Article 100731. <https://doi.org/10.1016/j.asw.2023.100731>
- Sinha, T. S., & Nassaji, H. (2022). ESL learners' perception and its relationship with the efficacy of written corrective feedback. *International Journal of Applied Linguistics*, 32(1), 41–56. <https://doi.org/10.1111/ijal.12378>
- Sippel, L., & Martin, I. A. (2024). Is corrective feedback during telecollaboration beneficial? The effects of peer and teacher corrections on L2 writing proficiency. *Journal of Second Language Writing*, 64, Article 101098. <https://doi.org/10.1016/j.jslw.2024.101098>
- Skar, G. B., & Huebner, A. (2022). Predicting first grade students' writing proficiency. *Assessment in Education: Principles, Policy & Practice*, 29(2), 219–237. <https://doi.org/10.1080/0969594X.2022.2057424>
- Sok, S., & Shin, H. W. (2023). Investigating the dimensions and determinants of children's narrative writing in Korean. *Assessing Writing*, 57, Article 100740. <https://doi.org/10.1016/j.asw.2023.100740>
- Steiss, J., Krishnan, J., Kim, Y.-S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, Article 100600. <https://doi.org/10.1016/j.asw.2021.100600>
- Su, H., & Lu, X. (2022). Assessing pragmatic performance in advanced L2 academic writing through the lens of local grammars: A case study of “exemplification.” *Assessing Writing*, 54, Article 100668. <https://doi.org/10.1016/j.asw.2022.100668>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Suhan, M., Papageorgiou, S., & Wolf, M. K. (2024). Mapping the scores of the TOEFL Primary® Writing test to the Common European Framework of Reference (CEFR) levels (Research Memorandum ETS RM-24-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-03.pdf>
- Sun, Y., Lan, G., & Zhang, L. (2023). Pedagogical values of translingual practices in improving student feedback literacy in academic writing. *Assessing Writing*, 56, Article 100715. <https://doi.org/10.1016/j.asw.2023.100715>
- Sung, H., Cho, S., & Kyle, K. (2024). An Empirical Evaluation of Lexical Diversity Indices in L2 Korean Writing Assessment. *Language Assessment Quarterly*, 21(2), 159–180. <https://doi.org/10.1080/15434303.2024.2311728>
- Tabari, M. A., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing*, 57, Article 100749. <https://doi.org/10.1016/j.asw.2023.100749>
- Tabari, M. A., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing*, 58, Article 100783. <https://doi.org/10.1016/j.asw.2023.100783>
- Tabari, M. A., Wang, Y., & Miller, M. (2024). Understanding the SSARC model of task sequencing: Assessing L2 writing development. *Assessing Writing*, 62, 100893. <https://doi.org/10.1016/j.asw.2024.100893>

- Talebinamvar, M., & Zarrabi, F. (2022). Clustering students' writing behaviors using keystroke logging: a learning analytic approach in EFL writing. *Language Testing in Asia*, 12, Article 6. <https://doi.org/10.1186/s40468-021-00150-5>
- Tan, T. X., Fan, X., Braunstein, L. B., & Lane-Holbert, M. (2022). Linguistic, cultural and substantive patterns in L2 writing: A qualitative illustration of MisLevy's sociocognitive perspective on assessment. *Assessing Writing*, 51, Article 100574. <https://doi.org/10.1016/j.asw.2021.100574>
- Tang, Y., & Ma, X. (2023). An interventionist dynamic assessment approach to college English writing in China. *Language Assessment Quarterly*, 20(1), 44-65. <https://doi.org/10.1080/15434303.2022.2155165>
- Tavassoli, K., & Rahmatollahi, M. (2024). Differential impact of asynchronous and synchronous computer-assisted dynamic assessment on higher-order and lower-order writing skills. *Journal of Language Horizons*, 8(1), 33-61. <https://doi.org/10.22051/lghor.2023.41130.1720>
- Tavassoli, K., Oskouiefar, M., & Ghamoushi, M. (2023). The effect of mobile-assisted learning-oriented assessment on EFL learners' writing ability. *Iranian Journal of Applied Linguistics*, 26(2), 6-6. <http://ijal.khu.ac.ir/article-1-3233-en.html>
- Teng, L. S. (2022). Explicit strategy-based instruction in L2 writing contexts: A perspective of self-regulated learning and formative assessment. *Assessing Writing*, 53, Article 100645. <https://doi.org/10.1016/j.asw.2022.100645>
- Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing*, 59, 100811. <https://doi.org/10.1016/j.asw.2024.100811>
- Teng, M. F., & Zhan, Y. (2023). Assessing self-regulated writing strategies, self-efficacy, task complexity, and performance in English academic writing. *Assessing Writing*, 57, Article 100728. <https://doi.org/10.1016/j.asw.2023.100728>
- Teng, M. F., Wang, C., & Zhang, L. J. (2022). Assessing self-regulatory writing strategies and their predictive effects on young EFL learners' writing performance. *Assessing Writing*, 51, Article 100573. <https://doi.org/10.1016/j.asw.2021.100573>
- Thi, N. K., Nikolov, M., & Simon, K. (2023). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 17(3), 690-705. <https://doi.org/10.1080/17501229.2022.2122476>
- Thwaites, P., Kollias, C., & Paquot, M. (2024). Is CJ a valid, reliable form of L2 writing assessment when texts are long, homogeneous in proficiency, and feature heterogeneous prompts? *Assessing Writing*, 60, 100843. <https://doi.org/10.1016/j.asw.2024.100843>
- Tinkle, T., Godfrey, J., Menon, A. R., Moos, A., Romaine, L., & Sprouse, M. (2022). (In)Equities in directed self-placement. *Assessing Writing*, 54, Article 100671. <https://doi.org/10.1016/j.asw.2022.100671>
- Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing*, 52, Article 100609. <https://doi.org/10.1016/j.asw.2022.100609>
- Uludag, P., Crawford, W. J., & McDonough, K. (2021). English for academic purposes writers' use of reporting verbs in argumentative and cause-and-effect essay exams. *BC TEAL Journal*, 6(1), 42-58. <https://doi.org/10.14288/bctj.v6i1.388>
- Uludag, P., McDonough, K., & Payant, C. (2021). Does prewriting planning positively impact English L2 students' integrated writing performance? *Canadian Journal of Applied Linguistics*, 24(3), 166-185. <https://doi.org/10.37213/cjal.2021.31313>
- Uzun, K. (2021). Performance prediction strengths of noun and verb phrases in L2 writing: Comparison of density and complexity variables. *Assessing Writing*, 50, Article 100572. <https://doi.org/10.1016/j.asw.2021.100572>
- Vakili, S., & Ebadi, S. (2022). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer*

- Assisted Language Learning, 35(3), 345-380.
<https://doi.org/10.1080/09588221.2019.1698616>
- Valentine, K. A., Truckenmiller, A. J., Troia, G. A., & Aldridge, S. (2021). What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change? *Assessing Writing*, 50, Article 100567.
<https://doi.org/10.1016/j.asw.2021.100567>
- Vanderwoude, A., & Shefarevich, L. (2024). Investigating the quality of source integration in an academic writing exam. *Studies in Language Assessment (SILA)*, 13(1), 37-68.
<https://doi.org/10.58379/CUUO1529>
- Vasu, K. A., Mei Fung, Y., Nimehchisalem, V., & Md Rashid, S. (2022). Self-regulated learning development in undergraduate ESL writing classrooms: Teacher feedback versus self-assessment. *RELC Journal*, 53(3), 612–626.
<https://doi.org/10.1177/0033688220957782>
- Vo, Y., Rickels, H., Welch, C., & Dunbar, S. (2023). Human scoring versus automated scoring for English learners in a statewide evidence-based writing assessment. *Assessing Writing*, 56, Article 100719. <https://doi.org/10.1016/j.asw.2023.100719>
- Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 17(1), 47–71. <https://doi.org/10.1080/17501229.2021.1914062>
- Wang, D., Zhang, L., & Huang, Y. (2024). Tracking the development of logical metaphor usage in argumentative writing: A longitudinal study with EFL learners. *System*, 122, 103292. <https://doi.org/10.1016/j.system.2024.103292>
- Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing*, 50, Article 100571.
<https://doi.org/10.1016/j.asw.2021.100571>
- Wang, X., & Wang, J. (2024). Comparing Chinese L2 writing performance in paper-based and computer-based modes: Perspectives from the writing product and process. *Assessing Writing*, 61, 100849. <https://doi.org/10.1016/j.asw.2024.100849>
- Wang, Y., & Jiang, J. (2024). Exploring the development of noun phrase complexity in L2 English writings across two genres. *Assessing Writing*, 62, 100892.
<https://doi.org/10.1016/j.asw.2024.100892>
- Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, 53, Article 100641.
<https://doi.org/10.1016/j.asw.2022.100641>
- Wang, Y., & Xie, Q. (2022). Diagnostic assessment of novice EFL learners' discourse competence in academic writing: a case study. *Language Testing in Asia*, 12, Article 47. <https://doi.org/10.1186/s40468-022-00197-y>
- Ward, E. (2021). An integrated mixed-methods study of contract grading's impact on adolescents' perceptions of stress in high school English: a pilot study. *Assessing Writing*, 48, Article 100508. <https://doi.org/10.1016/j.asw.2020.100508>
- Ward, E. (2021). Easing stress: Contract grading's impact on adolescents' perceptions of workload demands, time constraints, and challenge appraisal in high school English. *Assessing Writing*, 48, Article 100526. <https://doi.org/10.1016/j.asw.2021.100526>
- Watzinger-Tharp, J., Tharp, D. S., & Rubio, F. (2021). Sustaining dual language immersion: Partner language outcomes in a statewide program. *The Modern Language Journal*, 105(1), 194–217. <https://doi.org/10.1111/modl.12694>
- Wei, R., & Zhao, X. (2024). Effects of task-based language teaching on functional adequacy in L2 writing. *Assessing Writing*, 60, 100838.
<https://doi.org/10.1016/j.asw.2024.100838>
- Weng, F., Zhao, C. G., & Chen, S. (2024). Effects of peer feedback in English writing classes on EFL students' writing feedback literacy. *Assessing Writing*, 61, 100874.
<https://doi.org/10.1016/j.asw.2024.100874>

- Westbrook, C. (2023). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61, Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Wheeler, J. M., Engelhard, G., & Wang, J. (2022). Exploring rater accuracy using unfolding models combined with topic models: Incorporating supervised latent Dirichlet allocation. *Measurement: Interdisciplinary Research and Perspectives*, 20(1), 34–46. <https://doi.org/10.1080/15366367.2021.1915094>
- Wilson, J., & Huang, Y. (2024). Validity of automated essay scores for elementary-age English language learners: Evidence of bias?. *Assessing Writing*, 60, 100815. <https://doi.org/10.1016/j.asw.2024.100815>
- Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, 29(2), 183–199. <https://doi.org/10.1080/0969594X.2022.2025762>
- Wilson, J., Potter, A., Cordero, T. C., & Myers, M. C. (2023). Integrating goal-setting and automated feedback to improve writing outcomes: A pilot study. *Innovation in Language Learning and Teaching*, 17(3), 518–534. <https://doi.org/10.1080/17501229.2022.2077348>
- Wolf, M. K., & Oh, S. (2024). Investigating the types and use of feedback in middle-school English language learners' academic writing. *Language Learning & Technology*, 28(1), 1-25. <https://hdl.handle.net/10125/73583>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). Exploring the provision of AI-based feedback for TOEFL Junior Writing practice tasks (Research Memorandum No. RM 24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>
- Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). Design framework for the TOEFL Primary® Writing test (Research Memorandum ETS RM-24-02). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-02.pdf>
- Woods, K., Hashimoto, B., & Brown, E. K. (2023). A multi-measure approach for lexical diversity in writing assessments: Considerations in measurement and timing. *Assessing Writing*, 55, Article 100688. <https://doi.org/10.1016/j.asw.2022.100688>
- Wu, J., & Lu, X. (2024). Noun phrase complexity and second language Chinese proficiency: An analysis of narratives by Korean learners of Chinese. *Assessing Writing*, 59, 100810. <https://doi.org/10.1016/j.asw.2024.100810>
- Wu, M. Y., Steinkrauss, R., & Lowie, W. (2023). The reliability of single task assessment in longitudinal L2 writing research. *Journal of Second Language Writing*, 59, 100950. <https://doi.org/10.1016/j.jslw.2022.100950>
- Wu, P., & Lei, C. (2023). Developing feedback literacy through dialogue-supported performances of multi-draft writing in a postgraduate class. *Assessing Writing*, 57, Article 100759. <https://doi.org/10.1016/j.asw.2023.100759>
- Wu, P., Yu, S., & Luo, Y. (2023). The development of teacher feedback literacy in situ: EFL writing teachers' endeavor to human-computer-AWE integral feedback innovation. *Assessing Writing*, 57, Article 100739. <https://doi.org/10.1016/j.asw.2023.100739>
- Wu, Y-j. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology*, 25(2), 192–214. <http://hdl.handle.net/10125/73440>
- Xia, D., Chen, Y., & Pae, H. (2023). Lexical and grammatical collocations in beginning and intermediate L2 argumentative essays: a bigram study. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1421-1453. <https://doi.org/10.1515/iral-2021-0188>
- Xie, Q. (2024). Construct representation and predictive validity of integrated writing tasks: A study on the writing component of the Duolingo English Test. *Assessing Writing*, 61, 100846. <https://doi.org/10.1016/j.asw.2024.100846>
- Xie, Q., & Lei, Y. (2022). Diagnostic assessment of L2 academic writing product, process and self-regulatory strategy use with a comparative dimension. *Language*

- Assessment Quarterly*, 19(3), 231-263.
<https://doi.org/10.1080/15434303.2021.1903470>
- Xu, J. (2022). Incremental intelligence matters: How L2 writing mindsets impact feedback orientation and self-regulated learning writing strategies. *Assessing Writing*, 51, Article 100593. <https://doi.org/10.1016/j.asw.2021.100593>
- Xu, J., & Wang, Y. (2023). Individual differences in L2 writing feedback-seeking behaviors: The predictive roles of various motivational constructs. *Assessing Writing*, 56, Article 100698. <https://doi.org/10.1016/j.asw.2023.100698>
- Xu, J., & Wang, Y. (2024). Do I need feedback or avoid it in L2 writing? Impacts of self-efficacy and shyness on feedback-seeking behavior. *Assessing Writing*, 59, 100805. <https://doi.org/10.1016/j.asw.2023.100805>
- Xu, J., & Zheng, Y. (2024). Understanding the interconnectedness of L2 learners' writing assessment literacy and engagement: A longitudinal survey. *System*, 123, 103318. <https://doi.org/10.1016/j.system.2024.103318>
- Xu, Q., Sawaki, Y., & Zhu, Y. (2024). Linguistic complexity measures of short narrative writings for modeling overall Chinese proficiency. *System*, 121, 103220. <https://doi.org/10.1016/j.system.2024.103220>
- Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2023). A multidimensional approach to assessing the effects of task complexity on L2 students' argumentative writing. *Assessing Writing*, 55, Article 100690. <https://doi.org/10.1016/j.asw.2022.100690>
- Xu, Z., Zhang, L., & Parr, J. (2023). Incorporating peer feedback in writing instruction: examining its effects on Chinese English-as-a-foreign-language (EFL) learners' writing performance. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1337-1364. <https://doi.org/10.1515/iral-2021-0078>
- Yamashita, T. (2024). An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, 3(3), Article 100133. <https://doi.org/10.1016/j.rmal.2024.100133>
- Yan, X., & Lin, J. (2023). Chinese character matters!: An examination of linguistic accuracy in writing performances on the HSK test. *Assessing Writing*, 57, Article 100767. <https://doi.org/10.1016/j.asw.2023.100767>
- Yang, Y., Yap, N. T., & Ali, M. A. (2023). Predicting EFL expository writing quality with measures of lexical richness. *Assessing Writing*, 57, Article 100762. <https://doi.org/10.1016/j.asw.2023.100762>
- Yao, Y., Guo, N. S., Li, C., & McCampbell, D. (2021). How university EFL writers' beliefs in writing ability impact their perceptions of peer assessment: Perspectives from implicit theories of intelligence. *Assessment and Evaluation in Higher Education*, 46(1), 151-167. <https://doi.org/10.1080/02602938.2020.1750559>
- Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, 50, Article 100568. <https://doi.org/10.1016/j.asw.2021.100568>
- Yao, Y., Zhu, X., & Zhan, J. (2024). The associations among growth mindsets, the ideal L2 writing self, and L2 writing enjoyment and their impacts on L2 English writing performance: A gender difference perspective. *Assessing Writing*, 60, 100832. <https://doi.org/10.1016/j.asw.2024.100832>
- Yasuda, S. (2024). Does "more complexity" equal "better writing"? Investigating the relationship between form-based complexity and meaning-based complexity in high school EFL learners' argumentative writing. *Assessing Writing*, 61, 100867. <https://doi.org/10.1016/j.asw.2024.100867>
- Yu, C. (2023). Integrating information literacy and academic writing: Developing a self-assessment scale of information-based academic writing. *The Journal of Academic Librarianship*, 47(6), 102804. <https://doi.org/10.1016/j.acalib.2023.102804>

- Yu, C. (2023). Testing issues in writing. In H. Mohebbi, & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 56–70). Castledown. <https://doi.org/10.29140/9781914291159-5>
- Yu, C., & Xu, W. (2024). Writing assessment literacy and its impact on the learning of writing: A netnography focusing on Duolingo English Test examinees. *Language Testing in Asia*, 14, Article 24. <https://doi.org/10.1186/s40468-024-00297-x>
- Yu, L., & Poehner, M. E. (2023). Developing L2 writing abilities through an instructional enrichment program informed by dynamic assessment. *Journal of Second Language Writing*, 60, 101013. <https://doi.org/10.1016/j.jslw.2023.101013>
- Yu, S. (2021). Giving genre-based peer feedback in academic writing: sources of knowledge and skills, difficulties and challenges. *Assessment and Evaluation in Higher Education*, 46(1), 36–53. <https://doi.org/10.1080/02602938.2020.1742872>
- Yu, S., & Jiang, L. (2023). L2 university students' motivational self system in English writing: a sociocultural inquiry. *Applied Linguistics Review*, 14(3), 553-578. <https://doi.org/10.1515/applirev-2020-0110>
- Yu, S., & Lee, I. (2024). Writing assessment and feedback literacy: Where do we stand and where can we go? *Assessing Writing*, 60, 100829. <https://doi.org/10.1016/j.asw.2024.100829>
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48, Article 100525. <https://doi.org/10.1016/j.asw.2021.100525>
- Yu, S., Geng, F., Liu, C., & Zheng, Y. (2021). What works may hurt: The negative side of feedback in second language writing. *Journal of Second Language Writing*, 54, Article 100850. <https://doi.org/10.1016/j.jslw.2021.100850>
- Yu, S., Zhang, E. D., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing*, 53, Article 100643. <https://doi.org/10.1016/j.asw.2022.100643>
- Yu, S., Zheng, Y., Jiang, L., Liu, C., & Xu, Y. (2021). “I even feel annoyed and angry”: Teacher emotional experiences in giving feedback on student writing. *Assessing Writing*, 48, Article 100528. <https://doi.org/10.1016/j.asw.2021.100528>
- Yuan, Y., Engelhard, G., Raczynski, K., & Wang, J. (2022). Using unfolding models to identify targeted feedback strategies for student writing. *Assessing Writing*, 54, Article 100670. <https://doi.org/10.1016/j.asw.2022.100670>
- Zarrinabadi, N., Mahmoudi-Gahrouei, V., & Mohammadabadi, A. M. (2023). Your writing could have been better: Examining the effects of upward and downward counterfactual communication on the motivational aspects of L2 writing. *Assessing Writing*, 56, Article 100714. <https://doi.org/10.1016/j.asw.2023.100714>
- Zeng, X., Liu, J., & Bowen, N. E. J. A. (2024). A comparative study of voice in Chinese English-major undergraduates' timed and untimed argument writing. *Assessing Writing*, 62, 100896. <https://doi.org/10.1016/j.asw.2024.100896>
- Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47, Article 100505. <https://doi.org/10.1016/j.asw.2020.100505>
- Zesch, T., Horbach, A., & Zehner, F. (2023). To score or not to score: Factors influencing performance and feasibility of automatic content scoring of text responses. *Educational Measurement, Issues and Practice*, 42(1), 44–58. <https://doi.org/10.1111/emip.12544>
- Zhang-Wu, Q. (2023). Exploring multilingual students' feedback literacy in an asynchronous online writing course. *Assessing Writing*, 56, Article 100718. <https://doi.org/10.1016/j.asw.2023.100718>
- Zhang, E. D., & Yu, S. (2024). Understanding L2 student writers' self-assessment in digital multimodal composing: A process-oriented approach. *System*, 121, 103219. <https://doi.org/10.1016/j.system.2024.103219>

- Zhang, F., Min, H., He, P., Chen, S., & Ren, S. (2023). Understanding EFL students' feedback literacy development in academic writing: A longitudinal case study. *Assessing Writing*, 58, Article 100770. <https://doi.org/10.1016/j.asw.2023.100770>
- Zhang, L. J., & Zhang, J. (2024). EFL students' syntactic complexity development in argumentative writing: A latent class growth analysis (LCGA) approach. *Assessing Writing*, 61, 100877. <https://doi.org/10.1016/j.asw.2024.100877>
- Zhang, L., & Zhang, L. (2023). Development and validation of the questionnaire on EFL students' perceptions of authorial stance in academic writing. *Applied Linguistics Review*, 14(5), 1219–1251. <https://doi.org/10.1515/applirev-2021-0003>
- Zhang, M., & Chen, W. (2022). Assessing collaborative writing in the digital age: An exploratory study. *Journal of Second Language Writing*, 57, Article 100868. <https://doi.org/10.1016/j.jslw.2022.100868>
- Zhang, M., & Sinharay, S. (2022). Investigating the writing performance of educationally at-risk examinees using technology. *International Journal of Testing*, 22(3-4), 312-347. <https://doi.org/10.1080/15305058.2022.2050734>
- Zhang, P., Fan, J., & Jia, W. (2022). Assessing L2 integrated writing self-efficacy: Development and validation of a scale. *Assessing Writing*, 54, Article 100665. <https://doi.org/10.1016/j.asw.2022.100665>
- Zhang, T., & Mao, Z. (2023). Exploring the development of student feedback literacy in the second language writing classroom. *Assessing Writing*, 55, Article 100697. <https://doi.org/10.1016/j.asw.2023.100697>
- Zhang, T., & Weng, Z. (2024). Implementing rubric co-construction in ESL writing teaching. *ELT Journal*, 78(1), 32-41. <https://doi.org/10.1093/elt/ccad033>
- Zhang, X. (2021). Assessing EFL students' writing development as they are exposed to the integrated use of drama-based pedagogy and SFL-based teaching. *Assessing Writing*, 50, Article 100569. <https://doi.org/10.1016/j.asw.2021.100569>
- Zhang, X. (2022). The relationship between lexical use and L2 writing quality: A case of two genres. *International Journal of Applied Linguistics*, 32(3), 371-396. <https://doi.org/10.1111/ijal.12420>
- Zhang, X., & Li, W. (2021). Effects of n-grams on the rated L2 writing quality of expository essays: A conceptual replication and extension. *System*, 97, Article 102437. <https://doi.org/10.1016/j.system.2020.102437>
- Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, 51, Article 100597. <https://doi.org/10.1016/j.asw.2021.100597>
- Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing*, 60, 100842. <https://doi.org/10.1016/j.asw.2024.100842>
- Zhang, Y., & Ouyang, J. (2023). Linguistic complexity as the predictor of EFL independent and integrated writing quality. *Assessing Writing*, 56, Article 100727. <https://doi.org/10.1016/j.asw.2023.100727>
- Zhang, Y., & Xi, J. (2023). Fostering self-regulated young writers: Dynamic assessment of metacognitive competence in secondary school EFL class. *Language Assessment Quarterly*, 20(1), 88–107. <https://doi.org/10.1080/15434303.2022.2103702>
- Zhang, Z. V., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, 58, Article 100784. <https://doi.org/10.1016/j.asw.2023.100784>
- Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, Article 100586. <https://doi.org/10.1016/j.asw.2021.100586>
- Zhao, C. G., & Wu, J. (2022). Perceptions of authorial voice: Why discrepancies exist. *Assessing Writing*, 53, Article 100632. <https://doi.org/10.1016/j.asw.2022.100632>

- Zhao, H., & Zhao, B. (2023). Co-constructing the assessment criteria for EFL writing by instructors and students: A participative approach to constructively aligning the CEFR, curricula, teaching and learning. *Language Teaching Research*, 27(3), 765–793. <https://doi.org/10.1177/1362168820948458>
- Zhi, M., & Huang, B. (2021). Investigating the authenticity of computer-and paper-based ESL writing tests. *Assessing Writing*, 50, Article 100548. <https://doi.org/10.1016/j.asw.2021.100548>
- Zhou, J., Deneen, C., Tai, J., & Dawson, P. (2023). Feedback seeking by first-year Chinese international students: Understanding practices and challenges. *Assessing Writing*, 57, Article 100757. <https://doi.org/10.1016/j.asw.2023.100757>
- Zhou, Q., Chen, L., & Hou, X. (2022). Exploring Chinese EFL undergraduates' writing from sources: Self-efficacy and performance. *Assessing Writing*, 54, Article 100663. <https://doi.org/10.1016/j.asw.2022.100663>
- Zhu, M., Zhang, M., & Gu, L. (2023). Insights into editing and revising in writing process using keystroke logs. *Language Assessment Quarterly*, 20(4-5), 445–468. <https://doi.org/10.1080/15434303.2023.2291478>
- Zhu, X., Guan, Y., & Yao, Y. (2022). Transfer of ideal L1 and L2 writing selves and their impacts on L2 writing enjoyment and integrated writing performance. *Assessing Writing*, 54, Article 100674. <https://doi.org/10.1016/j.asw.2022.100674>
- Zhu, X., Li, G. Y., Cheong, C. M., & Wen, H. (2021). Effects of L1 single-text and multiple-text comprehension on L2 integrated writing. *Assessing Writing*, 49, Article 100546. <https://doi.org/10.1016/j.asw.2021.100546>
- Zou, S., Yan, X., & Fan, J. (2024). Establishing analytic score profiles for large-scale L2 writing assessment: The case of the CET-4 writing test. *Assessing Writing*, 60, 100826. <https://doi.org/10.1016/j.asw.2024.100826>

Assessing Integrated Skills

- Barkaoui, K. (2024). Examining performance on an integrated writing task from a Canadian English language proficiency test. *Canadian Modern Language Review*, 80(2), 77–115. <https://doi.org/10.3138/cmlr-2023-0022>
- Chan, S., & Yamashita, J. (2022). Integrated writing and its correlates: A meta-analysis. *Assessing Writing*, 54, Article 100662. <https://doi.org/10.1016/j.asw.2022.100662>
- Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487–2512. <https://doi.org/10.1177/13621688211046740>
- Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, 60, Article 101003. <https://doi.org/10.1016/j.jslw.2023.101003>
- Cliff, A. (2021). The use of mediation and feedback in a standardised test of academic literacy: Theoretical and design considerations. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 75–92). Multilingual Matters.
- Davis, L., & Norris, J. (2021). Developing an innovative elicited imitation task for efficient English proficiency assessment. *ETS Research Report Series*, 2021(1), 1–30. <https://doi.org/10.1002/ets2.12338>
- Delgado-Osorio, X., Koval, V., Hartig, J., & Harsch, C. (2023). Strategic processing of source text in reading-into-writing tasks: A comparison between summary and argumentative tasks. *Journal of English for Academic Purposes*, 62, Article 101227. <https://doi.org/10.1016/j.jeap.2023.101227>

- Douglas, D. (2023). Specificity, authenticity, and inseparability: Assessing integrated oral skills for pilots and air traffic controllers. *Applied Linguistics Papers*, 27(1), 4–13. <https://doi.org/10.32612/uw.25449354>
- Frost, K., Wigglesworth, G., & Clothier, J. (2021). Relationships between comprehension, strategic behaviours and content-related aspects of test performances in integrated speaking tasks. *Language Assessment Quarterly*, 18(2), 133–153. <https://doi.org/10.1080/15434303.2020.1835918>
- Golparvar, S. E., & Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assessing Writing*, 47, Article 100504. <https://doi.org/10.1016/j.asw.2020.100504>
- Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple-text source-based writing. *System*, 99, Article 102524. <https://doi.org/10.1016/j.system.2021.102524>
- Green, J., Weideman, A., Davis, C., Judith, K., & Harmes, M. (2024). Using a five-phase applied linguistics design to develop a contextualized academic literacy placement test for pre-university pathway students. *Literacy Research and Instruction*, 1-27. <https://doi.org/10.1080/19388071.2024.2340031>
- Hammad Al-Rashidi, A., Vadivel, B., Ramadan Khalil, N., & Basim, N. (2023). The comparative impacts of portfolio-based assessment, self-assessment, and scaffolded peer assessment on reading comprehension, vocabulary learning, and grammatical accuracy: insights from working memory capacity. *Language Testing in Asia*, 13, Article 24 (2023). <https://doi.org/10.1186/s40468-023-00237-1>
- Harsch, C., Koval, V., Delgado-Osorio, X. & Hartig, J. (2024). Usability of CEFR Companion Volume scales for the development of an analytic rating scale for academic integrated writing assessment. In *CEFR Journal - Research and Practice*. Vol. 6, 155-177, <https://doi.org/10.37546/JALTSIG.CEFR6-9>
- Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing*, 62, 100894. <https://doi.org/10.1016/j.asw.2024.100894>
- Harvey, R. (2022). *Implementing an integrated skills test in a Direct Entry project* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Inoue, C., & Lam, D. M. K. (2021). *The effects of extended planning time on candidates' performance, processes, and strategy use in the lecture listening-into-speaking tasks of the TOEFL iBT Test*. (TOEFL Research Report, TOEFL-RR-93, ETS Research Report No. RR-21-09). <https://doi.org/10.1002/ets2.12322>
- Jia, W., & Zhang, P. (2023). Rater cognitive processes in integrated writing tasks: from the perspective of problem-solving. *Language Testing in Asia*, 13, Article 50 (2023). <https://doi.org/10.1186/s40468-023-00265-x>
- Karimi, M. N., & Mousavi-Moghadam, T. (2022). Integrated writing across conflicting L2 texts: the effects of prior beliefs on perspectivism and propositional content. *Assessing Writing*, 53, Article 100648. <https://doi.org/10.1016/j.asw.2022.100648>
- Kazemian, M. & Khonamri, F. (2022). Past, present, and future of language assessment: An interview with Dr. Hossein Farhady, *MEXTESOL Journal*, 46(4), 1–8. https://www.mextesol.net/journal/index.php?page=journal&id_article=46395
- Khabbazzashi, N., Chan, S. and Clark, T. (2023). Towards the new construct of academic English in the digital age, *ELT Journal*, 77(2), 207–216

- <https://doi.org/10.1093/elt/ccac010>
- Kim, H., & Hwang, H. (2022). Assessing verb-construction integration in young learners of English as a foreign language: analyses of written and spoken production. *Language Learning*, 72(2), 497-533. <https://doi.org/10.1111/lang.12480>
- Kormos, J., Eberharder, K., Guggenbichler, E., Baumgartinger, S., Ebner, V., & Kremmel, B. (2024). First Language Literacy Skills, Listening Strategy Use, and Anxiety in Self-Paced versus Single-Play L2 Listening Tests. *Language Assessment Quarterly*, 21(4-5), 303-323. <https://doi.org/10.1080/15434303.2024.2435811>
- Lee, K. (2023). Diagnosing Chinese college-level English as a Foreign Language (EFL) learners' integrated writing capability: A Log-linear Cognitive Diagnostic Modeling (LCDM) study. *Assessing Writing*, 57, Article 100730. <https://doi.org/10.1016/j.asw.2023.100730>
- Li, X., & Wang, Y. (2024). Effects of input text genre on Chinese young EFL learners' performance on continuation tasks. *Language Testing in Asia*, 14, Article 3. <https://doi.org/10.1186/s40468-024-00273-5>
- Liao, R. J. T., Ohta, R., & Lee, K. (2024). The relationship between written discourse features and integrated listening-to-write scores for adolescent English language learners. *Language Testing*, 41(1), 35-59. <https://doi.org/10.1177/02655322231167629>
- Lin, R. (2023). Examining the scoring of content integration in a listening-speaking Test: A G-Theory analysis. *Language Assessment Quarterly*, 20(3), 319-338. <https://doi.org/10.1080/15434303.2023.2242334>
- Liu, Y., Cheong, C. M., & Zhang, J. (2024). The Role of Achievement Goals and Strategy Use in Secondary School students' Integrated Writing Performance. *Language Assessment Quarterly*, 21(4-5), 324-349. <https://doi.org/10.1080/15434303.2024.2435818>
- Lombardi, A. (2021). More is more: Explicit intertextuality in university writing placement exam essays. *Journal of English for Academic Purposes*, 50, Article 100955. <https://doi.org/10.1016/j.jeap.2020.100955>
- Myburgh-Smit J. & Weideman, A. (2021). How early should we measure academic literacy? The usefulness of an appropriate test of academic literacy for Grade 10 students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 117-131). Multilingual Matters. <https://doi.org/10.21832/9781788926218-007>
- Nawas, A., Darmawan, I.N., & Maadad, N. (2023). Indonesian secular vs. Madrasah schools: assessing the discrepancy in English reading and listening tests. *Language Testing in Asia*, 13, Article 52 (2023). <https://doi.org/10.1186/s40468-023-00266-w>
- Niimi, N., & Matsuura, N. (2022). Assessing Japanese junior high school students' English achievement through computer-based testing in the classroom: A case of integrated reading-into-writing continuous task. *Language Testing in Asia*, 12, Article 38. <https://doi.org/10.1186/s40468-022-00189-y>
- Oliver, G. P., & Gregory, K. M. (2024). Integrated Performance Assessments: providing equitable instruction and assessment for ELLs/MLLs. *TESOL Quarterly*, 58(4), 1815-1830. <https://doi.org/10.1002/tesq.3324>
- Pan, R., & Lu, X. (2023). The design and cognitive validity verification of reading-to-write tasks in L2 Chinese writing assessment. *Assessing Writing*, 56, Article 100699. <https://doi.org/10.1016/j.asw.2023.100699>

- Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly*, 39, 372–425. <https://doi.org/10.32038/ltrq.2024.39.23>
- Panahi, A., Mohebbi, H., & Oxford, R. L. (2024). Rebecca Oxford's 50 years of contributions to language education and related fields: A systematic review. *Language Teaching Research Quarterly*, 41, 169–230.
- Rukthong, A. (2021). MC listening questions vs. integrated listening-to-summarize tasks: What listening abilities do they assess? *System*, 97, Article 102439. <https://doi.org/10.1016/j.system.2020.102439>
- Safari, F., & Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *Journal of Second Language Writing*, 60, Article 101007. <https://doi.org/10.1016/j.jslw.2023.101007>
- Sebolai, K. (2021). Generic academic literacy testing: A logical precursor for faculty-specific language teaching and assessment. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 151-169). Multilingual Matters.
- Steyn, S. (2021). Pathways to parity between parallel tests of language ability: Lessons from a project. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 132-147). Multilingual Matters.
- Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing*, 52, Article 100609. <https://doi.org/10.1016/j.asw.2022.100609>
- Van Dyk, T. (2021). Postscript: What the data tell us: an overview of language assessment research in South Africa's multilingual context. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 217–235). Multilingual Matters.
- Van Dyk, T., & Weideman, A. (2022). Taaltoetsing en die assessering van taalvermoë [Language testing and the assessment of language ability]. In W. Carstens & T. van Dyk (Eds.), *Toegepaste Taalkunde in Afrikaans* [Applied linguistics in Afrikaans] (pp. 125-1420). Van Schaik. <https://toegepaste-taalkunde.co.za/hoofstuk-7/>
- Vanderwoude, A., & Shefarevich, L. (2024). Investigating the quality of source integration in an academic writing exam. *Studies in Language Assessment (SILA)*, 13(1), 37-68. <https://doi.org/10.58379/CUUO1529>
- Weideman, A. (2021). A skills-neutral approach to academic literacy assessment. In A. Weideman, J. Read, & T. du Plessis, T. (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 22-51). Multilingual Matters.
- Weideman, A. 2023. The practicality of theory: Reciprocity, assessment and applied linguistics. *Stellenbosch Papers in Linguistics Plus*, 66, 77-196. <https://doi.org/10.5842/66-1-935>
- Weideman, A., Read, J. & Du Plessis, T. (Eds.). (2021). *Assessing academic literacy in a multilingual society: Transition and transformation* (Volume 84). Multilingual Matters. <https://doi.org/10.21832/WEIDEM6201>.
- Westbrook, C. (2023). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61,

- Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Xie, Q. Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing*, 57, Article 100755. <https://doi.org/10.1016/j.asw.2023.100755>
- Yao, Y., Zhu, X., Zhu, S., & Jiang, Y. (2023). The impacts of self-efficacy on undergraduate students' perceived task value and task performance of L1 Chinese integrated writing: A mixed-method research. *Assessing Writing*, 55, Article 100687. <https://doi.org/10.1016/j.asw.2022.100687>
- Zhang, W., & Zhang, L. J. (2022). Understanding assessment tasks: Learners' and teachers' perceptions of cognitive load of integrated speaking tasks for TBLT implementation. *System*, 111, Article 102951. <https://doi.org/10.1016/j.system.2022.102951>
- Zhang, W., Zhao, M., & Zhu, Y. (2022). Understanding individual differences in metacognitive strategy use, task demand, and performance in integrated L2 speaking assessment tasks. *Frontiers in Psychology*, 13, Article 876208. <https://doi.org/10.3389/fpsyg.2022.876208>
- Zhang, Y., & Ouyang, J. (2023). Linguistic complexity as the predictor of EFL independent and integrated writing quality. *Assessing Writing*, 56, Article 100727. <https://doi.org/10.1016/j.asw.2023.100727>

Assessing Sign Languages

- Batty, A. O., Haug, T., Ebling, S., Tissi, K., & Sidler-Miserez, S. (2023). Challenges in rating signed production: A mixed-methods study of a Swiss German Sign Language form-recall vocabulary test. *Language Testing*, 40(2), 352–374. <https://doi.org/10.1177/02655322221122774>
- Buchanan, B. J., Pirone, J. S., Pudans-Smith, K. K., Leffler, B. W., Cohen, B. S., Simmons, P. K., Ballentine, P. M., Venable, T., & Soza, K. (2023). American Sign Language level 1 assessment: What is out there? *Foreign Language Annals*, 56(4), 912–926. <https://doi.org/10.1111/flan.12717>
- Fan, J., Jin, T., & Chen, I. (2022). *Applying the Lexile Framework to evaluate text complexity in the Chinese context: A corpus-based study* (Report 1/2022). https://www.britishcouncil.org/sites/default/files/rirg_fan_jin_chen_layout_1.pdf
- Han, C., & Xiao, X. (2022). A comparative judgment approach to assessing Chinese Sign Language interpreting. *Language Testing*, 39(2), 289–312. <https://doi.org/10.1177/02655322211038977>
- Haug, T., Boers-Visker, E., & Van den Bogaerde, B. (2021). Testing sign language learners. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 432–442). Routledge. <https://doi.org/10.4324/9781351034784-46>
- Haug, T., Holzknacht, F., & Mann, W. (2024). The impact of COVID-19 on sign language assessment. *Language Education & Assessment*, 7(1), 1316–1316. <https://doi.org/10.29140/lea.v7n1.1316>
- Haug, T., Mann, W., & Knoch, U. (2022). *The handbook of language assessment across modalities*. Oxford University Press.
- Holmström, I., Schönström, K., & Ryttervik, M. (2023). Development of a Sign Repetition Task for Novice L2 Signers. *Language Assessment Quarterly*, 21(1), 33–59. <https://doi.org/10.1080/15434303.2023.2256320>

- Holzknicht, F., Tornay, S., Battisti, A., Batty, A. O., Tissi, K., Haug, T., & Ebling, S. (2024). Automated Sign Language Vocabulary Assessment: Comparing Human and Machine Ratings and Studying Learner Perceptions. *Language Assessment Quarterly*, 21(3), 245–265. <https://doi.org/10.1080/15434303.2024.2364877>
- Kotowicz, J., Woll, B., & Herman, R. (2021). Adaptation of the British Sign Language Receptive Skills Test into Polish Sign Language. *Language Testing*, 38(1), 132–153. <https://doi.org/10.1177/0265532220924598>
- Occhino, C., Lidster, R., Geer, L. C., Listman, J., & Hauser, P. C. (2024). Development of the American Sign Language Fingerspelling and Numbers Comprehension Test (ASL FaN-CT). *Language Testing*, 41(1), 135–161. <https://doi.org/10.1177/02655322231179494>

Assessing Vocabulary and Grammar

- Akase, M. (2022). Longitudinal measurement of growth in vocabulary size using Rasch-based test equating. *Language Testing in Asia*, 12, Article 5. <https://doi.org/10.1186/s40468-022-00155-8>
- Akbarian, I. H., & Farrokhi, J. (2021). Lexical availability and writing ability of EFL learners. *International Journal of Applied Linguistics*, 31(1), 48–64. <https://doi.org/10.1111/ijal.12319>
- Allen, D., & Nakamura, K. (2023). The distribution of cognates and their impact on response accuracy in the EIKEN tests. *Language Testing*, 40(3), 771–795. <https://doi.org/10.1177/02655322231158551>
- Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, 42(2), 292–314. <https://doi.org/10.1093/applin/amaa020>
- Ardasheva, Y., Crosson, A. C., Carbonneau, K. J., & French, B. F. (2021). Unpacking contributions of morphosyntactic awareness and vocabulary to science reading comprehension among linguistically diverse students. *TESOL Quarterly*, 55(3), 931–965. <https://doi.org/10.1002/tesq.3039>
- Atak, N., & Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. *Assessing Writing*, 47, Article 100506. <https://doi.org/10.1016/j.asw.2020.100506>
- Bao, G. (2024). Differential effects of EFL proficiency and mode on noun phrase use. *System*, 121, 103244. <https://doi.org/10.1016/j.system.2024.103244>
- Bonvin, A., Brugger, L., & Berthele, R. (2023). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*, 61(2), 257–285. <https://doi.org/10.1515/iral-2020-0093>
- Cai, J., Zhang, X., & Wang, Y. (2024). Effects of processing tasks on children's learning of second language words. *System*, 125, 103435. <https://doi.org/10.1016/j.system.2024.103435>
- Canning, D., McLean, S., & Vitta, J. (2024). Relative complexity in a model of word difficulty: The role of loanwords in vocabulary size tests. *Studies in Second Language Learning and Teaching*, 14(4), 631–659. <https://doi.org/10.14746/ssllt.38492>
- Carney, N. (2024). Referencing word and phrase lists in an online vocabulary notebook. *Language Teaching Research*, 28(1), 253–272.

- <https://doi.org/10.1177/1362168821997613>
- Chedid, G., & Stephan, M. (2024). Adaptation and norm determination of the Boston Naming Test for healthy Lebanese adults aged between 50 and 88 years. *Language Testing in Asia*, 14, Article 25. <https://doi.org/10.1186/s40468-024-00294-0>
- Chernova, S., & Abbasi, M. (2022). Using ipsative assessment in teaching and learning English grammar through e-portfolios (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Clark, T., & Endres, H. (2021). Online diagnostic grammar testing for high school students—An international trial. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 602–632. <https://doi.org/10.1080/0969594X.2021.1970513>
- Cobb, T., & Laufer, B. (2021). The nuclear word family list: A list of the most frequent family members, including base and affixed words. *Language Learning*, 71(3), 834–871. <https://doi.org/10.1111/lang.12452>
- Consolini, C. H., & Kyle, K. (2024). The relationship between L2 Spanish proficiency and features of written lexical and lexicogrammatical use. *Applied Linguistics*, 45(6), 992–1009. <https://doi.org/10.1093/applin/amae032>
- Deygers, B., & Vanbuel, M. (2022). Gauging the impact of literacy and educational background on receptive vocabulary test scores. *Language Testing*, 39(2), 191–211. <https://doi.org/10.1177/02655322211049097>
- Di Zhang, E., & Yu, S. (2023). Investigating the relationship between linguistic changes in L2 writers' paraphrasing, paraphrasing performance and L2 proficiency. *Applied Linguistics Review*, 14(5), 1451–1473. <https://doi.org/10.1515/applirev-2021-0014>
- Drown, L., Giovannone, N., Pisoni, D., & Theodore, R. (2023). Validation of two measures for assessing English vocabulary knowledge on web-based testing platforms: brief assessments. *Linguistics Vanguard*, 9(1), 99–111. <https://doi.org/10.1515/lingvan-2022-0116>
- Duan, S., & Shi, Z. (2024). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Language Teaching Research*, 28(2), 497–530. <https://doi.org/10.1177/13621688211002942>
- Dudley, A., & Marsden, E. (2024). The lexical content of high-stakes national exams in French, German, and Spanish in England. *Foreign Language Annals*, 57(2), 311–338.
- Dudley, A., Marsden, E., & Bovolenta, G. (2024). A Context-Aligned Two Thousand Test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England. *Language Testing*, 41(4), 759–791. <https://doi.org/10.1177/02655322241261415>
- Dunn, K. J. (2024). Random-item Rasch models and explanatory extensions: A worked example using L2 vocabulary test item responses. *Research Methods in Applied Linguistics*, 3(3), Article 100143. <https://doi.org/10.1016/j.rmal.2024.100143>
- Durrant, P., & Durrant, A. (2022). Appropriateness as an aspect of lexical richness: What do quantitative measures tell us about children's writing? *Assessing Writing*, 51, Article 100596. <https://doi.org/10.1016/j.asw.2021.100596>
- Eguchi, M., Suzuki, S., & Suzuki, Y. (2022). Lexical competence underlying second language word association tasks: Examining the construct validity of response type and response time measures. *Studies in Second Language Acquisition*, 44(1), 112–142. <https://doi.org/10.1017/S0272263121000164>

- Eibenstein, L. (2023). L3 acquisition of aspect: the influence of structural similarity, analytic L2 and general L3 proficiency. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1827–1858. <https://doi.org/10.1515/iral-2021-0220>
- Ekiert, M., & di Gennaro, K. (2021). Focused written corrective feedback and linguistic target mastery: Conceptual replication of Bitchener and Knoch (2010). *Language Teaching*, 54(1), 71–89. <http://doi.org/10.1017/S0261444819000120>
- Erlam, R., & Wei, L. (2024). The importance of increased processing demands in the design of Elicited Imitation tests. *Language Teaching Research*, 28(4), 1472–1500. <https://doi.org/10.1177/13621688211026032>
- Fouché, I. (2023). Measuring the impact of academic literacy interventions: Refining an evaluation design through self-reflection and feedback. *Applied Linguistics*, 44(4), 718–746. <https://doi.org/10.1093/applin/amad025>
- François, J., & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. *Language Learning & Technology*, 25(2), 26–41. <http://hdl.handle.net/10125/73429>
- Fukunaga, T. (2023). L2 writing development through two types of writing task repetition. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1109–1138. <https://doi.org/10.1515/iral-2021-0144>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Goriot, C., Hout, R., Broersma, M., Lobo, V., McQueen, J., & Unsworth, S. (2021). Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, 24(4), 546–568. <https://doi.org/10.1080/13670050.2018.1494131>
- Guo, J., & East, M. (2024). The Role of Vocabulary in Speaking: Voices from Learners and Raters. *Language Assessment Quarterly*, 21(4–5), 452–472. <https://doi.org/10.1080/15434303.2024.2440887>
- Gyllstad, H., McLean, S., & Stewart, J. (2021). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing*, 38(4), 558–579. <https://doi.org/10.1177/0265532220979562>
- Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11, Article 16. <https://doi.org/10.1186/s40468-021-00132-7>
- Ha, H. T., & Nguyen, D. T. B. (2023). Using structural equation modeling to examine the relationship between receptive vocabulary knowledge and receptive language proficiency: A new light on an old issue. *International Journal of Applied Linguistics*, 33(2), 382–398. <https://doi.org/10.1111/ijal.12475>
- Ha, H. T., Nguyen, D. T. B., & Stoeckel, T. (2024). What is the best predictor of word difficulty? A case of data mining using random forest. *Language Testing*, 41(4), 828–844. <https://doi.org/10.1177/02655322241263628>
- Hashimoto, B. J. (2021). Is frequency enough?: The frequency model in vocabulary size testing. *Language Assessment Quarterly*, 18(2), 171–187. <https://doi.org/10.1080/15434303.2020.1860058>
- Heeren, J., Speelman, D., & Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university

- students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458–1473.
<https://doi.org/10.1080/13670050.2019.1709411>
- Hicks, N. S. (2024). Exploring systematic orthographic crosslinguistic similarities to enhance foreign language vocabulary learning. *Language Teaching Research*, 28(6), 2385–2407. <https://doi.org/10.1177/13621688211047353>
- Holzknicht, F., Tornay, S., Battisti, A., Batty, A. O., Tissi, K., Haug, T., & Ebling, S. (2024). Automated Sign Language Vocabulary Assessment: Comparing Human and Machine Ratings and Studying Learner Perceptions. *Language Assessment Quarterly*, 21(3), 245–265. <https://doi.org/10.1080/15434303.2024.2364877>
- Hopp, H., & Thoma, D. (2021). Effects of plurilingual teaching on grammatical development in early foreign-language learning. *The Modern Language Journal*, 105(2), 464–483. <https://doi.org/10.1111/modl.12709>
- Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 305–315). Routledge.
<https://doi.org/10.4324/9781351034784-33>
- Isbell, D. R., Delaney, M., & Park, L. (2024). Developing measures of L2 Korean vocabulary knowledge. *Vocabulary Learning and Instruction*, 13(1), 1–10.
<https://doi.org/10.29140/vli.v13n1.1227>
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, Article 102521.
<https://doi.org/10.1016/j.system.2021.102521>
- Jiang, J., Bi, P., Xie, N., & Liu, H. (2023). Phraseological complexity and low- and intermediate-level L2 learners' writing quality. *International Review of Applied Linguistics in Language Teaching*, 61(3), 765–790.
<https://doi.org/10.1515/iral-2019-0147>
- Kahng, J., & Otonya, M. (2021). What elicited imitation can show us: Quantitative and qualitative analyses of longitudinal changes. *TESOL Quarterly*, 55(1), 284–295.
<https://doi.org/10.1002/tesq.3020>
- Kaltsa, M., Prentza, A., Prela, L., & Tsimpli, I. M. (2024). The interplay of lexical and grammatical development in Greek-Albanian bilingual children: evidence from the majority and the heritage language. *International Journal of Bilingual Education and Bilingualism*, 27(10), 1390–1406. <https://doi.org/10.1080/13670050.2024.2359408>
- Karyolemou, M. (2022). Teaching endangered languages of oral tradition: how and what to assess?. *Applied Linguistics*, 43(2), 389–411.
<https://doi.org/10.1093/applin/amab045>
- Kasahara, K., & Kanayama, K. (2021). When to conduct a vocabulary quiz, before the review or after the review? *System*, 103, Article 102641.
<https://doi.org/10.1016/j.system.2021.102641>
- Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, 14, Article 48. <https://doi.org/10.1186/s40468-024-00324-x>
- Kim, H. S., Lee, J. H., & Lee, H. (2024). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, 28(1), 7–28.
<https://doi.org/10.1177/1362168820981394>

- Kim, K. M., Chen, X., & Liu, X. (2024). Accuracy scoring of elicited imitation: A tutorial of automating speech data with commercial NLP support. *Research Methods in Applied Linguistics*, 3(3), Article 100127. <https://doi.org/10.1016/j.rmal.2024.100127>
- Kim, S. (2021). Generalizability of CEFR criterial grammatical features in a Korean EFL corpus across A1, A2, B1, and B2 levels. *Language Assessment Quarterly*, 18(3), 273–295. <https://doi.org/10.1080/15434303.2020.1855647>
- Kim, S., & Kessler, M. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, Article 100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170. <https://doi.org/10.1080/15434303.2020.1844205>
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781–812. <https://doi.org/10.1017/S0272263120000546>
- Lam, B. P., Sheng, L., & Zhang, X. (2023). Children's likelihood to perform adult-like in word association test: Effects of bilingualism and distributional properties of word relationships. *Bilingualism: Language and Cognition*, 26(1), 216–230. <https://doi.org/10.1017/S1366728922000463>
- Lau, W. M. E., Aubrey, S., & Ellis, R. (2024). Error logs in the second language classroom: Exploring the relationship between learner engagement with written corrective feedback and improvements in writing accuracy. *System*, 123, 103329. <https://doi.org/10.1016/j.system.2024.103329>
- Lecouvet, M. (2023). Non-canonical word order as a measure of syntactic complexity in advanced L2 German. *International Review of Applied Linguistics in Language Teaching*, 61(3), 877–907. <https://doi.org/10.1515/iral-2021-0029>
- Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, 97, Article 102461. <https://doi.org/10.1016/j.system.2021.102461>
- Lee, M., & Jung, J. (2024). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*, 28(4), 1552–1571. <https://doi.org/10.1177/13621688211034640>
- Lee, S., & Shin, S. Y. (2021). Towards improved assessment of L2 collocation knowledge. *Language Assessment Quarterly*, 18(4), 419–445. <https://doi.org/10.1080/15434303.2021.1908295>
- Leontjev, D., Huhta, A., & Tolvanen, A. (2023). L2 English vocabulary breadth and knowledge of derivational morphology: One or two constructs? *Language Testing*, 40(2), 300–324. <https://doi.org/10.1177/02655322221114015>
- Li, Y. K., Lin, S., Liu, Y., & Lu, X. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, 56, Article 100707. <https://doi.org/10.1016/j.asw.2023.100707>
- Li, Y., & Yang, R. (2023). Assessing the writing quality of English research articles based on absolute and relative measures of syntactic complexity. *Assessing Writing*, 55, Article

100692. <https://doi.org/10.1016/j.asw.2022.100692>
- Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*, 55(3), 1011–1023. <https://doi.org/10.1002/tesq.3055>
- Luft Baker, D., McCoach, B. D., Ware, S., Coyne, M. D., & Rattan, S. M. (2022). Effects of Spanish vocabulary knowledge on the English word knowledge and listening comprehension of bilingual students. *International Journal of Bilingual Education and Bilingualism*, 25(6), 2269–2283. <https://doi.org/10.1080/13670050.2021.1908219>
- Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, Article 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Mangold, M., Lenhard, W., Schindler, J., Schulz, D., & Richter, T. (2024). Assessing vocabulary of bilingual German-Turkish preschool children. *Bilingualism: Language and Cognition*, 27(4), 678–689. <https://doi.org/10.1017/S1366728923000822>
- Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *Modern Language Journal*, 107, 669–692. <https://doi.org/10.1111/modl.12866>
- Monster, I., Tellings, A., Burk, W. J., Keuning, J., Segers, E., & Verhoeven, L. (2021). Assessing children's incremental word knowledge in the upper primary grades. *Language Testing*, 38(4), 536–557. <https://doi.org/10.1177/0265532220961541>
- Monteiro, K., Crossley, S., Botarleanu, R.-M., & Dascălu, M. (2023). L2 and L1 semantic context indices as automated measures of lexical sophistication. *Language Testing*, 40(3), 576–606. <https://doi.org/10.1177/02655322221147924>
- Morea, N., Kasprowicz, R. E., Morrison, A., & Silvestri, C. (2024). Diverse population, homogenous ability: The development of a new receptive vocabulary size test for young language learners in England using Rasch analysis. *Research Methods in Applied Linguistics*, 3(3), 100166. <https://doi.org/10.1016/j.rmal.2024.100166>
- Muszyńska, K., Łuniewska, M., Dynak, A., Kolak, J., Lohrum, R., Otwinowska, A., ... Haman, E. (2024). Bilinguals' knowledge of 'home' and 'school' words revisited: evidence from Polish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 28(1), 73–91. <https://doi.org/10.1080/13670050.2024.2399639>
- Nakata, T., Tada, S., Mclean, S., & Kim, Y. A. (2021). Effects of distributed retrieval practice over a semester: Cumulative tests as a way to facilitate second language vocabulary learning. *TESOL Quarterly*, 55(1), 248–270. <https://doi.org/10.1002/tesq.596>
- Nguyen, B. T. T., & Nguyen, L. Q. (2024). Measuring productive derivative knowledge of vietnamese EFL learners: The role of headword scoring, vocabulary breadth, and headword familiarity. *System*, 123, 103322. <https://doi.org/10.1016/j.system.2024.103322>
- Nicklin, C., & Vitta, J. P. (2022). Assessing Rasch measurement estimation methods across R packages with yes/no vocabulary test data. *Language Testing*, 39(4), 513–540. <https://doi.org/10.1177/02655322211066822>
- O'Reilly, D., & Marsden, E. (2023). Elicited metaphoric competence in a second language: a construct associated with vocabulary knowledge and general proficiency?. *International Review of Applied Linguistics in Language Teaching*, 61(2), 287–327. <https://doi.org/10.1515/iral-2020-0054>
- O'Shannessy, C., Davis, V., Bartlett, J., Nelson, A., & Foster, D. (2023). Developing the Little Kids' Word List app, a fair assessment tool of communicative development for

- young Aboriginal children in multilingual families in Central Australia. *Studies in Language Assessment*, 12(2), 286-320.
- Owen, N., Shrestha, P., & Bax, S. (2021). *Researching lexical thresholds and lexical profiles across the Common European Framework of References for Languages (CEFR) levels assessed in the Aptis test* (Report No. AR-G/2021/1). British Council. https://www.britishcouncil.org/sites/default/files/owen_shrestha_and_bax_layout.pdf
- Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly*, 39, 372–425. <https://doi.org/10.32038/ltrq.2024.39.23>
- Prichard, C., & Atkins, A. (2021). Evaluating the vocabulary coping strategies of L2 readers: An eye tracking study. *TESOL Quarterly*, 55(2), 593–620. <https://doi.org/10.1002/tesq.3005>
- Purpura, J. E., & Oh, S. (2024). Assessing the linguistic resources of meaningful communication. In A. J. Kunnan (Ed.), *Companion to Language Assessment* (pp. 100-124). Oxford, UK: John Wiley & Sons.
- Qian, L. (2023). Use of lexical features in high-stakes tests: Evidence from the perspectives of complexity, accuracy and fluency. *Assessing Writing*, 57, Article 100758. <https://doi.org/10.1016/j.asw.2023.100758>
- Rahmati, T., Kazemian, M., Khonamri, F., & Daryaback, A.R. (2023). A chaos complexity perspective on EFL learners' sensitivity dependence on initial conditions in vocabulary tests, *The Journal of Asia TEFL*, 20(4), 895-903. <http://dx.doi.org/10.18823/asiatefl.2023.20.4.9.895>
- Rahmati, T., Kazemian, M., Khonamri, F., & Daryabak, A. (2023). A chaos complexity perspective on EFL learners' sensitivity dependence on initial conditions in vocabulary tests. *The Journal of Asia TEFL*, 20(4), 895-903. <http://dx.doi.org/10.18823/asiatefl.2023.20.4.9.895>
- Ramezanali, N., Uchiyara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, 55(1), 105–133. <https://doi.org/10.1002/tesq.579>
- Randall, T. S., & Urbanski, K. (2023). Development of a Computerized Dynamic Assessment Program for Second Language Grammar Instruction and Assessment. *Language and Sociocultural Theory*, 10(1), 50-81. <https://doi.org/10.1558/lst.21006>
- Read, J. (2023). Towards a new sophistication in vocabulary assessment. *Language Testing*, 40(1), 40-46. <https://doi.org/10.1177/02655322221125698>
- Révész, A., Bunting, L., Florea, A., Gilabert, R., Hård af Segerstad, Y., Mihu, I. P., & Vasalou, A. (2021). The effects of multiple-exposure textual enhancement on child L2 learners' development in derivational morphology: A multi-site study. *TESOL Quarterly*, 55(3), 901–930. <https://doi.org/10.1002/tesq.3036>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2021). Linguistic complexity in teachers' assessment of German essays in high stakes testing. *Assessing Writing*, 50, Article 100561. <https://doi.org/10.1016/j.asw.2021.100561>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2024). The interplay of task characteristics, linguistic complexity, and language proficiency in High-Stakes English as a Foreign language writing. *TESOL Quarterly*, 58(2), 775–801. <https://doi.org/10.1002/tesq.3254>
- Robles-García, P., McLean, S., Stewart, J., Shin, J. young, & Sánchez-Gutiérrez, C. H.

- (2024). The Development and Initial Validation of O-WSVLT, a Meaning-Recall Online L2 Spanish Vocabulary Levels Test. *Language Assessment Quarterly*, 21(2), 181–205. <https://doi.org/10.1080/15434303.2024.2311724>
- Sarac, M., & Loken, E. (2023). Examining patterns of omitted responses in a large-scale English language proficiency test. *International Journal of Testing*, 23(1), 56–72. <https://doi.org/10.1080/15305058.2022.2070756>
- Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency, and task effects in Spanish Dual Language Immersion. *Language Learning & Technology*, 25(1), 53–72. <https://hdl.handle.net/10125/73425>
- Shafipoor, M., Ravand, H., & Maftoon, P. (2021). Test-level and item-level model fit comparison of general vs. specific diagnostic classification models: A case of true DCM. *Language Testing in Asia*, 11, Article 33. <https://doi.org/10.1186/s40468-021-00148-z>
- Shamshiri, F., Esfahani, F. R., & Hosseini, S. E. (2023). Models of assessment in the classroom: A comparative research of CALL-based vs. traditional assessment on vocabulary learning among Iranian EFL learners. *Language Testing in Asia*, 13(1), Article 43. <https://doi.org/10.1186/s40468-023-00259-9>
- Sok, S., & Shin, H. W. (2023). Investigating the dimensions and determinants of children's narrative writing in Korean. *Assessing Writing*, 57, Article 100740. <https://doi.org/10.1016/j.asw.2023.100740>
- Stewart, J., Gyllstad, H., Nicklin, C., & McLean, S. (2024). Establishing meaning recall and meaning recognition vocabulary knowledge as distinct psychometric constructs in relation to reading proficiency. *Language Testing*, 41(1), 89–108. <https://doi.org/10.1177/02655322231162853>
- Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, 43(2), 462–471. <https://doi.org/10.1017/S0272263121000437>
- Stewart, J., Vitta, J. P., Nicklin, C., McLean, S., Pinchbeck, G. G., & Kramer, B. (2022). The relationship between word difficulty and frequency: a response to Hashimoto (2021). *Language Assessment Quarterly*, 19(1), 90–101. <https://doi.org/10.1080/15434303.2021.1992629>
- Stoeckel, T., Ha, H. T., Nguyen, D. T. B., & Nicklin, C. (2024). Can an extended-matching second-language vocabulary test format bridge the gap between meaning-recognition and meaning-recall?. *Research Methods in Applied Linguistics*, 3(2), Article 100109. <https://doi.org/10.1016/j.rmal.2024.100109>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>
- Sung, H., Cho, S., & Kyle, K. (2024). An Empirical Evaluation of Lexical Diversity Indices in L2 Korean Writing Assessment. *Language Assessment Quarterly*, 21(2), 159–180. <https://doi.org/10.1080/15434303.2024.2311728>
- Szabo, C., Stickler, U., & Adinolfi, L. (2021). Predicting the academic achievement of multilingual students of English through vocabulary testing. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1531–1542. <https://doi.org/10.1080/13670050.2020.1814196>

- Tan, L. Y., McLean, S., Kim, Y. A., & Vitta, J. P. (2024). Rasch modelling vs. item facility: Implications on the validity of assessments of Asian EFL/ESL vocabulary knowledge and lexical sophistication modelling. *Language Testing in Asia*, 14, Article 55. <https://doi.org/10.1186/s40468-024-00327-8>
- Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in second language Chinese. *The Modern Language Journal*, 105(3), 740–759. <https://doi.org/10.1111/modl.12725>
- Uzun, K. (2021). Performance prediction strengths of noun and verb phrases in L2 writing: Comparison of density and complexity variables. *Assessing Writing*, 50, Article 100572. <https://doi.org/10.1016/j.asw.2021.100572>
- Wang, M., Shen, H., Boland, J. E., & Wei, H. (2021). Visual context modulates L2 long-term structural priming for the Chinese ba construction. *The Modern Language Journal*, 105(2), 526–551. <https://doi.org/10.1111/modl.12715>
- Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programmes. *Journal of English for Academic Purposes*, 61, Article 101203. <https://doi.org/10.1016/j.jeap.2022.101203>
- Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/S0272263121000449>
- Weideman, A. & Van Dyk, T. (2023). Achieving technical economy: A modification of cloze procedure. *Language Teaching Research Quarterly*, 37, 144–160. <https://doi.org/10.32038/ltrq.2023.37.07>
- Weiler, T. (2021). *Testing lexicogrammar: An investigation into the construct tested in the Language in Use section of the Austrian Matura in English*. Peter Lang.
- Weisleder, A., Friend, M., Tsui, A. S. M., & Marchman, V. A. (2024). Using parent report to measure vocabulary in young bilingual children: A scoping review. *Language Learning*, 74(2), 468–505. <https://doi.org/10.1111/lang.12617>
- Wilsenach, C., & Schaefer, M. (2022). Development and initial validation of productive vocabulary tests for isiZulu, Siswati and English in South Africa. *Language Testing*, 39(4), 567–592. <https://doi.org/10.1177/02655322211063785>
- Wood, C., Hoge, R., Schatschneider, C., & Castilla-Earls, A. (2021). Predictors of item accuracy on the Test de Vocabulario en Imágenes Peabody for Spanish-English speaking children in the United States. *International Journal of Bilingual Education and Bilingualism*, 24(8), 1178–1192. <https://doi.org/10.1080/13670050.2018.1547266>
- Woods, K., Hashimoto, B., & Brown, E. K. (2023). A multi-measure approach for lexical diversity in writing assessments: Considerations in measurement and timing. *Assessing Writing*, 55, Article 100688. <https://doi.org/10.1016/j.asw.2022.100688>
- Wu, Y.-j. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology*, 25(2), 192–214. <http://hdl.handle.net/10125/73440>
- Xia, T., Chen, X., Parsaei, H. R., & Qiu, F. (2023). An intelligent vocabulary size measurement method for second language learner. *Language Testing in Asia*, 13, Article 45 (2023). <https://doi.org/10.1186/s40468-023-00258-w>
- Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2023). A multidimensional approach to assessing the effects of task complexity on L2 students' argumentative writing. *Assessing Writing*, 55, Article 100690. <https://doi.org/10.1016/j.asw.2022.100690>
- Yang, Y., Yap, N. T., & Ali, M. A. (2023). Predicting EFL expository writing quality with

- measures of lexical richness. *Assessing Writing*, 57, Article 100762.
<https://doi.org/10.1016/j.asw.2023.100762>
- Zhang, X., Zhao, B., & Li, W. (2023). N-gram use in EFL learners' retelling and monologic tasks. *International Review of Applied Linguistics in Language Teaching*, 61(3), 939–965. <https://doi.org/10.1515/iral-2021-0080>
- Zhang, Y., & Papi, M. (2024). Regulatory focus predicts individual differences in pragmatic versus grammatical awareness and sensitivity. *Studies in Second Language Learning and Teaching*, 14(2), 267-290. <https://doi.org/10.14746/ssllt.35073>
- Zhang, Y., & Wu, W. (2021). How effective are lexical richness measures for differentiations of vocabulary proficiency? A comprehensive examination with clustering analysis. *Language Testing in Asia*, 11, Article 15.
<https://doi.org/10.1186/s40468-021-00133-6>

Assessing Interactional Competence and Pragmatics

- Ahmadi, A. (2024). Patterns of interaction in a paired speaking test: Comparing L1 and L2 interactions. *Language Testing in Asia*, 14, Article 31.
<https://doi.org/10.1186/s40468-024-00299-9>
- Antoniou, K., & Milaki, E. (2021). Irony comprehension in bidialectal speakers. *The Modern Language Journal*, 105(3), 697–719. <https://doi.org/10.1111/modl.12724>
- Barkaoui, K. (2023). Exploring task effects on register variation in second language learners' writing. *TASK: Journal on Task-Based Language Teaching and Learning*, 3, 109-139. <https://doi.org/10.1075/task.22009.bar>
- Burton, J. D. (2021). The face of communication breakdown: Multimodal repair in L2 oral proficiency interviews. *Papers in Language Testing and Assessment*, 10(2), 30–61.
http://www.altanz.org/uploads/5/9/0/8/5908292/3._plta_10_2__burton.pdf
- Canals, E. (2023). Modified output and metalanguage during conversational interaction: A qualitative look at interactional feedback. *Studies in Second Language Learning and Teaching*, 13(3), 601-626. <https://doi.org/10.14746/ssllt.31128>
- Carrió-Pastor, M. L. (2021). The assessment of metadiscourse devices in English as a foreign language. *Assessing Writing*, 50, Article 100560.
<https://doi.org/10.1016/j.asw.2021.100560>
- Chan, J. Y. H. (2021). Bridging the gap between ELF and L2 learners' use of communication strategies: Rethinking current L2 assessment and teaching practices. *System*, 101, Article 102609.
<https://doi.org/10.1016/j.system.2021.102609>
- Chukharev-Hudilainen, E., & Ockey, G. J. (2021). The development and evaluation of interactional competence elicitor for oral language assessments. *ETS Research Report Series*, 2021(1), 1–20. <https://doi.org/10.1002/ets2.12319>
- Dai, D. W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test*. Peter Lang. <https://doi.org/10.3726/b21295>
- De John, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9(1), 541-560.
<https://doi.org/10.1146/annurev-linguistics-030521-052114>
- Ellis, R., Roever, C., Shintani, N., & Zhu, Y. (2024). *Measuring second language pragmatic competence: A psycholinguistic perspective*. Multilingual Matters.
<https://doi.org/10.2307/jj.22679670>

- Galaczi, E. D., & Taylor, L. (2021). Measuring interactional competence. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 338–348). Routledge.
<https://doi.org/10.4324/9781351034784-36>
- Gallardo del Puerto, F., & Basterrechea, M. (2024). The role of proficiency and pair formation method in language-related episodes: A study of young CLIL learners' interaction. *Language Teaching Research*, 28(3), 1087–1105.
<https://doi.org/10.1177/13621688211024363>
- Garcia-Ponce, E. E., & Tavakoli, P. (2022). Effects of task type and language proficiency on dialogic performance and task engagement. *System*, 105, Article 102734.
<https://doi.org/10.1016/j.system.2022.102734>
- Gesuato, S., Pagliarini, E., & Sanfelici, E. (2023). YOUNG ADULTS' INFERENTIAL READING SKILLS - RECOGNISING IMPLICIT AND UNSTATED CONTENT IN ITALIAN L1 AND ENGLISH L2. *Italiano LinguaDue*, 15(2), 20–20.
<https://doi.org/10.54103/2037-3597/22011>
- Gokturk, N., & Chukharev, E. (2024). Exploring the Potential of a Spoken Dialog System-Delivered Paired Discussion Task for Assessing Interactional Competence. *Language Assessment Quarterly*, 21(1), 60–99.
<https://doi.org/10.1080/15434303.2023.2289173>
- Guo, Y., Xu, J., & Chen, C. (2023). Measurement of engagement in the foreign language classroom and its effect on language achievement: the case of Chinese college EFL students. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1225–1270. <https://doi.org/10.1515/iral-2021-0118>
- Harding, L., Macqueen, S., & Pill, J. (2023). Assessing communicative competence. In M. Solon & M. Kanwit (Eds.), *Communicative competence in a second language: Theory, method, and applications* (pp. 187–207). Routledge.
<http://doi.org/10.4324/9781003160779-14>
- Jung, J., & Fu, X. (2023). The impact of pragmalinguistic support on video-conferenced collaborative suggestion-giving task. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1083–1107. <https://doi.org/10.1515/iral-2021-0138>
- Kang, O., Kermad, A., & Taguchi, N. (2021). The interplay of proficiency and study abroad experience on the prosody of L2 speech acts. *Journal of Second Language Pronunciation*, 7(3), 343–369. <https://doi.org/10.1075/jslp.20024.kan>
- Kazemian, M., Khodareza, M.R., Khonamri, F., & Rahimy, R. (2022) Problematizing language testing and assessment syllabi through intercultural competence assessment perspective in an EFL context. *The Journal of Asia TEFL*, 19(1), 309–316. <http://dx.doi.org/10.18823/asiatefl.2022.19.1.23.309>
- Kazemian, M., Khodareza, M.R., Khonamri, F., & Rahimy, R. (2023). ELT scholars' attitudes towards inclusion of intercultural competence assessment in language proficiency tests. *TESL-EJ Journal*, 26(4), 1–21.
<https://tesl-ej.org/wordpress/issues/volume26/ej104/ej104a6/>
- Lee, H.-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring interaction in video-call paired speaking tests: A look at scores, language, and perceptions. *Applied Linguistics*, 45(3), 533–552. <https://doi.org/10.1093/applin/amad046>
- Li, Y. K., Lin, S., Liu, Y., & Lu, X. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, 56, Article 100707.

- <https://doi.org/10.1016/j.asw.2023.100707>
- Lv, Z. (2024). Prompt effects on assessment of interactional competence in a group oral test. *International Journal of Applied Linguistics*, 34(3), 1089-1107. <https://doi.org/10.1111/ijal.12554>
- McCloy R. A., Everson H. T., & Purpura, J. E. (2023). Next Steps to Support CAT ECL 2.0: Report from a Workshop. HumRRO Technical Report 166. Alexandria, VA: Human Development Resources Research Organization.
- Montasseri, Z., & Ahmadi, A. (2024). Validity argument of an Interactional Competence checklist: Perspectives and many-facets Rasch analysis. *Studies in Language Assessment (SILA)*, 13(1), 69-107. <https://doi.org/10.58379/MEZB3039>
- Mootoosamy, K., & Aryadoust, V. (2024). Transitioning from communicative competence to multimodal and intercultural competencies: A systematic review. *Societies*, 14(115). <https://doi.org/10.3390/soc14070115>
- Nakatsuhara, F., May, L., Inoue, C., Willcox-Ficzere, E., Westbrook, C., & Spiby, R. (2021). *Exploring the potential for assessing interactional and pragmatic competence in semi-direct speaking tests* (Report No. VS/2021/001). British Council. https://www.britishcouncil.org/sites/default/files/nakatsuhara_et_al_2021_-_assessing_interactional_and_pragmatic_competence_in_semi-direct_speaking_tests_v2.pdf
- Nicholas, A., & Blake, J. (2024). Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing. *Research Methods in Applied Linguistics*, 3(3), Article 100164. <https://doi.org/10.1016/j.rmal.2024.100164>
- Ockey, G. J., Chukharev-Hudilainen, E., & Hirsch, R. R. (2023). Assessing interactional competence: ICE versus a human partner. *Language Assessment Quarterly*, 20(4-5), 377-398. <https://doi.org/10.1080/15434303.2023.2237486>
- Qin, L., & Nian, Y. (2024). Unpacking Chinese lower-proficiency EFL learners' conceptual development of English subjunctives with group dynamic assessment: Efficacy, sustainability and reciprocity. *System*, 121, 103248. <https://doi.org/10.1016/j.system.2024.103248>
- Qin, T., & van Compernelle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, 25(2), 55-74. <http://hdl.handle.net/10125/73433>
- Roevers, C., & Ikeda, N. (2022). What scores from monologic speaking tests can(not) tell us about interactional competence. *Language Testing*, 39(1), 7-29. <https://doi.org/10.1177/02655322211003332>
- Roevers, C., & Ikeda, N. (2024). The relationship between L2 interactional competence and proficiency. *Applied Linguistics*, 45(4), 676-698. <https://doi.org/10.1093/applin/amad053>
- Su, H., & Lu, X. (2022). Assessing pragmatic performance in advanced L2 academic writing through the lens of local grammars: A case study of "exemplification." *Assessing Writing*, 54, Article 100668. <https://doi.org/10.1016/j.asw.2022.100668>
- Su, Y., & Chen, X. (2024). Examining refusal and acceptance sequences employing a data-driven binary rating approach. *Research Methods in Applied Linguistics*, 3(3), Article 100161. <https://doi.org/10.1016/j.rmal.2024.100161>
- Su, Y., & Shin, S.-Y. (2024). Comparing two formats of data-driven rating scales for classroom assessment of pragmatic performance with roleplays. *Language Testing*, 41(2), 357-383. <https://doi.org/10.1177/02655322231210217>

- Syquia, J., & Leeming, P. (2024). Assessing interactional competence through group discussion: A mixed methods validation. *Research Methods in Applied Linguistics*, 3(3), Article 100144. <https://doi.org/10.1016/j.rmal.2024.100144>
- Taguchi, N. (2023). Using immersive virtual reality for the assessment of intercultural conflict mediation. *Language Learning & Technology*, 27(3), 108–128. <https://hdl.handle.net/10125/73534>
- Tai, H., Liu, C., & Chen, Y. (2023). Validation of a multiple-choice implicature test: insights from Chinese EFL learners' cognitive processes. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1481-1506. <https://doi.org/10.1515/iral-2021-0190>
- Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in second language Chinese. *The Modern Language Journal*, 105(3), 740–759. <https://doi.org/10.1111/modl.12725>
- Tavakoli, P. & Cooke, S. (2024). *Comprehensibility in Language Assessment: A Broader Perspective*. University of Toronto Press.
- Timpe-Laughlin, V., & Youn, S. J. (2021). Measuring L2 pragmatics. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.254–264). Routledge. <https://doi.org/10.4324/9781351034784-28>
- Trofimovich, P., Tekin, O., & McDonough, K. (2021). Task engagement and comprehensibility in interaction: Moving from what second language speakers say to what they do. *Journal of Second Language Pronunciation*, 7(3), 435–461. <https://doi.org/10.1075/jslp.21006.tro>
- Uskokovic, B., & Schirm, S. (2024). Assessment of Interactional Competence in L2 German: Integrating an innovative rubric to help language teachers effectively evaluate pragmatics. *System*, 125, 103400. <https://doi.org/10.1016/j.system.2024.103400>
- Uyar, G., & Balaman, U. (2024). The Role of Gazing Behaviors in Navigating Paired Role-Play Interactional Competence Assessment Tasks. *Language Assessment Quarterly*, 21(3), 266–286. <https://doi.org/10.1080/15434303.2024.2375651>
- Vössing, L., & Kern, F. (2023). The co-construction of pragmatic competencies in different settings: The case of two children with autism spectrum disorder. *Journal of Interactional Research in Communication Disorders*, 14(2), 195-219. <https://doi.org/10.1558/jircd.24423>
- Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing*, 50, Article 100571. <https://doi.org/10.1016/j.asw.2021.100571>
- Wang, Y., & Ren, W. (2024). L2 grit and pragmatic comprehension among Chinese postgraduate learners: Enjoyment and anxiety as mediators. *International Journal of Applied Linguistics*, 34(1), 224-241. <https://doi.org/10.1111/ijal.12489>
- Weideman, A. (2022). Context, construct, and validation: A perspective from South Africa. *Language Assessment Quarterly*, 19(2), 124-141. <https://doi.org/10.1080/15434303.2020.1860991>
- Weideman, A., Read, J. & Du Plessis, T. (Eds.). (2021). *Assessing academic literacy in a multilingual society: Transition and transformation* (Volume 84). Multilingual Matters. <https://doi.org/10.21832/WEIDEM6201>.

- Williams, K. (2024). Exploring English for academic purposes instructors' perceptions of speech fluency through developing and piloting a rating scale for a paired conversational task. *System*, 121, 103266. <https://doi.org/10.1016/j.system.2024.103266>
- Zhang, L., & Jin, Y. (2021). Assessing interactional competence in the computer-based CET-SET: An investigation of the use of communication strategies. *Assessment in Education: Principles, Policy & Practice*, 28(4), 389–410. <https://doi.org/10.1080/0969594X.2021.1976107>
- Zhang, S., & Isaacs, T. (2023). Can interactions happen across the screens? The use of videoconferencing technology in assessing second language pragmatic competence. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts* (pp. 196–211). Routledge.
- Zhang, Y., & Papi, M. (2024). Regulatory focus predicts individual differences in pragmatic versus grammatical awareness and sensitivity. *Studies in Second Language Learning and Teaching*, 14(2), 267-290. <https://doi.org/10.14746/ssllt.35073>

Assessing Mediation, Translation, and Interpretation

- Badham, L., & Furlong, A. (2023). Summative assessments in a multilingual context: What comparative judgment reveals about comparability across different languages in Literature. *International Journal of Testing*, 23:2, 111-134. <https://doi.org/10.1080/15305058.2022.2149536>
- Cheng, L., & Im, G.-H. (2021). Translation of surveys for international research: Beyond equivalence. In G. Harbaugh & U. Luhanga (Eds.). *Basic elements of survey research in education: Addressing the problems your advisor never told you about* (pp. 219–236). Information Age Publishing.
- Chovancová, B., Bilová, Š., & Hradilová, A. (2023). Turning the tables on online exam cheating via language mediation tasks. *Language Learning in Higher Education*, 13(2), 517-533. <https://doi.org/10.1515/cercles-2023-2033>
- Colman, A., Segers, W., & Verplaetse, H. (2023). Extending Preselected Items Evaluation (PIE) with the ATA (American Translators Association) error categories for objective and targeted translation evaluation. *Journal of Applied Linguistics and Professional Practice*, 17(3), 274-293. <https://doi.org/10.1558/jalpp.21681>
- Han, C. (2021). Analytic rubric scoring versus comparative judgment: A comparison of two approaches to assessing spoken-language interpreting. *Meta: Translators' Journal*, 66(2), 337-361. <https://doi.org/10.7202/1083182ar>
- Han, C. (2021). Assessing spoken-language interpretation: The method of comparative judgment. *Interpreting*, 24(1), 59-83. <https://doi.org/10.1075/intp.00068.han>
- Han, C. (2022). Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, 39(1), 30–55. <https://doi.org/10.1177/02655322211036100>
- Han, C. (2023). Interrogating the predictive validity of aptitude testing for interpreting: A systematic methodological review. *The Interpreter and Translator Trainer*, 17 (1), 7–28. <https://doi.org/10.1080/1750399X.2023.2170049>
- Han, C., & Lu, X.-L. (2021). Interpreting quality assessment re-imagined: The synergy between human and machine scoring. *Interpreting and Society*, 1(1), 70-90. <https://doi.org/10.1177/27523810211033670>

- Han, C., & Lu, X.-L. (2023). Can automated machine translation evaluation metrics be used to assess students' interpretation in the language learning classroom? *Computer Assisted Language Learning*, 36(5-6), 1064–1087. <https://doi.org/10.1080/09588221.2021.1968915>
- Han, C., & Xiao, X. (2022). A comparative judgment approach to assessing Chinese Sign Language interpreting. *Language Testing*, 39(2), 289–312. <https://doi.org/10.1177/02655322211038977>
- Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment and Evaluation in Higher Education*, 46(8), 1299–1313. <https://doi.org/10.1080/02602938.2020.1855624>
- Han, C., Hu, J., & Deng, Y. (2023). Effects of language background and directionality on raters' assessments of spoken-language interpretation. *Revista Española de Lingüística Aplicada*, 36(2), 556–584. <https://doi.org/10.1075/resla.21009.han>
- Han, C., Xiao, R., & Su, W. (2021). Assessing the fidelity of consecutive interpretation: The effects of using source versus target text as the reference material. *Interpreting*, 23(2), 245–268. <https://doi.org/10.1075/intp.00058.han>
- Han, C., Zheng, B.-H., Xie, M.-Q., & Chen, S.-R. (2024). Raters' scoring process in assessment of interpreting: an empirical study based on eye tracking and retrospective verbalisation. *Interpreter and Translator Trainer*, 18(3), 400–422. <https://doi.org/10.1080/1750399X.2024.2326400>
- Khabbazzbashi, N., Chan, S. and Clark, T. (2023). Towards the new construct of academic English in the digital age, *ELT Journal*, 77(2), 207–216 <https://doi.org/10.1093/elt/ccac010>
- Liang, Y., Li, Y., & Sang, Z. (2023). A study on peer mediation in dynamic assessment of translation revision competence. *Language Assessment Quarterly*, 20(1), 108–126. <https://doi.org/10.1080/15434303.2022.2153050>
- Liu, Y., Cheong, C. M., & Zhang, J. (2024). The Role of Achievement Goals and Strategy Use in Secondary School students' Integrated Writing Performance. *Language Assessment Quarterly*, 21(4–5), 324–349. <https://doi.org/10.1080/15434303.2024.2435818>
- Lu, X.-L. & Han, C. (2023). Automatic assessment of spoken-language interpreting based on machine translation evaluation metrics: A multi-scenario exploratory study. *Interpreting: International Journal of Research and Practice in Interpreting*, 25(1), 109–143. <https://doi.org/10.1075/intp.00076.lu>
- Moradkhan, D., & Shahmirzadi, N. (2012). Utilizing measures of (CAF) complexity, accuracy and fluency for interpreter evaluation. *Translation Studies (Motale'at-e-Tarjome Quarterly)*, 10(39), 87–104, (ISSN: 1735–0212).
- Tavakoli, P. & Cooke, S. (2024). *Comprehensibility in Language Assessment: A Broader Perspective*. University of Toronto Press.
- Van Dyk, T., Murre, P., & Kotzé, H. (2021). Does one size fit all? Some considerations for test translation. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 52–74). Multilingual Matters.
- Wang, Z., & Zhang, J. (2023). Mediation and learner reciprocity: Applying dynamic assessment to the oral proficiency interview framework. *Language and Sociocultural Theory*, 10(1), 82–105. <https://doi.org/10.1558/lst.22181>

Yang, Z., & Yu, C.* (2024). Is the use of multiple-choice items and a holistically-scored paragraph translation task fair? Examining a large-scale translation subtest. *Asia Pacific Education Review*. Online First. <https://doi.org/10.1007/s12564-024-09993-y>

Assessing Young Learners

- Allal, L. (2021). Involving primary school students in the co-construction of formative assessment in support of writing. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 584–601. <https://doi.org/10.1080/0969594X.2021.1951164>
- Angelo, D., & Hudson, C. (2023). Phantom young English language learners: The shadowy presence of second language proficiency in (English medium) early childhood assessment. *Studies in Language Assessment*, 12(2), 1–27. <https://doi.org/10.58379/KCDT5699>
- Ardasheva, Y., Crosson, A. C., Carbonneau, K. J., & French, B. F. (2021). Unpacking contributions of morphosyntactic awareness and vocabulary to science reading comprehension among linguistically diverse students. *TESOL Quarterly*, 55(3), 931–965. <https://doi.org/10.1002/tesq.3039>
- Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners' reading comprehension: Exploring the effects of working memory, grade level, and reading task. *Language Testing*, 38(3), 356–377. <https://doi.org/10.1177/0265532221991480>
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455. <https://doi.org/10.1177/0265532221992274>
- Buzick, H. M., Wolf, M. K., & Ballard, L. (2024). Achievement and growth on English language proficiency and content assessments for English learners in elementary grades. *Educational Measurement: Issues and Practice*, 43(1), 83–95. <http://doi.org/10.1111/emip.12588>
- Chapman, M., Kemp, J. A., Kim, A. A., MacGregor, D., & MacMillan, F. Testing young multilingual learners remotely: Prioritizing equity, fairness, and social justice. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 153–170). Routledge. <https://doi.org/10.4324/9781003292395-13>
- Chen Hsieh, J. (2021). Digital storytelling outcomes and emotional experience among middle school EFL Learners: Robot-assisted versus PowerPoint-assisted mode. *TESOL Quarterly*, 55(3), 994–1010. <https://doi.org/10.1002/tesq.3043>
- Choi, J. S., & Loewen, S. (2022). Exploring young learners' strategic behaviors in a speaking test. *TESOL Quarterly*, 56(4), 1384–1396. <https://doi.org/10.1002/tesq.313>
- Deng, Z., Uccelli, P., & Snow, C. (2022). Diversity of Advanced Sentence Structures (DASS) in writing predicts argumentative writing quality and receptive academic language skills of fifth-to-eighth grade students. *Assessing Writing*, 53, Article 100649. <https://doi.org/10.1016/j.asw.2022.100649>
- Devine, A., Lopes, S., & Harrison, G. (2021). Young learners' positive attitudes to learning in South America. *Research Notes*, 80, 21–28. Cambridge Assessment English.
- Ding, W., & Yu, G. (2023). Using causal explanation speaking tasks to assess young EFL learners' speaking ability: The effects of age, cognitive, and L2 linguistic development. *Language Assessment Quarterly*, 20(3), 251–276.

- <https://doi.org/10.1080/15434303.2023.2206811>
- Durrant, P., & Brenchley, M. (2021). The development of academic collocations in children's writing. In P. Szudarski & S. Barclay (Eds.), *Vocabulary theory, patterning and teaching* (pp. 99–120). Multilingual Matters.
- Durrant, P., & Durrant, A. (2022). Appropriateness as an aspect of lexical richness: What do quantitative measures tell us about children's writing? *Assessing Writing*, 51, Article 100596. <https://doi.org/10.1016/j.asw.2021.100596>
- Fladung, I. T., Gruhn, S., Österbauer, V., & Jost, J. (2023). Assessing writing in fourth grade: Rhetorical specification effects on text quality. *Assessing Writing*, 57, Article 100764. <https://doi.org/10.1016/j.asw.2023.100764>
- Gallardo del Puerto, F., & Basterrechea, M. (2024). The role of proficiency and pair formation method in language-related episodes: A study of young CLIL learners' interaction. *Language Teaching Research*, 28(3), 1087-1105. <https://doi.org/10.1177/13621688211024363>
- Getman, E. P. (2024). Four task characteristics and their influence on speaking performance and engagement. *English Teaching & Learning*, 48(2), 265–290. <https://doi.org/10.1007/s42321-024-00176-x>
- Getman, E. P. (2024). Speaking test tasks for young language learners: Development and validation. In L. Davis & J. M. Norris (Eds.), *Challenges and Innovations in Speaking Assessment*. Routledge. <https://doi.org/10.4324/9781315201504-7>
- Goriot, C., Hout, R., Broersma, M., Lobo, V., McQueen, J., & Unsworth, S. (2021). Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, 24(4), 546–568. <https://doi.org/10.1080/13670050.2018.1494131>
- Graham, K. M. (2024). Examining the criterion validity of curriculum-based measurement of writing (CBM-W) with young EFL learners. *Research Methods in Applied Linguistics*, 3(2), Article 100108. <https://doi.org/10.1016/j.rmal.2024.100108>
- Grapin, S. E., & Llosa, L. (2022). Dynamic assessment of English learners in the content areas: An exploratory study in fifth-grade science. *TESOL Quarterly*, 56(1), 201-229. <https://doi.org/10.1002/tesq.3059>
- Guo, J. & Cheng, L. (2024). Chinese Parent Involvement in Young Learners' English Tests: A Systematic Review of the Literature. *語言測試和評估 (Language Testing and Assessment)*, 4, 1-20.
- Hannah, L., Kim, H., & Jang, E. E. (2022). Investigating the effects of task type and linguistic background on accuracy in automated speech recognition systems: Implications for use in language assessment of young learners. *Language Assessment Quarterly*, 19(3), 289–313. <https://doi.org/10.1080/15434303.2022.2038172>
- Hsieh, C.-N. (2024). Building a validity argument for the TOEFL Junior® tests. ETS Research Report Series, 2024(1), 1–22. <https://doi.org/10.1002/ets2.12379>
- Hu, X., Zhang, X., & McGeown, S. (2024). Foreign language anxiety and achievement: A study of primary school students learning English in China. *Language Teaching Research*, 28(4), 1594-1615. <https://doi.org/10.1177/13621688211032332>
- Huang, B. H., Bailey, A. L., Sass, D. A., & Shawn Chang, Y. (2021). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401–428. <https://doi.org/10.1177/0265532220925731>
- Huang, Y., Wu, J., Li, M., & Yang, Y. (2022). The trajectory of syntactic complexity

- development in L1 Chinese narrative writings of primary school children: A systematic 5-year longitudinal study. *Assessing Writing*, 52, Article 100622. <https://doi.org/10.1016/j.asw.2022.100622>
- Hunte, M. R., McCormick, S., Shah, M., Lau, C., & Jang, E. E. (2021). Investigating the potential of NLP-driven linguistic and acoustic features for predicting human scores of children's oral language proficiency. *Assessment in Education: Principles, Policy & Practice*, 28(4), 477–505. <https://doi.org/10.1080/0969594X.2021.1999209>
- Ignacio, A., Jang, E. E., Heywood, A., Hunte, M. R., & Lai, H. (2024). Fostering young students' writing skill development and self-regulation through Dynamic Assessment feedback during pandemic disruptions. *Studies in Language Assessment (SILA)*, 13(2), 147–186. <https://doi.org/10.58379/asqh2857>
- Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, 55, Article 100676. <https://doi.org/10.1016/j.asw.2022.100676>
- Kahn-Horwitz, J., & Goldstein, Z. (2024). English foreign language reading and spelling diagnostic assessments informing teaching and learning of young learners. *Language Testing*, 41(1), 60–88. <https://doi.org/10.1177/02655322231162838>
- Khabbazzbashi, N., Nakatsuhara, F., Inoue, C., Kaplan, G., & Green, A. (2022). The design and validation of an online speaking test for young learners in Uruguay: challenges and innovations. *International Journal of TESOL Studies*, 4(1), 141-168. <https://doi.org/10.46451/ijts.2022.01.10>
- Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, 14, Article 48. <https://doi.org/10.1186/s40468-024-00324-x>
- Kim, A. A., Tywoniw, R. L., & Chapman, M. (2022). Technology-enhanced items in Grades 1–12 English language proficiency assessments. *Language Assessment Quarterly*, 19(4), 343–367. <https://doi.org/10.1080/15434303.2022.2039659>
- Kim, H., & Hwang, H. (2022). Assessing verb-construction integration in young learners of English as a foreign language: analyses of written and spoken production. *Language Learning*, 72(2), 497-533. <https://doi.org/10.1111/lang.12480>
- Kotowicz, J., Woll, B., & Herman, R. (2021). Adaptation of the British Sign Language Receptive Skills Test into Polish Sign Language. *Language Testing*, 38(1), 132–153. <https://doi.org/10.1177/0265532220924598>
- Lázaro-Ibarrola, A. (2023). Model texts in collaborative and individual writing among EFL children: noticing, incorporations, and draft quality. *International Review of Applied Linguistics in Language Teaching*, 61(2), 329-359. <https://doi.org/10.1515/iral-2020-0160>
- Luquin, M., & García Mayo, M. D. P. (2024). A longitudinal study of the effects of model texts on EFL children's written production. *System*, 120, 103190. <https://doi.org/10.1016/j.system.2023.103190>
- Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, 49, Article 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- Madyarov, I., Movsisyan, V., Madoyan, H., Galikyan, I., & Gasparyan, R. (2021). New validity evidence on the TOEFL Junior® standard test as a measure of progress.

- ETS Research Report Series, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12334>
- Mangold, M., Lenhard, W., Schindler, J., Schulz, D., & Richter, T. (2024). Assessing vocabulary of bilingual German-Turkish preschool children. *Bilingualism: Language and Cognition*, 27(4), 678–689. <https://doi.org/10.1017/S1366728923000822>
- Maplethorpe, L., Kim, H., Hunte, M. R., Vincett, M., & Jang, E. E. (2022). Student-generated questions in literacy education and assessment. *Journal of Literacy Research*, 54(1), 74–97. <https://doi.org/10.1177/1086296X221076436>
- Monster, I., Tellings, A., Burk, W. J., Keuning, J., Segers, E., & Verhoeven, L. (2021). Assessing children's incremental word knowledge in the upper primary grades. *Language Testing*, 38(4), 536–557. <https://doi.org/10.1177/0265532220961541>
- Moraes, I. T. & Batista, E. G. (2020). Letramento em avaliação para professores de línguas estrangeiras para crianças: orientações teórico-práticas [Assessment literacy for foreign language teachers of children: theoretical and practical guidelines]. *Revista Horizontes de Linguística Aplicada*, 19(2), 15–42. <https://doi.org/10.26512/rhla.v19i2.26804>
- Morea, N., Kasprowicz, R. E., Morrison, A., & Silvestri, C. (2024). Diverse population, homogenous ability: The development of a new receptive vocabulary size test for young language learners in England using Rasch analysis. *Research Methods in Applied Linguistics*, 3(3), 100166. <https://doi.org/10.1016/j.rmal.2024.100166>
- Nakanishi, Y., & Takeuchi, O. (2024). Validating the foreign language enjoyment scale for young learners: An exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100167. <https://doi.org/10.1016/j.rmal.2024.100167>
- Nikolov, M., & Timpe-Laughlin, V. (2021). Assessing young learners of foreign language abilities. *Language Teaching*, 54(1), 1–37. <https://doi.org/10.1017/S0261444820000294>
- O'Shannessy, C., Davis, V., Bartlett, J., Nelson, A., & Foster, D. (2023). Developing the Little Kids' Word List app, a fair assessment tool of communicative development for young Aboriginal children in multilingual families in Central Australia. *Studies in Language Assessment*, 12(2), 286–320.
- Patekar, J. (2021). A look into the practices and challenges of assessing young EFL learners' writing in Croatia. *Language Testing*, 38(3), 456–479. <https://doi.org/10.1177/0265532221990657>
- Peterson, S. S., Altidor, A., & Kerwood, J. (2021). Young children's written and verbal responses in a dynamic assessment context. *Assessing Writing*, 49, Article 100543. <https://doi.org/10.1016/j.asw.2021.100543>
- Pinar, Y., Ünal, F., & Pinar, N. K. (2021). Language proficiency analysis of Turkish children from high-income and highly educated families aged 5 in Antalya. *Language Testing in Asia*, 11, Article 28. <https://doi.org/10.1186/s40468-021-00138-1>
- Poetsch, S. (2023). Unguided. Assessing young EAL/D learners' achievements in mainstream curriculum areas. *Studies in Language Assessment*, 12(2), 28–58.
- QUEVEDO-CAMARGO, G., & TONELLI, J. R. A. (2024). Inglês e avaliação no ensino fundamental I: Propostas para o PNE 2025-2035 [English and assessment in primary school I: proposals for the National Plan of Education 2025-2035]. *Estudos Em Avaliação Educacional*, 35, e10547. <https://doi.org/10.18222/eae.v35.10547>
- Reed, D. J. and Stansfield, C. W. (2023). The MLAT-Elementary: History, adaptations and current uses. In E. Wen, P. Skehan, & R. Sparks (Eds.), *Language aptitude: Theory*

- and practice (pp. 25-46). Cambridge University Press & Assessment
- Reed, D. K., Ma, J., & Gerde, H. K. (2023). Resiliency and vulnerability in early grades writing performance during the COVID-19 pandemic. *Assessing Writing*, 57, Article 100741. <https://doi.org/10.1016/j.asw.2023.100741>
- Révész, A., Bunting, L., Florea, A., Gilabert, R., Hård af Segerstad, Y., Mihiu, I. P., & Vasalou, A. (2021). The effects of multiple-exposure textual enhancement on child L2 learners' development in derivational morphology: A multi-site study. *TESOL Quarterly*, 55(3), 901–930. <https://doi.org/10.1002/tesq.3036>
- Ryan, M., Khosronejad, M., Barton, G., Myhill, D., & Kervin, L. (2022). Reflexive writing dialogues: Elementary students' perceptions and performances as writers during classroom experiences. *Assessing Writing*, 51, Article 100592. <https://doi.org/10.1016/j.asw.2021.100592>
- Shin, S., & Lee, S. (2024). Effects of the type of audio-visual input and listenability on young L2 learners' listening comprehension. *TESOL Quarterly*, 58(4), 1580–1608. <https://doi.org/10.1002/tesq.3298>
- Sok, S., & Shin, H. W. (2023). Investigating the dimensions and determinants of children's narrative writing in Korean. *Assessing Writing*, 57, Article 100740. <https://doi.org/10.1016/j.asw.2023.100740>
- Sok, S., Shin, H. W., & Do, J. (2021). Exploring which test-taker characteristics predict young L2 learners' performance on listening and reading comprehension tests. *Language Testing*, 38(3), 378–400. <https://doi.org/10.1177/0265532221991134>
- Suárez, M. M. (2022). A need for an aptitude test for young learners in Catalan: The case of the Modern Language Aptitude Test - Elementary in Catalan. *Language Teaching Research Quarterly*, 31, 101-118. <https://doi.org/10.32038/ltrq.2022.31.08>
- Suárez, M.-d.-M., & Stansfield, C. W. (2023). Young learners' bilingual status and cognitive development in foreign language aptitude testing. *ITL - International Journal of Applied Linguistics*, 174(2), 291-315. <https://doi.org/https://doi.org/10.1075/itl.22017.sua>
- Suhan, M., Papageorgiou, S., & Wolf, M. K. (2024). Mapping the scores of the TOEFL Primary® Writing test to the Common European Framework of Reference (CEFR) levels (Research Memorandum ETS RM-24-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-03.pdf>
- Teng, M. F., & Zhang, L. J. (2024). Validating a new method for assessing young foreign language learners' metacognitive knowledge. *Research Methods in Applied Linguistics*, 3(2), Article 100121. <https://doi.org/10.1016/j.rmal.2024.100121>
- Teng, M. F., Wang, C., & Zhang, L. J. (2022). Assessing self-regulatory writing strategies and their predictive effects on young EFL learners' writing performance. *Assessing Writing*, 51, Article 100573. <https://doi.org/10.1016/j.asw.2021.100573>
- Timpe-Laughlin, V., & Cho, Y. (2021). Reflecting on assessing young foreign language learners. *Language Testing*, 38(3), 343–355. <https://doi.org/10.1177/0265532221998264>
- Torregrossa, J., Listanti, A., Bongartz, C., & Narinis T. (2024). Adding discourse to sentence repetition tasks: Under which conditions does bilingual children's performance improve?. *Research Methods in Applied Linguistics*, 3(2), Article 100107. <https://doi.org/10.1016/j.rmal.2024.100107>
- Tsagari, D. & H. Bøhn (eds.) (2022) Recent Developments in Language Testing and Assessment (Special Issue) Issue 7. Languages (open access journal)

https://www.mdpi.com/journal/languages/special_issues/Language_Testing_Assessment?fbclid=IwAR2_A6SIGEKehpZ95o5AQvPmM3kzi6O3rqypHjoTPA22LIdl6F_dHJLpnas

- Valentine, K. A., Truckenmiller, A. J., Troia, G. A., & Aldridge, S. (2021). What is the nature of change in late elementary writing and are curriculum-based measures sensitive to that change? *Assessing Writing*, 50, Article 100567. <https://doi.org/10.1016/j.asw.2021.100567>
- Wallace, M.P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72(1), 5-44. <https://doi.org/10.1111/lang.12424>
- Ward, E. (2021). An integrated mixed-methods study of contract grading's impact on adolescents' perceptions of stress in high school English: a pilot study. *Assessing Writing*, 48, Article 100508. <https://doi.org/10.1016/j.asw.2020.100508>
- Ward, E. (2021). Easing stress: Contract grading's impact on adolescents' perceptions of workload demands, time constraints, and challenge appraisal in high school English. *Assessing Writing*, 48, Article 100526. <https://doi.org/10.1016/j.asw.2021.100526>
- Weisleder, A., Friend, M., Tsui, A. S. M., & Marchman, V. A. (2024). Using parent report to measure vocabulary in young bilingual children: A scoping review. *Language Learning*, 74(2), 468-505. <https://doi.org/10.1111/lang.12617>
- Wolf, M. K. (2022). Interconnection between constructs and consequences: a key validity consideration in K–12 English language proficiency assessments. *Language Testing in Asia*, 12, Article 44. <https://doi.org/10.1186/s40468-022-00194-1>
- Wolf, M. K. (2024). Assessment of young language learners. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 312-325). Wiley-Blackwell.
- Wolf, M. K., & Oh, S. (2024). Investigating the types and use of feedback in middle-school English language learners' academic writing. *Language Learning & Technology*, 28(1), 1-25. <https://hdl.handle.net/10125/73583>
- Wolf, M. K., Lopez, A. A., & Lee, J. (2023). An investigation of the use of standardized and local assessments for young EAL students. In G. Brooks, J. Clenton, & S. Fraser (Eds.), *EAL research for the classroom: Practical and pedagogical implications* (pp. 164-184). Routledge. <https://doi.org/10.4324/9781003274889-13>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). Exploring the provision of AI-based feedback for TOEFL Junior Writing practice tasks (Research Memorandum No. RM 24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>
- Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). Design framework for the TOEFL Primary® Writing test (Research Memorandum ETS RM-24-02). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-02.pdf>
- Wolf, M. K., Yoo, H., & Ballard, L. (2024). The relationship between English language proficiency and academic performance among English learners: A subgroup analysis by home language background. *Learning and Individual Differences*, 109, 102388. <https://doi.org/10.1016/j.lindif.2023.102388>
- Wood, C., Hoge, R., Schatschneider, C., & Castilla-Earls, A. (2021). Predictors of item accuracy on the Test de Vocabulario en Imágenes Peabody for Spanish-English speaking children in the United States. *International Journal of Bilingual Education and Bilingualism*, 24(8), 1178–1192. <https://doi.org/10.1080/13670050.2018.1547266>

Xie, Q. (2022). An online database of English assessment tasks for developing student teachers' assessment literacy. *International Journal of Mobile Learning and Organization*, 16, Article 4. <https://doi.org/10.1504/IJMLO.2022.10048804>

Assessing Learners with Specific Learning Differences

- Alicia Kim, A., Yumsek, M., Kemp, J. A., Chapman, M., & Gary Cook, H. (2023). Universal tools activation in English language proficiency assessments: A comparison of Grades 1–12 English learners with and without disabilities. *Language Testing*, 40(4), 877–903. <https://doi.org/10.1177/02655322221149009>
- Altherr Flores, J. A. (2021). The interplay of text and image on the meaning-making processes of adult L2 learners with emerging literacy: Implications for test design and evaluation frameworks. *Language Assessment Quarterly*, 18(5), 508–529. <https://doi.org/10.1080/15434303.2021.1984491>
- An, H., & Li, S. (2024). Task-specific writing anxiety and self-efficacy are separate from general L2 writing anxiety and self-efficacy and they have differential associations with the effects of written corrective feedback in pre-task and within-task planning. *System*, 126, Article 103480. <https://doi.org/10.1016/j.system.2024.103480>
- Deygers, B., Bigelow, M., Lo Bianco, J., Nadarajan, D., & Tani, M. (2021). Low print literacy and its representation in research and policy. *Language Assessment Quarterly*, 18(5), 463–476. <https://doi.org/10.1080/15434303.2021.1903471>
- Du Plessis, C. (2021). Basic education and academic literacy: conflicting constructs in the South African National Senior Certificate (NSC) Language Examination. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 95–116). Multilingual Matters.
- Eberharter, K., Kormos, J., Guggenbichler, E., Ebner, V. S., Suzuki, S., Moser-Frötscher, D., Konrad, E., & Kremmel, B. (2023). Investigating the impact of self-pacing on the L2 listening performance of young learner candidates with differing L1 literacy skills. *Language Testing*, 40(4), 960–983. <https://doi.org/10.1177/02655322221149642>
- Guzman-Orth, D., Steinberg, J., & Albee, T. (2023). English learners who are blind or visually impaired: A participatory design approach to enhancing fairness and validity for language testing accommodations. *Language Testing*, 40(4), 933–959. <https://doi.org/10.1177/02655322231159143>
- Hooft, H., Schiepers, M., & Vandommele, G. (2021). Developing and validating a multilingual literacy test for asylum seekers. *Language Assessment Quarterly*, 18(5), 530–546. <https://doi.org/10.1080/15434303.2021.1931230>
- Kormos, J., & Smith, A. M. (2024). Assessment of the language skills of language learners with SpLDs. In J. Kormos, & A. M. Smith (Eds.), *Teaching languages to students with specific learning differences* (2nd ed., pp. 143–164). Multilingual Matters. <https://doi.org/10.2307/jj.22679668.10>
- Lindeberg, S., Müller, N., & Samuelsson, C. (2023). Multimodality in PPA: Hand movements as resources in conversations and testing. *Journal of Interactional Research in Communication Disorders*, 14(2), 268–291. <https://doi.org/10.1558/jircd.24306>
- Marinho, N. L., Witmer, S. E., Jess, N., & Roschmann, S. (2023). A meta-Analysis of accommodation effects for English learners: Considering possible moderators. *Language Assessment Quarterly*, 20(3), 296–318. <https://doi.org/10.1080/15434303.2023.2237499>

- Motteram, J., Spiby, R., Bellhouse, G., & Sroka, K. (2023). Implementation of an accommodations policy for candidates with diverse needs in a large-scale testing system. *Language Testing*, 40(4), 904-932. <https://doi.org/10.1177/02655322231166587>
- O'Sullivan, B., & Chalhoub-Deville, M. (2021). Language testing for migrants: Co-constructing validation. *Language Assessment Quarterly*, 18(5), 547–557. <https://doi.org/10.1080/15434303.2021.1986513>
- Ożańska-Ponikwia, K., Piechurska-Kuciel, E., & Skalacka, K. (2023). Emotional intelligence as a mediator in the relationship between neuroticism and L2 achievement. *Applied Linguistics Review*, 14(1), 67-86. <https://doi.org/10.1515/applirev-2020-0008>
- Pelleriti, M. (2023). Item writing and dyslexic students: Promoting best practices. *Iperstoria*, 21. 217–231. <https://iperstoria.it/article/view/1285/1315>
- Pomarolli, G., Artoni, D., Boschiero, M., & Piccinin, S. (2023). Inclusive and accessible teaching of Russian L2 to absolute beginners: A comparison between TBLT and PPP. *Instructed Second Language Acquisition*, 7(2), 166-193. <https://doi.org/10.1558/isla.26047>
- Randez, R. A., & Cornell, C. (2023). Advancing equity in language assessment for learners with disabilities. *Language Testing*, 40(4), 984-999. <https://doi.org/10.1177/02655322231169442>
- Roschmann, S., Witmer, S. E., & Volker, M. A. (2021). Examining provision and sufficiency of testing accommodations for English learners. *International Journal of Testing*, 21(1), 32–55. <https://doi.org/10.1080/15305058.2021.1884872>
- Rüsseler, J., Arendt, D., Münte, T. F., Mohammadi, B., & Boltzmann, M. (2021). Literacy affects brain structure—What can we learn for language assessment in low literates? *Language Assessment Quarterly*, 18(5), 492–507. <https://doi.org/10.1080/15434303.2021.1931231>
- Schmidgall, J., Huo, Y., Cid, J., & Wei, Y. (2024). Investigating fairness claims for a general-purposes assessment of English proficiency for the international workplace: Do full-time employees have an unfair advantage over full-time students? ETS Research Report Series, 2024(1), 1–20. <https://doi.org/10.1002/ets2.12380>
- Suárez, M.-d.-M., & Stansfield, C. W. (2023). Young learners' bilingual status and cognitive development in foreign language aptitude testing. *ITL - International Journal of Applied Linguistics*, 174(2), 291-315. <https://doi.org/https://doi.org/10.1075/itl.22017.sua>
- Tsagari, D. (2022). Inclusive and accessible assessment. In D. Tsagari, J. Nijakowska, & E. Guz. (Eds.), *Good practices in inclusive language education and foreign language teacher training - universal design in education and increasing accessibility, especially for marginalized groups*. https://scaled.uw.edu.pl/wp-content/uploads/sites/115/2022/11/SCALED_Output_1_Report.pdf
- Vanbuel, M., & Deygers, B. (2024). Assessing Receptive Skills in an Adult L2 Learner Population with Diverse Educational Backgrounds. *Language Assessment Quarterly*, 21(4–5), 379–403. <https://doi.org/10.1080/15434303.2024.2438151>
- Viera, C., & Arispe, K. (2022). Can Spanish programs evaluate oral proficiency gains for all learners? An empirical analysis of traditional and nontraditional Spanish majors' oral proficiency development. *Foreign Language Annals*, 55(3), 832– 852. <https://doi.org/10.1111/flan.12635>

- Vössing, L., & Kern, F. (2023). The co-construction of pragmatic competencies in different settings: The case of two children with autism spectrum disorder. *Journal of Interactional Research in Communication Disorders*, 14(2), 195-219. <https://doi.org/10.1558/jircd.24423>
- Wallace, M.P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72(1), 5-44. <https://doi.org/10.1111/lang.12424>

Assessing Digital and Multimodal Literacy

- Beltrán-Palanques, V. (2024). Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP. *Assessing Writing*, 60, 100809. <https://doi.org/10.1016/j.asw.2024.100809>
- Burton, J. D. (2021). The face of communication breakdown: Multimodal repair in L2 oral proficiency interviews. *Papers in Language Testing and Assessment*, 10(2), 30–61. http://www.altanz.org/uploads/5/9/0/8/5908292/3._plta_10_2__burton.pdf
- Burton, J. D. (2021). The face of communication breakdown: Multimodal repair in L2 oral proficiency interviews. *Papers in Language Testing and Assessment*, 10(2), 30–61. http://www.altanz.org/uploads/5/9/0/8/5908292/3._plta_10_2__burton.pdf
- Chen Hsieh, J. (2021). Digital storytelling outcomes and emotional experience among middle school EFL Learners: Robot-assisted versus PowerPoint-assisted mode. *TESOL Quarterly*, 55(3), 994–1010. <https://doi.org/10.1002/tesq.3043>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487-2512. <https://doi.org/10.1177/13621688211046740>
- Chong, J. J. Q., & Aryadoust, V. (2023). Investigating the effect of multimodality and sentiments on speaking assessments: A facial emotional analysis. *Education and Information Technologies*, 28, 7413–7436. <https://doi.org/10.1007/s10639-022-11478-7>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- Gràcia, M., Adam-Alcocer, A., & Castillo Mardones, P. (2023). Exploring the impact of a teacher development programme using a digital application on linguistic interactions in the classroom: a multiple case study. *Applied Linguistics Review*, 14(5), 1305–1343. <https://doi.org/10.1515/applirev-2020-0132>
- Greenblatt, P., & McDonald, P. (2024). *Using Multimodal Public Service Announcement*

- Video Projects as Communicative Language Assessments. RELC Journal*, 55(2), 559–566. <https://doi.org/10.1177/00336882221102234>
- Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing*, 57, Article 100869. <https://doi.org/10.1016/j.jslw.2022.100869>
- Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing*, 57, Article 100869. <https://doi.org/10.1016/j.jslw.2022.100869>
- Kang, S., & Kim, Y. (2024). Examining the quality of mobile-assisted, video-making task outcomes: The role of proficiency, narrative ability, digital literacy, and motivation. *Language Teaching Research*, 28(6), 2326–2353. <https://doi.org/10.1177/13621688211047984>
- Kessler, M. (2024). Assessment. In M. Kessler (Ed.), *Digital multimodal composing: Connecting theory, research and practice in second language acquisition* (pp. 109–120). Multilingual Matters. <https://doi.org/10.2307/jj.22679723.11>
- Kessler, M. (2024). Assessment. In M. Kessler (Ed.), *Digital multimodal composing: Connecting theory, research and practice in second language acquisition* (pp. 109–120). Multilingual Matters. <https://doi.org/10.2307/jj.22679723.11>
- Khabbazzbashi, N., Chan, S. and Clark, T. (2023). Towards the new construct of academic English in the digital age, *ELT Journal*, 77(2), 207–216 <https://doi.org/10.1093/elt/ccac010>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73–95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247. <https://doi.org/10.1016/j.jeap.2023.101247>
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247. <https://doi.org/10.1016/j.jeap.2023.101247>
- Kılıçkaya, F. (2022). Pre-service EFL teacher candidates' concerns over digital language assessment and test anxiety. In W. M. El-Henawy & M. del Mar Suárez (Eds.), *English as a foreign language in a new-found post-pandemic world* (pp. 164–181). IGI Global. <https://doi.org/10.4018/978-1-6684-4205-0>
- Lindeberg, S., Müller, N., & Samuelsson, C. (2023). Multimodality in PPA: Hand movements as resources in conversations and testing. *Journal of Interactional Research in Communication Disorders*, 14(2), 268–291. <https://doi.org/10.1558/jircd.24306>
- Mootosamy, K., & Aryadoust, V. (2024). Transitioning from communicative competence to multimodal and intercultural competencies: A systematic review. *Societies*, 14(115). <https://doi.org/10.3390/soc14070115>

- Nakatsuhara, F., May, L., & Khabbazzbashi, N. (2023). *Exploring the speaking construct in academic settings in a digital age*. British Council, Cambridge Assessment English, and IDP: IELTS Australia.
<https://search.informit.org/doi/epdf/10.3316/informit.262806705843094>
- Palmour, L. (2024). Assessing speaking through multimodal oral presentations: The case of construct underrepresentation in EAP contexts. *Language Testing*, 41(1), 9–34.
<https://doi.org/10.1177/02655322231183077>
- Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, 55(1), 105–133.
<https://doi.org/10.1002/tesq.579>
- Ryu, J., Ae Kim, Y., Eum, S., Park, S., Chun S., & Yang, S. (2022). The assessment of memes as digital multimodal composition in L2 classrooms. *Journal of Second Language Writing*, 57, Article 100914. <https://doi.org/10.1016/j.jslw.2022.100914>
- Ryu, J., Ae Kim, Y., Eum, S., Park, S., Chun S., & Yang, S. (2022). The assessment of memes as digital multimodal composition in L2 classrooms. *Journal of Second Language Writing*, 57, Article 100914. <https://doi.org/10.1016/j.jslw.2022.100914>
- Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in second language Chinese. *The Modern Language Journal*, 105(3), 740–759.
<https://doi.org/10.1111/modl.12725>
- Yu, G. & Xu, J. (Eds.). (2024). Language test validation in a digital age. Studies in Language Testing Volume 52. Cambridge University Press & Assessment.
- Zhang, E. D., & Yu, S. (2024). Understanding L2 student writers' self-assessment in digital multimodal composing: A process-oriented approach. *System*, 121, 103219.
<https://doi.org/10.1016/j.system.2024.103219>
- Zhang, E. D., & Yu, S. (2024). Understanding L2 student writers' self-assessment in digital multimodal composing: A process-oriented approach. *System*, 121, 103219.
<https://doi.org/10.1016/j.system.2024.103219>
- Zhang, M., & Chen, W. (2022). Assessing collaborative writing in the digital age: An exploratory study. *Journal of Second Language Writing*, 57, Article 100868.
<https://doi.org/10.1016/j.jslw.2022.100868>
- Zimmermann, S., Schwarz, L., Peters, A. & Depner, G. (2024). Enhancing teachers and test takers' assessment literacy? Insights from test preparation for the digital TestDaF. In: B. Baker & L. Taylor (Eds.), *Language Assessment Literacy and Competence Volume 1: Research and Reflections from the Field* (pp. 110–128). Cambridge: Cambridge University Press.

Language Test Construction and Development

Validity and Validation

- Abdullaeva, B. S., Çakmak, F., & Abdullaev, D. (2024). Paper assessment or online assessment: Exploring the impact of assessment modes on EFL students' language learning outcomes and personal development. *Language Testing in Asia*, 14(, Article 35. <https://doi.org/10.1186/s40468-024-00309-w>
- Ali, M. M., & Hamid, M. O. (2023). Teacher agency and washback: Insights from classrooms in rural Bangladesh. *ELT Journal*, 77(1), 52-61. <https://doi.org/10.1093/elt/ccac005>
- Allen, D., & Koizumi, R. (2024). The English Speaking Achievement Test for Japanese Junior High School Students (ESAT-J): An analysis of public criticisms using the socio-cognitive framework. *Language Assessment Quarterly*, 21(4-5), 404-426. <https://doi.org/10.1080/15434303.2024.2438982>
- Aryadoust, V. (2023). The vexing problem of validity and the future of second language assessment. *Language Testing*, 40(1), 8-14. <https://doi.org/10.1177/02655322221125204>
- Aryadoust, V. (2024). Topic and accent coverage in a commercialized L2 listening test: Implications for test-takers' identity. *Applied Linguistics*, 45(5), 765-785. <https://doi.org/10.1093/applin/amad062>
- Aryadoust, V., Foo, S., & Ng, L. Y. (2022). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Language Testing*, 39(1), 56-89. <https://doi.org/10.1177/02655322211026876>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6-40. <https://doi.org/10.1177/0265532220927487>
- Aull, L. (2021). Directed Self-Placement: Subconstructs and group differences at a US university. *Assessing Writing*, 48, Article 100522. <https://doi.org/10.1016/j.asw.2021.100522>
- Barkaoui, K. (2021). *Evaluating tests of second language development*. Peter Lang. <https://doi.org/10.3726/b18059>
- Barkaoui, K. (2024). Examining performance on an integrated writing task from a Canadian English language proficiency test. *Canadian Modern Language Review*, 80(2), 77-115. <https://doi.org/10.3138/cmlr-2023-0022>
- Barkaoui, K. (2024). Exploring the effects of task difficulty and learner variables on performance on picture description writing tasks. *Assessing Writing*, 60, 100827. <https://doi.org/10.1016/j.asw.2024.100827>
- Barkaoui, K. (2024). Multi-faceted Rasch analysis for evaluating second language tests. In Kunnan, A. (Ed.), *The concise companion to language assessment* (pp. 602-618). Wiley-Blackwell.
- Barkaoui, K. (2024). The academic achievement of undergraduate students with different score profiles on the TOEFL iBT: A replication study. *Language Assessment Quarterly*. <https://doi.org/10.1080/15434303.2024.2438142>
- Barkaoui, K. & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language*

- Assessment, 12(1), 26-63. 2._sila_12_1__barkaoui_and_woodworth.pdf (altaanz.org)
- Barkaoui, K., & Hadidi, A. (2021). *Assessing changes in second Language writing performance*. Routledge. <https://doi.org/10.4324/9781003092346>
- Barkaoui, K., & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26–63. <https://doi.org/10.58379/QCFS2805>
- Batty, A. O. (2021). Shūjukudo to seibetsu ga bideo L2 chōkairyoku tesuto ni okeru shikaku no kōka ni ataeru eikyō [The impact of proficiency and gender on the effect of visuals in video L2 listening comprehension tests]. *Journal of Foreign Language Education*, 17, 1–18.
- Batty, A. O., Haug, T., Ebling, S., Tissi, K., & Sidler-Miserez, S. (2023). Challenges in rating signed production: A mixed-methods study of a Swiss German Sign Language form-recall vocabulary test. *Language Testing*, 40(2), 352-374. <https://doi.org/10.1177/02655322221122774>
- BijanKhan, M., ShayesteFar, P., & Mohebbi, H. (2023). Assessing the content typicality and construct of Persian language proficiency test (PLPT) for non-Persian speakers: A corpus-informed study. *Language Testing in Asia*, 13, Article 11. <https://doi.org/10.1186/s40468-023-00217-5>
- Binti Hasrol, S., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), Article 100023. <https://doi.org/10.1016/j.rmal.2022.100023>
- Bourdeaud'Hui, H., Aesaert, K., & van Braak, J. (2021). Exploring the validity of a comprehensive listening test to identify differences in primary school students' listening skills. *Language Assessment Quarterly*, 18(3), 228–252. <https://doi.org/10.1080/15434303.2020.1860059>
- Brown, J. D., Panahi, A., & Mohebbi, H. (2023). James Dean Brown's 50 years of work in language testing: A systematic review. *Language Teaching Research Quarterly*, 37, 4–75. <https://doi.org/10.32038/ltrq.2023.37.02>
- Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, 40(1), 15-23. <https://doi.org/10.1177/02655322221127896>
- Brunsting, N., Yu, Q., Smart, J., and Bingham, W.P. (2022). Investigating linkages between international students' English language proficiency, social-contextual outcomes, and wellbeing in U.S. universities. IELTS Research Reports Online Series, No. 4/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/brunsting-yu-smart-bingham-2022.ashx>
- Buffington, J., Demos, A., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, 43(3), 635–662. <https://doi.org/10.1017/S0272263121000127>
- Buzick, H. M., Wolf, M. K., & Ballard, L. (2024). Achievement and growth on English language proficiency and content assessments for English learners in elementary grades. *Educational Measurement: Issues and Practice*, 43(1), 83-95. <http://doi.org/10.1111/emip.12588>
- Chalhoub-Deville, M., & O'Sullivan, B. (2020). *Validity: Theoretical development and*

- integrated arguments*. University of Toronto Press.
- Chan, S., & May, L. (2023). Towards more valid scoring criteria for integrated reading-writing and listening-writing summary tasks. *Language Testing*, 40(2), 410–439. <https://doi.org/10.1177/02655322221135025>
- Chapelle, C. A., & Lee, H. (2022). Conceptions of validity. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 17–31). Routledge. <https://doi.org/10.4324/9781003220756-3>
- Chapelle, C. A., & Lee, H.-W. (2021). Understanding argument-based validity in language testing. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 19–44). Cambridge University Press. <https://doi.org/10.1017/9781108669849>
- Chapelle, C. A., & Voss, E. (Eds.). (2021). *Validity argument in language testing: Case studies of argument-based validation research*. Cambridge University Press. <https://doi.org/10.1017/9781108669849>
- Chapelle, C. A., Voss, E., & Kim, H. (2024). Designing evaluations for validation of language assessments. In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment*. (pp. 67–79). Wiley-Blackwell.
- Cheng, L. (2024). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–1146). Chichester, UK: John Wiley & Sons. doi:10.1002/9781118411360.wbcla071. <https://onlinelibrary.wiley.com/doi/10.1002/9781118411360.wbcla071>
- Cheng, L., & Im, G.-H. (2021). Translation of surveys for international research: Beyond equivalence. In G. Harbaugh & U. Luhanga (Eds.), *Basic elements of survey research in education: Addressing the problems your advisor never told you about* (pp. 219–236). Information Age Publishing.
- Cheng, L., & Zheng, Y. (2021). Measuring anxiety. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 177–186). <https://doi.org/10.4324/9781351034784-20>
- Cheng, L., Im, G.-H., & Jabeen, R. (2021). Whose English(es) are we assessing and by whom? In B. Lantaigne, C. Coombe., & J.D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp. 169–179). Springer. https://doi.org/10.1007/978-981-33-4232-3_12
- Choi, Y. (2021). What interpretations can we make from scores on graphic-prompt writing (GPW) tasks? An argument-based approach to test validation. *Assessing Writing*, 48, Article 100523. <https://doi.org/10.1016/j.asw.2021.100523>
- Choi, Y. D. (2022). Validity of score interpretations on an online English placement writing test. *Language Testing in Asia*, 12, Article 42. <https://doi.org/10.1186/s40468-022-00187-0>
- Choi, Y., Vo, S., & Ockey, G. J. (2024). Investigation into the factor structure of a local English placement test of oral communication. *International Journal of Applied Linguistics*, 34(1), 63–85. <https://doi.org/10.1111/ijal.12478>
- Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, 60, Article 101003. <https://doi.org/10.1016/j.jslw.2023.101003>
- Coniam, D., Lee, T., Milanovic, M., Pike, N., & Zhao, W. (2022). Role of expert judgement in language test validation. *Language Education and Assessment*, 5(1), 18–33. <https://doi.org/10.29140/lea.v5n1.769>

- Crowther, D., Isbell, D. R., & Nishizawa, H. (2023). Second language speech comprehensibility and acceptability in academic settings: Listener perceptions and speech stream influences. *Applied Psycholinguistics*, 44(5), 858–888. <https://doi.org/10.1017/S0142716423000346>
- Daller, M., Müller, A., & Wang-Taylor, Y. (2021). The C-test as predictor of the academic success of international students. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1502–1511. <https://doi.org/10.1080/13670050.2020.1747975>
- Dang, C. N., & Dang, T. N. Y. (2023). The predictive validity of the IELTS test and contribution of IELTS preparation courses to international students' subsequent academic study: Insights from Vietnamese international students in the UK. *RELC Journal*, 54(1), 84-98. <https://doi.org/10.1177/0033688220985533>
- Darabad, M. A., Abbasian, G., Mowlaie, B., & Abusaeedi, A. R. (2023). On the construct and perceptual validity measures of L1-based vs. L2-based elicitation as a measure of L2 classroom performance assessment. *Language Testing in Asia*, 13, Article 13 (2023). <https://doi.org/10.1186/s40468-023-00218-4>
- Devos, N., Nizonkiza, D., & Lynch, S. (2024). Developing and validating a post-admission screening-diagnostic assessment procedure to offer language support in college diploma programs. *Canadian Journal of Applied Linguistics*, 27(1), 51-77. <https://doi.org/10.37213/cjal.2024.33259>
- Deygers, B., & Vanbuel, M. (2022). Gauging the impact of literacy and educational background on receptive vocabulary test scores. *Language Testing*, 39(2), 191–211. <https://doi.org/10.1177/02655322211049097>
- Deygers, B., Bigelow, M., Lo Bianco, J., Nadarajan, D., & Tani, M. (2021). Low print literacy and its representation in research and policy. *Language Assessment Quarterly*, 18(5), 463–476. <https://doi.org/10.1080/15434303.2021.1903471>
- Drackert, A., Zimmermann, S., Marks, D. & Korflür, D. (2023). New test format – new research agenda: An overview of the technology-related research at g.a.s.t. In: *Conference Proceedings of the ALTE 8th International Conference: Language Assessment Fit for the Future* (pp. 34–38).
- Drown, L., Giovannone, N., Pisoni, D., & Theodore, R. (2023). Validation of two measures for assessing English vocabulary knowledge on web-based testing platforms: brief assessments. *Linguistics Vanguard*, 9(1), 99-111. <https://doi.org/10.1515/lingvan-2022-0116>
- Dudley, A., Marsden, E., & Bovolenta, G. (2024). A Context-Aligned Two Thousand Test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England. *Language Testing*, 41(4), 759–791. <https://doi.org/10.1177/02655322241261415>
- Dunlea, J., O'Sullivan, B., Dunn, K., Felice, M., Lightfoot, A., Patel, M., Spiby, R., & Westbrook, C. (2024). Assessing English as an international language. In: Galloway, N. & Fuad Selvi, A. (Eds.). *Routledge Handbook of English as an International Language* (pp.184-198), Routledge. DOI: 10.4324/9781003203261-16
- Edmonds, A., Clenton, J., & Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System*, 108, 102855. <https://doi.org/10.1016/j.system.2022.102855>
- Eerdemutu, L., Dewaele, J.-M., & Wang, J. (2024). Developing a short language classroom engagement scale (Lces) and linking it with needs satisfaction and achievement. *System*, 120, 103189. <https://doi.org/10.1016/j.system.2023.103189>

- Eguchi, M., Suzuki, S., & Suzuki, Y. (2022). Lexical competence underlying second language word association tasks: Examining the construct validity of response type and response time measures. *Studies in Second Language Acquisition*, 44(1), 112-142. <https://doi.org/10.1017/S0272263121000164>
- Elosua, P. (2022). Validity evidences for scoring procedures of a writing assessment task. A case study on consistency, reliability, unidimensionality and prediction accuracy. *Assessing Writing*, 54, Article 100669. <https://doi.org/10.1016/j.asw.2022.100669>
- Elosua, P. (2024). A Three-Step DIF Analysis of a Reading Comprehension Test Across Regional Dialects to Improve Test Score Validity. *Language Assessment Quarterly*, 21(2), 141–158. <https://doi.org/10.1080/15434303.2024.2307621>
- Erickson, G., & Tholin, J. (2022). Overall, a good test, but...—Swedish lower secondary teachers' perceptions and use of national test results of English. *Languages*, 7(1), Article 64. <https://doi.org/10.3390/languages7010064>
- Erlam, R., & Rau, W. T. (2023). Fairer for all: Teachers identifying and solving problems with the assessment of English language learners in the mainstream classroom. *Studies in Language Assessment*, 12(2), 94-116.
- Estaji, M., & Banitalebi, Z. (2023). A study of Test-Taking strategies of Iranian IELTS repeaters: Any change in the strategy use? *International Journal of Testing*, 23(3), 205-230. <https://doi.org/10.1080/15305058.2023.2195662>
- Fan, J., & Zhang, X. (2024). Applying the Rasch model in language assessment. Foreign Language Teaching and Research Press.
- Fox, J., & Artemeva, N. (2022). *Reconsidering context in language assessment: Transdisciplinary perspectives, social theories, and validity*. Routledge. <https://doi.org/10.4324/9781351184571>
- Franken, M. (2023). The new national literacy tests for post-primary students in Aotearoa New Zealand: How process and design issues undermine principles of a strong and fair qualification. *Studies in Language Assessment*, 12(2), 59-93.
- García Laborda, J., & Fernández Álvarez, M. (2021). Multilevel Language tests: Walking into the land of the unexplored. *Language Learning & Technology*, 25(2), 1–25. <http://hdl.handle.net/10125/73428>
- George, A. C., & Robitzsch, A. (2021). Validating theoretical assumptions about reading with cognitive diagnosis models. *International Journal of Testing*, 21(2), 105–129. <https://doi.org/10.1080/15305058.2021.1931238>
- Getman, E. P. (2024). Four task characteristics and their influence on speaking performance and engagement. *English Teaching & Learning*, 48(2), 265–290. <https://doi.org/10.1007/s42321-024-00176-x>
- Ghamoushi, M., Zenouzagh, Z. M., & Hashamdar, M. (2022). Development and validation of a potential assessment inventory for assessing EFL teachers' ecological agency. *Language Testing in Asia*, 12, Article 37. <https://doi.org/10.1186/s40468-022-00190-5>
- Ghiasvand, F., Jahanbakhsh, A. A. & Sharifpour, P. (2023). Designing and validating an assessment agency questionnaire for EFL teachers: an ecological perspective. *Language Testing in Asia*, 13, Article 41. <https://doi.org/10.1186/s40468-023-00255-z>
- Graham, S., Harbaugh-Schattenkirk, A. G., Aitken, A., Harris, K. R., Ng, C., Ray, A., Wilson, J. M., & Wdowin, J. (2022). Writing motivation questionnaire: Validation and application as a formative assessment. *Assessment in Education: Principles, Policy & Practice*, 29(2), 238–261. <https://doi.org/10.1080/0969594X.2022.2080178>

- Grocott, C., Moe, E., & Verhelst, N. (2024). Investigating level descriptors and their impact on validity. *Studies in Language Assessment (SILA)*, 13(1), 199-236. <https://doi.org/10.58379/LIHU1555>
- Guapacha Chamorro, M. E. (2022). Cognitive validity evidence of computer- and paper-based writing tests and differences in the impact on EFL test-takers in classroom assessment. *Assessing Writing*, 51, Article 100594. <https://doi.org/10.1016/j.asw.2021.100594>
- Guzman-Orth, D., Steinberg, J., & Albee, T. (2023). English learners who are blind or visually impaired: A participatory design approach to enhancing fairness and validity for language testing accommodations. *Language Testing*, 40(4), 933–959. <https://doi.org/10.1177/02655322231159143>
- Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11, Article 16. <https://doi.org/10.1186/s40468-021-00132-7>
- Hadidi, A. (2023). Comparing summative and dynamic assessments of L2 written argumentative discourse: Microgenetic validity evidence. *Assessing Writing*, 55, Article 100691. <https://doi.org/10.1016/j.asw.2022.100691>
- Hamada, A., & Takaki, S. (2021). Approximate replication of Matsuda and Gobel (2004) for psychometric validation of the Foreign Language Reading Anxiety Scale. *Language Teaching*, 54(4), 535–551. <http://doi.org/10.1017/S0261444819000296>
- Handley, Z. L., & Wang, H. (2023). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency? *Language Assessment Quarterly*, 21(1), 3–32. <https://doi.org/10.1080/15434303.2023.2283839>
- Handley, Z. L., & Wang, H. (2024). What Do the Measures of Utterance Fluency Employed in Automatic Speech Evaluation (ASE) Tell Us About Oral Proficiency?. *Language Assessment Quarterly*, 21(1), 3-32.
- Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity arguments for automated essay scoring of young students' writing traits. *Language Assessment Quarterly*, 20(4-5), 399–420. <https://doi.org/10.1080/15434303.2023.2288253>
- Heeren, J., Speelman, D., & Wachter, L. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458–1473. <https://doi.org/10.1080/13670050.2019.1709411>
- Hidri, S. (2021). Linking the International English Language Competency Assessment suite of examinations to the Common European Framework of Reference. *Language Testing in Asia*, 11, Article 9. <https://doi.org/10.1186/s40468-021-00123-8>
- Hilden, R., Pulkkinen, J., & Rautopuro, J. (2022). The meaningfulness of two curriculum-based national tests of English as a foreign language. *Studies in Language Assessment*, 11(2), 130–163.
- Hirai, A., Oka, H., Kato, T., & Maeda, H. (2022). Development and validation of an English test measuring EFL learners' critical thinking skills. *Language Testing in Asia*, 12, Article 45. <https://doi.org/10.1186/s40468-022-00193-2>
- Hoeve, K. B. (2022). A validity framework for accountability: educational measurement and language testing. *Language Testing in Asia*, 12, Article 3. <https://doi.org/10.1186/s40468-021-00153-2>

- Hsieh, C.-N. (2023). *Evaluating the use and interpretations of the TOEIC Listening and Reading test score report: Perspectives of test takers in Japan*. ETS Research Report Series (RR-23-02). <https://doi.org/10.1002/ets2.12364>
- Hsieh, C.-N. (2023). *Interpretation and use of a workplace English language proficiency test score report: Perspectives of test takers and score users in Taiwan*. ETS Research Report Series (RR-23-10). <https://doi.org/10.1002/ets2.12373>
- Hsieh, C.-N. (2024). Building a validity argument for the TOEFL Junior® tests. ETS Research Report Series, 2024(1), 1–22. <https://doi.org/10.1002/ets2.12379>
- Hu, T & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486-1501, DOI: 10.1080/13670050.2019.1691498
- Huang, B. H., Bailey, A. L., Sass, D. A., & Shawn Chang, Y. (2021). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401–428. <https://doi.org/10.1177/0265532220925731>
- Huang, H. T. D., Hung, S. T. A., Chao, H. Y., Chen, J. H., Lin, T. P., & Shih, C. L. (2022). Developing and validating a computerized adaptive testing system for measuring the English proficiency of Taiwanese EFL university students. *Language Assessment Quarterly*, 19(2), 162-188. <https://doi.org/10.1080/15434303.2021.1984490>
- Huang, J., Shu, T., Dong, Y., & Zhu, D. (2023). Constructing and validating a self-assessment scale for Chinese college English-major students' feedback knowledge repertoire in EFL academic writing: Item response theory and factor analysis approaches. *Assessing Writing*, 56, Article 100716. <https://doi.org/10.1016/j.asw.2023.100716>
- Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276-299. <https://doi.org/10.1177/02655322221112364>
- In'nami, Y., & Koizumi, R. (2021). *Factor structure and four-skill profiles of the Aptis test*. (ARAGs research reports, AR-G/2021/2). British Council. <https://www.britishcouncil.org/factor-structure-and-four-skill-profiles-aptis-test>
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2023). Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach. *Language Learning*, 73(3), 759–798. <https://doi.org/10.1111/lang.12548>
- Inoue, C., & Lam, D. M. K. (2021). *The effects of extended planning time on candidates' performance, processes, and strategy use in the lecture listening-into-speaking tasks of the TOEFL iBT Test*. (TOEFL Research Report, TOEFL-RR-93, ETS Research Report No. RR-21-09). <https://doi.org/10.1002/ets2.12322>
- Inoue, C., & Nakatsuhara, F. (2022). Validation of a large-scale task-based test: Functional progression in dialogic speaking performance. In N. P. Sudharshana & L. Mukhopadhyay (eds.), *Task-based language teaching and assessment: Contemporary reflections from across the world* (pp. 217 - 247). Springer Nature.
- Inoue, C., Khabbazzashi, N., Lam, D. M. K., & Nakatsuhara, F. (2021). Towards new avenues for the IELTS Speaking Test: Insights from examiners' voices. *IELTS Research Reports Online Series*, 2. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/online-series-2021-2.ashx>

- Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, 40(3), 748–770. <https://doi.org/10.1177/02655322231158550>
- Isbell, D. R. (2021). Can the test support student learning? Validating the use of a second language pronunciation diagnostic. *Language Assessment Quarterly*, 18(4), 331–356. <https://doi.org/10.1080/15434303.2021.1874382>
- Isbell, D. R., & Son, Y. A. (2022). Measurement properties of a standardized elicited imitation test: An integrative data analysis. *Studies in Second Language Acquisition*, 44(3), 859–885. <https://doi.org/10.1017/S0272263121000383>
- Isbell, D. R., & Son, Y.-A. (2023). Explanatory item response models for instrument validation: A tutorial based on an elicited imitation test. *Research Methods in Applied Linguistics*, 2(3), Article 100080. <https://doi.org/10.1016/j.rmal.2023.100080>
- Isbell, D. R., Crowther, D., & Nishizawa, H. (2024). Speaking performances, stakeholder perceptions, and test scores: Extrapolating from the Duolingo English test to the university. *Language Testing*, 41(2), 233–262. <https://doi.org/10.1177/02655322231165984>
- Isbell, D. R., Kim, K. M., & Chen, X. (2023). Exploring the potential of automated speech recognition for scoring the Korean Elicited Imitation Test. *Research Methods in Applied Linguistics*, 2(3), Article 100076. <https://doi.org/10.1016/j.rmal.2023.100076>
- Jamalzadeh, M., Lotfi, A. R., & Rostami, M. (2021). Assessing the validity of an IAU General English Achievement Test through hybridizing differential item functioning and differential distractor functioning. *Language Testing in Asia*, 11, Article 8. <https://doi.org/10.1186/s40468-021-00124-7>
- Jung Youn, S. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing*, 40(1), 54–60. <https://doi.org/10.1177/02655322221126606>
- Kamiya, N. (2024). Is the Common Test for University Admissions in Japan enough to measure students' general English proficiency? The case of the TOEIC Bridge. *Language Testing in Asia*, 14, Article 2. <https://doi.org/10.1186/s40468-024-00272-6>
- Karakolidis, A., O'Leary, M., & Scully, D. (2021). Animated videos in assessment: comparing validity evidence from and test-takers' reactions to an animated and a text-based situational judgment test. *International Journal of Testing*, 21(2), 57–79. <https://doi.org/10.1080/15305058.2021.1916505>
- Karatay, Y. (2023). Using spoken dialog systems to assess L2 learners' oral skills in a local language testing context. In Y. Xun, D. Slobodanka & J. G. April (Eds.), *Local language testing: Practice across contexts* (pp, 231–252). Springer. https://link.springer.com/chapter/10.1007/978-3-031-33541-9_12
- Kaya, E., O'Grady, S., & Kalender, İ. (2022). IRT-based classification analysis of an English language reading proficiency subtest. *Language Testing*, 39(4), 541–566. <https://doi.org/10.1177/02655322211068847>
- Kennedy, E., & Shiel, G. (2022). Writing assessment for communities of writers: rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy & Practice*, 29(2), 127–149. <https://doi.org/10.1080/0969594X.2022.2047608>
- Khabbazzbashi, N. (2021). *On topic validity of speaking test*. Cambridge University Press.
- Kho, S. Q. E., Aryadoust, V., & Foo, S. (2023). An eye-tracking investigation of the keyword-matching strategy in listening assessment. *Education and Information*

- Technologies, 28, 3739–3763.
<https://link.springer.com/article/10.1007/s10639-022-11322-y>
- Khodi, A., Ponniah, L. S., Farrokhi, A. H., & Sadeghi, F. (2024). Test review of Iranian English language proficiency test: MSRT test. *Language Testing in Asia*, 14, Article 4. <https://doi.org/10.1186/s40468-023-00270-0>
- Kim, H., Song, C., Kim, J., Jeong, H., & Park, J. (2024). Validation of an elicited Imitation test as a measure of Korean language proficiency. *Language Testing in Asia*, 14, Article 17. <https://doi.org/10.1186/s40468-024-00287-z>
- Kim, M., Nam, Y., & Crossley, S. A. (2022). Roles of working memory, syllogistic inferencing ability, and linguistic knowledge on second language listening comprehension for passages of different lengths. *Language Testing*, 39(4), 593–617.
<https://doi.org/10.1177/02655322211060076>
- Kim, S. (2021). Prepping for the TOEFL iBT Writing test, Gangnam style. *Assessing Writing*, 49, Article 100544. <https://doi.org/10.1016/j.asw.2021.100544>
- Knoch, U., & Fan, J. (2024). Assessing clinician language skills. In Ortega, P., Martínez, G., Lor, M., & Ramírez, A. S. *The Handbook of Language in Public Health and Healthcare*. Wiley.
<https://doi.org/10.1002/9781119853855.ch12>
- Koizumi, R. (2021). The Telephone Standard Speaking Test: An outside evaluator's investigation of a rebuttal to the generalization inference. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of argument-based validation research* (pp. 154–175). Cambridge University Press.
<https://doi.org/10.1017/9781108669849.010>
- Koizumi, R., & Watanabe, A. (2021). Rater reliability in classroom speaking assessment in a Japanese senior high school. *Annual Review of English Language Education in Japan*, 32, 129–144. https://doi.org/10.20581/arele.32.0_129
- Kongsuwanakul, K. (2022). Relating a concordance-based cloze test to the model of communicative language ability: a verbal protocol study. *Language Testing in Asia*, 12, Article 1. <https://doi.org/10.1186/s40468-021-00151-4>
- Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: a meta-analysis. *Studies in Second Language Acquisition*, 44(3), 886–911. <https://doi.org/10.1017/S0272263121000395>
- Kyle, K., Choe, A. T., Eguchi, M., LaFlair, G., & Ziegler, N. (2021). A comparison of spoken and written language use in traditional and technology-mediated learning environments. *ETS Research Report Series*, 2021(1), 1–29.
<https://doi.org/10.1002/ets2.12329>
- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170.
<https://doi.org/10.1080/15434303.2020.1844205>
- Kyle, K., Eguchi, M., Choe, A. T., & LaFlair, G. (2022). Register variation in spoken and written language use across technology-mediated and non-technology-mediated learning environments. *Language Testing*, 39(4), 618–648.
<https://doi.org/10.1177/02655322211057868>
- Latimer, N., & Chan, S. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: Benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Lee, S. (2022). Evaluating the new TOEFL ITP® speaking test: insights from field test

- takers. *ETS Research Report Series*, 2022(1), 1–19.
<https://doi.org/10.1002/ets2.12352>
- Li, S. (2024). Using the Rasch model to evaluate instrument and data quality in assessing requests in L2 Chinese. In S. Li (Ed.), *Pragmatics of Chinese as a second language* (pp. 161–184). Multilingual Matters. <https://doi.org/10.2307/jj.22679704.12>
- Li, X., & Wang, Y. (2024). Effects of input text genre on Chinese young EFL learners' performance on continuation tasks. *Language Testing in Asia*, 14, Article 3.
<https://doi.org/10.1186/s40468-024-00273-5>
- Liao, R. J. T. (2021). Advancing the international recognition of the locally-produced GEPT: An interview with Jessica Wu. *Language Assessment Quarterly*, 18(4), 446–455.
<https://doi.org/10.1080/15434303.2021.1919116>
- Liao, Y. F. (2022). Using the English GSAT for placement into EFL classes: Accuracy and validity concerns. *Language Testing in Asia*, 12, Article 31.
<https://doi.org/10.1186/s40468-022-00181-6>
- Liu, T., Aryadoust, V., & Foo, S. (2022). Examining the factor structure and its replicability across multiple listening test forms: Validity evidence for the Michigan English test. *Language Testing*, 39(1), 142–171. <https://doi.org/10.1177/02655322211018139>
- Lu, Y. (2022). *Assessing learners' competence in L2 Chinese*. Routledge.
- Lukácsi, Z. (2021). Developing a level-specific checklist for assessing EFL writing. *Language Testing*, 38(1), 86–105. <https://doi.org/10.1177/0265532220916703>
- Macqueen, S. (2024). [Review of the book *Reconsidering Context in Language Assessment: Transdisciplinary Perspectives, Social Theories, and Validity* by J. Fox and N. Artemeva]. *Language Testing*, 41(1), 220–226.
<https://doi.org/10.1177/02655322231199501>
- Madyarov, I., Movsisyan, V., Madoyan, H., Galikyan, I., & Gasparyan, R. (2021). New validity evidence on the TOEFL Junior® standard test as a measure of progress. *ETS Research Report Series*, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12334>
- Matta, M., Mercer, S. H., & Keller-Margulis, M. A. (2022). Evaluating validity and bias for hand-calculated and automated written expression curriculum-based measurement scores. *Assessment in Education: Principles, Policy & Practice*, 29(2), 200–218.
<https://doi.org/10.1080/0969594X.2022.2043240>
- McElwee, S., Y.F. Cheung, K., R.T. Cromie, S., Shannon, M., & Gallacher, T. (2021). Revising the BioMedical Admissions Test (BMAT) to improve impact and washback for candidates and support fair access to test preparation. *Assessment in Education: Principles, Policy & Practice*, 28(1), 77–95.
<https://doi.org/10.1080/0969594X.2019.1639133>
- Meng, Y., & Fu, H. (2023). Modeling mediation in the dynamic assessment of listening ability from the cognitive diagnostic perspective. *Modern Language Journal*, 107, 137–160. <https://doi.org/10.1111/modl.12820>
- Menold, N., & Raykov, T. (2022). On the relationship between item stem formulation and criterion validity of multiple-component measuring instruments. *Educational and Psychological Measurement*, 82(2), 356–375.
<https://doi.org/10.1177/0013164420988169>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68, Article 100963.
<https://doi.org/10.1016/j.stueduc.2020.100963>

- Mohammadi, M., Abbasian, G. R., & Siyyari, M. (2022). Adaptation and validation of a critical thinking scale to measure the 3D critical thinking ability of EFL readers. *Language Testing in Asia*, 12, Article 24. <https://doi.org/10.1186/s40468-022-00173-6>
- Monfils, L. F., & Manna, V. F. (2021). Time to achieving a designated criterion score level: A survival analysis study of test taker performance on the TOEFL iBT® test. *Language Testing*, 38(1), 154–176. <https://doi.org/10.1177/0265532220940709>
- Montasseri, Z., & Ahmadi, A. (2024). Validity argument of an Interactional Competence checklist: Perspectives and many-facets Rasch analysis. *Studies in Language Assessment (SILA)*, 13(1), 69–107. <https://doi.org/10.58379/MEZB3039>
- Nakanishi, Y., & Takeuchi, O. (2024). Validating the foreign language enjoyment scale for young learners: An exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100167. <https://doi.org/10.1016/j.rmal.2024.100167>
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2021). Video-conferencing speaking tests: Do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice*, 28(4), 369–388. <https://doi.org/10.1080/0969594X.2021.1951163>
- Nayernia, A., Nosrati, R., & Mohebbi, H. (2022). The factors contributing to language teachers' effectiveness in an EFL learning context: A questionnaire validation study. *Profile: Issues in Teachers' Professional Development*, 24(2), 63–79. <https://doi.org/10.15446/profile.v24n2.93571>
- Nguyen, T. H. H., Nguyen, B. T. T., Hoang, G. T. L., Pham, N. T. H., & Dang, T. T. C. (2024). Computer-delivered vs. face-to-face score comparability and test takers' perceptions: The case of the two English speaking proficiency tests for Vietnamese EFL learners. *Language Testing in Asia*, 14, Article 6. <https://doi.org/10.1186/s40468-024-00277-1>
- Nguyen, T. M. H., Gu, P., & Coxhead, A. (2024). Argument-based validation of Academic Collocation Tests. *Language Testing*, 41(3), 459–505. <https://doi.org/10.1177/02655322231198499>
- Nikmard, F., & Tavassoli, K. (2021). A three-parameter logistic IRT calibration of the items of the TEFL MA admission test as a high-stakes test in Iran. *Iranian Journal of Applied Linguistics (IJAL)*, 24(2), 135–161. <http://ijal.khu.ac.ir/article-1-3173-en.html>
- Nikmard, F., Tavassoli, K., & Pourdana, N. (2023). Designing and validating a scale for evaluating the sources of unreliability of a high-stakes test. *Language Testing in Asia*, 13, Article 2 (2023). <https://doi.org/10.1186/s40468-023-00215-7>
- Nishizawa, H., Isbell, D. R., & Suzuki, Y. (2022). Review of the Japanese-language proficiency test. *Language Testing*, 39(3), 494–503. <https://doi.org/10.1177/02655322221080898>
- Noori, M. (2022). Road not taken" in language testing: Sociocultural implications of test and teaching contents. *TESOL Quarterly*, 56(4), 1486–1503. <https://doi.org/10.1002/tesq.3189>
- Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of IELTS preparation materials from a critical perspective. *Language Assessment Quarterly*, 18(4), 382–397. <https://doi.org/10.1080/15434303.2021.1883618>
- Noroozi, S., & Karami, H. (2024). A Rasch-based validation of the University of Tehran English Proficiency Test (UTEPT). *Language Testing in Asia*, 14, Article 18.

- <https://doi.org/10.1186/s40468-024-00290-4>
- Nour, P., Esfandiari, R., & Zarei, A. A. (2021). Development and validation of a metamemory maturity questionnaire in the context of English as a foreign language. *Language Testing in Asia*, 11, Article 24. <https://doi.org/10.1186/s40468-021-00141-6>
- O'Sullivan, B., & Chalhoub-Deville, M. (2021). Language testing for migrants: Co-constructing validation. *Language Assessment Quarterly*, 18(5), 547–557. <https://doi.org/10.1080/15434303.2021.1986513>
- O'Sullivan, B., & Chalhoub-Deville, M. (2023). Validity: An integrated arguments approach. In: Dobric, N., Cesnik, H., and Harsch, C (Eds). *Festschrift in honor of Günther Sigott: Advanced Methods in Language Testing*. Peter Lang.
- O'Sullivan, B., & Cheng, L. (2022). Lessons from the Chinese imperial examination system. *Language Testing in Asia*, 12, Article 52. <https://doi.org/10.1186/s40468-022-00201-5>
- O'Sullivan, B., Breakspear, T., & Bayliss, W. (2023). Validating an AI-Driven Scoring System: The Model Card Approach. In Karim Sadeghi & Dan Douglas (eds), *Fundamental Considerations in Technology Mediated Language Assessment*, (pp.115-134). New York: Routledge. DOI: 10.4324/9781003292395-10
- Ockey, G. J., & Neiriz, R. (2021). Evaluating technology-mediated second language oral communication assessment delivery models. *Assessment in Education: Principles, Policy & Practice*, 28(4), 350–368. <https://doi.org/10.1080/0969594X.2021.1976106>
- Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, 40(3), 521-547. <https://doi.org/10.1177/02655322221139162>
- Owen, N., Shrestha, P. N., & Hultgren, A. K. (2021). Researching academic reading in two contrasting English as a medium of instruction contexts at a university level. *ETS Research Report Series*, 2021(1), 1–28. <https://doi.org/10.1002/ets2.12317>
- Owen, N., Shrestha, P., & Bax, S. (2021). *Researching lexical thresholds and lexical profiles across the Common European Framework of References for Languages (CEFR) levels assessed in the Aptis test* (Report No. AR-G/2021/1). British Council. https://www.britishcouncil.org/sites/default/files/owen_shrestha_and_bax_layout.pdf
- Pan, R., & Lu, X. (2023). The design and cognitive validity verification of reading-to-write tasks in L2 Chinese writing assessment. *Assessing Writing*, 56, Article 100699. <https://doi.org/10.1016/j.asw.2023.100699>
- Park, Y., Lee, S., & Shin, S.-Y. (2022). Developing a local academic English listening test using authentic unscripted audio-visual texts. *Language Testing*, 39(3), 401–424. <https://doi.org/10.1177/02655322221076024>
- Pearson, W. S. (2022). Test Review: LanguageCert IESOL B1 (Achiever) SELT. *Language Assessment Quarterly*, 19(5), 526-537. <https://doi.org/10.1080/15434303.2022.2103420>
- Pelleriti, M. (2023). Item writing and dyslexic students: Promoting best practices. *Iperstoria*, 21. 217–231. <https://iperstoria.it/article/view/1285/1315>
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang H. (2023). Predicting Chinese EFL Learners' Human-rated Writing Quality in Argumentative Writing Through Multidimensional Computational Indices of Lexical Complexity. *Assessing Writing*, 56, Article 100722. <https://doi.org/10.1016/j.asw.2023.100722>
- Phakiti, A., & Isaacs, T. (2021). Classroom assessment and validity: Psychometric and

- edumetric approaches. *The European Journal of Applied Linguistics and TEFL*, 10(1), 3–21.
- Prabowo, M. Y. (2022). Assessing equivalency between paper-and-pencil and computer-based English competency test. In R. E. Capel & Q. S. Hoon (Eds.), *Conference proceedings iELT-CON2021* (pp. 42–51). Penang English Language Learning and Teaching Association.
https://pellta.org/uploads/8/1/8/9/81899300/ielt-con_2021_proceedings.pdf
- Pusey, K., & Butler, Y. G. (2023). Investigating the ecological validity of second language writing assessment tasks. *System*, 119, Article 103174.
<https://doi.org/10.1016/j.system.2023.103174>
- QUEVEDO-CAMARGO, G. & SANTOS, Gabriela M. (2020) Validade de conteúdo das provas de um curso de compreensão oral para fins acadêmicos: relato de uma experiência [Content validity of tests in a listening comprehension course for academic purposes: report of an experience]. *Revista Olhares e Trilhas*, v. 22, p. 289-308. DOI: <https://doi.org/10.14393/OT2020v22.n.2.55520>
- Rafatbakhsh, E., Ahmadi, A., Moloodi, A., & Mehrpour, S. (2021). Development and validation of an automatic item generation system for English idioms. *Educational Measurement: Issues and Practice*, 40(2), 49–59.
<https://doi.org/10.1111/emip.12401>
- Randall, J., Poe, M., Slomp, D., & Oliveri, M. E. (2024). Our validity looks like justice. Does yours? *Language Testing*, 41(1), 203–219.
<https://doi.org/10.1177/02655322231202947>
- Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2024). Classroom Assessment Fairness Inventory: A New Instrument to Support Fairness in Classroom Assessment. *Assessment in Education*, 30-5-6. <https://doi.org/10.1080/0969594X.2023.2255936>
- Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education: Principles, Policy & Practice*, 29(3), 383-394.
<https://doi.org/10.1080/0969594X.2022.2103516>
- Read, J. (2022). Test Review: The International English Language Testing System (IELTS). *Language Testing*, 39(4), 679–694. <https://doi.org/10.1177/02655322221086211>
- Révész, A., & Brunfaut, T. (2021). Validating assessments for research purposes. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 21–32). Routledge.
<https://doi.org/10.4324/9781351034784-4>
- Robles-García, P., McLean, S., Stewart, J., Shin, J. young, & Sánchez-Gutiérrez, C. H. (2024). The Development and Initial Validation of O-WSVLT, a Meaning-Recall Online L2 Spanish Vocabulary Levels Test. *Language Assessment Quarterly*, 21(2), 181–205. <https://doi.org/10.1080/15434303.2024.2311724>
- Roohr, K., Olivera-Aguilar, M., Bochenek, J., & Belur, V. (2022), Exploring GRE® and TOEFL® Score Profiles of International Students Intending to Pursue a Graduate Degree in the United States. *ETS Research Report Series*, 2022(1), 1–27.
<https://doi.org/10.1002/ets2.12343>
- Russell, B., Barron, C., Kim, H., & Jang, E. E. (2022). A mixed-method investigation into international university students' experience with academic language demands. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2022.934692>
- Sadoughi, M., & Hejazi, S. Y. (2022). Development and validation of Foreign Language Teacher Support Scale (FLTSS). *Language Testing in Asia*, 12, Article 30.

- <https://doi.org/10.1186/s40468-022-00186-1>
- Saito, H., Sawaki, Y., & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and use for Japanese university admission. *Language Assessment Quarterly*, 19(4), 422-448.
<https://doi.org/10.1080/15434303.2022.2083965>
- Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 46(4), 1206–1230. <https://doi.org/10.1017/S027226312200047X>
- Saito, K., Macmillan, K., Kachlicka, M., Kuniyara, T., & Minematsu, N. (2023). Automated assessment of second language comprehensibility: Review, training, validation, and generalization studies. *Studies in Second Language Acquisition*, 45(1), 234-263.
<https://doi.org/10.1017/S0272263122000080>
- Salaberry, M. R. & Weideman, A. (2024). Context, construct and ethics. In M. R., Salaberry, A. Weideman, A. & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 3-27). Routledge.
<https://doi.org/10.4324/9781003384922-2>
- Sawaki, Y. (2024). [Review of the book *Validity Argument in Language Testing: Case Studies of Validation Research* by C. A. Chapelle and E. Voss]. *Language Testing*, 41(1), 227–229. SAGE Publications. <https://doi.org/10.1177/02655322231193705>
- Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, 14, Article 50. <https://doi.org/10.1186/s40468-024-00325-w>
- Sayyadi, A., & Rezvani, R. (2021). Questioning in TOEFL iBT speaking test: A case of washback and construct underrepresentation. *Language Testing in Asia*, 11, Article 21. <https://doi.org/10.1186/s40468-021-00137-2>
- Schmidgall, J., & Powers, D. E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC® Speaking test scores from linguistic laypersons. *Language Testing*, 38(2), 302–325.
<https://doi.org/10.1177/0265532220941803>
- Schmidgall, J., Cid, J., Carter Grissom, E., & Li, L. (2021). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge® tests. *ETS Research Report Series*, 2021(1), 1–22.
<https://doi.org/10.1002/ets2.12335>
- Schmitt, Romain & Saif, Shahrzad. 2022. L'utilisation du TFLM comme mesure de compétence langagière des candidats aux programmes de formation initiale en enseignement au Québec est-elle justifiée? Un argumentaire d'utilisation des évaluations. In N. Talbot et S. Eleuch (Eds.), *L'évaluation des compétences: une pluralité de défis*, Éditions JFD.
- Shafipoor, M., Ravand, H., & Maftoon, P. (2021). Test-level and item-level model fit comparison of general vs. specific diagnostic classification models: A case of true DCM. *Language Testing in Asia*, 11, Article 33.
<https://doi.org/10.1186/s40468-021-00148-z>
- Shahmirzadi, N. (2023). Validation of a language center placement test: Differential item functioning. *International Journal of Language Testing*, 13(1), 1-17.
<https://doi.org/10.22034/IJLT.2022.336779.1151>

- Shahmirzadi, N. (2024). Simulation study: Seeking fairness in a high-stakes placement test. *Language Testing in Focus: An International Journal*, 10, 18–28, (e-ISSN: 2717-9087). <https://doi.org/10.32038/ltf.2024.10.02>
- Shahmirzadi, N., & Marashi, H. (2023). Cognitive diagnostic assessment of reading comprehension for high-stakes tests: Using GDINA model. *Language Testing in Focus: An International Journal*, (8), 1–16. <https://doi.org/10.32038/ltf.2023.08.01>
- Shahmirzadi, N., Siyyari, M., Marashi, H., & Gramipour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA. *Journal of Foreign Languages Research*, 10(1), 152–165, (ISC), (ISSN: 2588–4123, Online ISSN: 2588–7521). DOI: 10.22059/JFLR.2019.282328.634
- Shi, H., & Aryadoust, V. (2023). A systematic review of automated writing evaluation within the argument-based validation framework. *Education and Information Technology*, 28, 771–795. <https://link.springer.com/article/10.1007/s10639-022-11200-7>
- Shin, D. (2024). Critical discursive approaches to evaluating policy-driven testing: Social impact as a target for validation. *Language Testing*, 41(1), 162–180. <https://doi.org/10.1177/02655322231163863>
- Shin, J., Rodríguez-Fuentes, R. A., Swatek, A. M., & Ginther, A. (2022). Aptis test review. *Language Testing*, 39(1), 172–187. <https://doi.org/10.1177/02655322211032873>
- Shoahosseini, R., Baghaei, P., Khodabakhshzadeh, H., & Ashraf, H. (2024). C-Test construct validity: Evidence from nonparametric item response theory. *Language Testing in Asia*, 14, Article 10. <https://doi.org/10.1186/s40468-024-00279-z>
- Skaggs, G. (2021). *Test development and validation*. SAGE Publications.
- Solon, M., Park, H. I., Dehghan-Chaleshtori, M., Carver, C., & Long, A. Y. (2022). Exploring an elicited imitation task as a measure of heritage language proficiency. *Studies in Second Language Acquisition*, 44(4), 1095–1123. <https://doi.org/10.1017/S0272263121000905>
- Stansfield, C.W., Gao, J., & Rivers, W.P. (2010). A concurrent validity study of self-assessments and the federal interagency roundtable oral proficiency interview. *Russian Language Journal / Русский Язык*, 60, 299–315. <http://www.jstor.org/stable/43669189>
- Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, 43(2), 462–471. <https://doi.org/10.1017/S0272263121000437>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>
- Su, X., & Lee, I. (2024). Is the magic in the mix? The development and validation of the L2 teachers' blended assessment literacy scale. *System*, 123, 103354. <https://doi.org/10.1016/j.system.2024.103354>
- Su, Y., & Chen, X. (2024). Examining refusal and acceptance sequences employing a data-driven binary rating approach. *Research Methods in Applied Linguistics*, 3(3), Article 100161. <https://doi.org/10.1016/j.rmal.2024.100161>
- Suárez, M. M. (2022). A need for an aptitude test for young learners in Catalan: The case of the Modern Language Aptitude Test - Elementary in Catalan. *Language Teaching Research Quarterly*, 31, 101–118. <https://doi.org/10.32038/ltrq.2022.31.08>

- Sudina, E., Vernon, T., Foster, H., Del Villano, H., Hernandez, S., Beck, D., & Plonsky, L. (2021). Development and initial validation of the L2-teacher grit scale. *TESOL Quarterly*, 55(1), 156–184. <https://doi.org/10.1002/tesq.581>
- Sureeyatanapas, P., Sureeyatanapas, P., Panitanarak, U., Kraisriwattana, J., Sarootyanapat, P., & O'Connell, D. (2024). The analysis of marking reliability through the approach of gauge repeatability and reproducibility (GR&R) study: A case of English-speaking test. *Language Testing in Asia*, 14, Article 1. <https://doi.org/10.1186/s40468-023-00271-z>
- Suto, I., Williamson, J., Ireland, J., & Macinska, S. (2023). On reducing errors in assessment instruments. *Research Papers in Education*, 38(3), 357-377. <https://doi.org/10.1080/02671522.2021.1968940>
- Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, 43(3), 663–676. <https://doi.org/10.1017/S0272263120000704>
- Suzumura, N. (2022). Content analysis of test taker responses on an AP Japanese computer-simulated conversation test: A mixed methods approach for a validity argument. *Language Assessment Quarterly*, 19(5), 480-502. <https://doi.org/10.1080/15434303.2022.2130326>
- Syquia, J., & Leeming, P. (2024). Assessing interactional competence through group discussion: A mixed methods validation. *Research Methods in Applied Linguistics*, 3(3), Article 100144. <https://doi.org/10.1016/j.rmal.2024.100144>
- Tai, H., Liu, C., & Chen, Y. (2023). Validation of a multiple-choice implicature test: insights from Chinese EFL learners' cognitive processes. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1481-1506. <https://doi.org/10.1515/iral-2021-0190>
- Tan, L. Y., McLean, S., Kim, Y. A., & Vitta, J. P. (2024). Rasch modelling vs. item facility: Implications on the validity of assessments of Asian EFL/ESL vocabulary knowledge and lexical sophistication modelling. *Language Testing in Asia*, 14, Article 55. <https://doi.org/10.1186/s40468-024-00327-8>
- Tan, T. X., Fan, X., Braunstein, L. B., & Lane-Holbert, M. (2022). Linguistic, cultural and substantive patterns in L2 writing: A qualitative illustration of MisLevy's sociocognitive perspective on assessment. *Assessing Writing*, 51, Article 100574. <https://doi.org/10.1016/j.asw.2021.100574>
- Taylor, L., & Banerjee, J. (2023). Language assessment accommodations: Issues and challenges for the future. *Language Testing*, 40(4), 1000-1006. <https://doi.org/10.1177/02655322231186222>
- Teng, M. F., & Teng, L. S. (2024). Validating the multi-dimensional structure of self-efficacy beliefs in peer feedback for L2 writing: A bifactor-exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100136. <https://doi.org/10.1016/j.rmal.2024.100136>
- Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, 5, 1-15. <https://doi.org/10.22599/jesla.74>
- Tsagari, D., & Demetriou, T. (2022). Investigating the parameters accounting for language proficiency: The case of an English language test used for academic purposes. *Research Papers in Language Teaching and Learning*, 12(1), 89-110.

- https://rpltl.eap.gr/images/2022/01_loannou__Tsagari.pdf
- Vanbuel, M., & Deygers, B. (2024). Assessing Receptive Skills in an Adult L2 Learner Population with Diverse Educational Backgrounds. *Language Assessment Quarterly*, 21(4–5), 379–403. <https://doi.org/10.1080/15434303.2024.2438151>
- Vogt, K. & Tsagari, D. (2022). Language assessment literacy and test validation: Highlighting the role of language teachers. *Studies in Language Assessment*, 11(1), 1–15. https://arts.unimelb.edu.au/__data/assets/pdf_file/0006/4282557/SiLA-11.1-Vogt_and_Tsagari.pdf
- Wang, P. & Cheng, L. (2024). The Impact of TOEFL iBT Test Preparation on Chinese Test-Takers' Perceived Test Value. *Critical Inquiry in Language Studies*.
- Wang, S. (2021). *Consequential validity of the Chinese Proficiency Test (HSK) from macro and micro perspectives*. Cambridge Scholars Publishing.
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, Article 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/S0272263121000449>
- Weideman, A. (2022). Context, construct, and validation: A perspective from South Africa. *Language Assessment Quarterly*, 19(2), 124–141. <https://doi.org/10.1080/15434303.2020.1860991>
- Weideman, A. (2024). Yardsticks for the future of language assessment: Disclosing the meaning of measurement. In M. R. Salaberry, A. Weideman, A., & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 220-234). Routledge. <https://doi.org/10.4324/9781003384922-12>
- Weideman, A., & Deygers, B. (2024). Validity and validation: An alternative perspective. In M. R. Salaberry, A. Weideman, & W-L. Hsu.(Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 28-50). Routledge. <https://doi.org/10.4324/9781003384922-3>
- Weiler, T. (2021). *Testing lexicogrammar: An investigation into the construct tested in the Language in Use section of the Austrian Matura in English*. Peter Lang.
- Weir, C. & Chan, S. (2019). *Research and practice in assessing academic English: The case of IELTS*. Cambridge University Press.
- Weisi, H., & Zandi, M. (2024). A mixed-method concurrent transformative study of metadiscourse markers employed by L2 speakers: Does proficiency level matter? *Language Testing in Asia*, 14, Article 60. <https://doi.org/10.1186/s40468-024-00330-z>
- Williams, K. (2024). Exploring English for academic purposes instructors' perceptions of speech fluency through developing and piloting a rating scale for a paired conversational task. *System*, 121, 103266. <https://doi.org/10.1016/j.system.2024.103266>
- Wilsenach, C., & Schaefer, M. (2022). Development and initial validation of productive vocabulary tests for isiZulu, Siswati and English in South Africa. *Language Testing*, 39(4), 567–592. <https://doi.org/10.1177/02655322211063785>
- Winke, P., & Brunfaut, T. (2021). Perspectives on “knowing” a second language: What are we seeking to measure? In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.1–8). Routledge.

- <https://doi.org/10.4324/9781351034784-1>
- Winke, P., & Brunfaut, T. (Eds.). (2021). *The Routledge handbook of second language acquisition and language testing*. Routledge.
<https://doi.org/10.4324/9781351034784>
- Wolf, M. K., Bailey, A. L., & Ballard, L. (2023). Aligning English language proficiency assessments to standards: Conceptual and technical issues. *TESOL Quarterly*, 57(2), 670-685. <https://doi.org/10.1002/tesq.3199>
- Wolf, M. K., Yoo, H., & Ballard, L. (2024). The relationship between English language proficiency and academic performance among English learners: A subgroup analysis by home language background. *Learning and Individual Differences*, 109, 102388. <https://doi.org/10.1016/j.lindif.2023.102388>
- Wolf, M. K., Yoo, H., Guzman-Orth, D., & Abedi, J. (2022). Investigating the effects of test accommodations with process data for English learners in a mathematics assessment. *Educational Assessment*, 27(1), 27-45.
<https://doi.org/10.1080/10627197.2021.1982693>
- Wu, M. Y., Steinkrauss, R., & Lowie, W. (2023). The reliability of single task assessment in longitudinal L2 writing research. *Journal of Second Language Writing*, 59, 100950.
<https://doi.org/10.1016/j.jslw.2022.100950>
- Wu, S. L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, 44(1), 271-300. <https://doi.org/10.1017/S0272263121000103>
- Xie, Q. (2024). Construct representation and predictive validity of integrated writing tasks: A study on the writing component of the Duolingo English Test. *Assessing Writing*, 61, 100846. <https://doi.org/10.1016/j.asw.2024.100846>
- Yan, X., Chuang, P., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment (SILA)*, 13(1), 108-140. <https://doi.org/10.58379/DLDI3148>
- Yang, X., & Osborne, C. (2023). The development and validation of a C-test and a pseudo-character test for online CFL assessment. *Journal of China Computer-Assisted Language Learning*, 3(1), 101-131.
<https://doi.org/10.1515/jccall-2022-0019>
- Yang, Z., & Yu, C.* (2024). Is the use of multiple-choice items and a holistically-scored paragraph translation task fair? Examining a large-scale translation subtest. *Asia Pacific Education Review*. Online First. <https://doi.org/10.1007/s12564-024-09993-y>
- Yeager, R. R., Park, G., & Liao, R. J. T. (2024). Notetaking as Validity Evidence: A Mixed-Methods Investigation of Question Preview in Listening Assessment. *Journal of English for Academic Purposes*, 68(3), 101346.
<https://doi.org/10.1016/j.jeap.2024.101346>
- Yeager, R., Park, G., & Liao, R. J. (2024). Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment. *Journal of English for Academic Purposes*, 68, Article 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Yeom, S., & Llosa, L. (2024). Comparability of reading tasks in high-stakes English proficiency tests and university courses in Korea. *Language Testing in Asia*, 14, Article 8. <https://doi.org/10.1186/s40468-024-00281-5>
- Yu, G., & Xu, J. (Eds.). (2024). *Language test validation in a digital age*. Cambridge University Press & Assessment.
- Yu, S., Zhang, E. D., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A

- scale development and validation study. *Assessing Writing*, 53, Article 100643. <https://doi.org/10.1016/j.asw.2022.100643>
- Yu, W., & Iwashita, N. (2021). Comparison of test performance on paper-based testing. (PBT). and computer-based testing. (CBT). by English-majored undergraduate students in China. *Language Testing in Asia*, 11, Article 32. <https://doi.org/10.1186/s40468-021-00147-0>
- Zakaria, A., & Aryadoust, V. (2024). From colonialism to self-determination: A historical study of language testing and institutional policy. *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amae057>
- Zhang, C., Xu, X., & Zhang, Y. (2023). A communicative validation study on an English listening test in Korea. *Language Testing in Asia*, 13, Article 26 (2023). <https://doi.org/10.1186/s40468-023-00238-0>
- Zhang, L. (2022). College English Test–Spoken English Test (CET–SET). *Studies in Language Assessment*, 11(2), 164–180. https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4396265/SiLA_11.2_Test-review.pdf
- Zhang, L., & Jin, Y. (2021). Assessing interactional competence in the computer-based CET–SET: An investigation of the use of communication strategies. *Assessment in Education: Principles, Policy & Practice*, 28(4), 389–410. <https://doi.org/10.1080/0969594X.2021.1976107>
- Zhang, L., & Zhang, L. (2023). Development and validation of the questionnaire on EFL students’ perceptions of authorial stance in academic writing. *Applied Linguistics Review*, 14(5), 1219–1251. <https://doi.org/10.1515/applirev-2021-0003>
- Zhang, P., Fan, J., & Jia, W. (2021). A bibliographic analysis of hotspots and trends in international language testing research. (2008–2018). *Foreign Language Teaching and Research*, 53(4), 618–627.
- Zhang, P., Fan, J., & Jia, W. (2022). Assessing L2 integrated writing self-efficacy: Development and validation of a scale. *Assessing Writing*, 54, Article 100665. <https://doi.org/10.1016/j.asw.2022.100665>
- Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. *Educational Science*, 12(10), Article 723. <https://doi.org/10.3390/educsci12100723>

Test Design and Development

- Al Lawati, Z. A. (2023). Investigating the characteristics of language test specifications and item writer guidelines, and their effect on item development: A mixed-method case study. *Language Testing in Asia*, 13, Article 21. <https://doi.org/10.1186/s40468-023-00233-5>
- Almukhaild, H., & King, J. (2024). Scaling the apex of engagement: Anxious language learners’ positive flow experiences during speaking tasks. *System*, 124, 103373. <https://doi.org/10.1016/j.system.2024.103373>
- Altherr Flores, J. A. (2021). The interplay of text and image on the meaning-making processes of adult L2 learners with emerging literacy: Implications for test design and evaluation frameworks. *Language Assessment Quarterly*, 18(5), 508–529. <https://doi.org/10.1080/15434303.2021.1984491>

- Amjadi, A., & Talebi, S. H. (2024). Extending the efficacy of Collaborative Strategic Reading (CSR) in teaching reading strategies: A socio-cultural emotional perspective. *Language Teaching Research*, 28(2), 714–748. <https://doi.org/10.1177/13621688211011501>
- Baker, B., Arias, A., Bibeau, L. D., Qin, Y., Norenberg, M., & St-John, J. (2023). Rethinking student placement to enhance efficiency and student agency. *Language Testing*, 41(1), 181–191. <https://doi.org/10.1177/02655322231179128>
- Baker, B., Arias, A., Bibeau, L.-D., Qin, Y. (Coral), Norenberg, M., & St-John, J. (2024). Rethinking student placement to enhance efficiency and student agency. *Language Testing*, 41(1), 181–191. <https://doi.org/10.1177/02655322231179128>
- Banerjee, J., Lestari, S. B., & Rossi, O. (2021). Choosing test formats and task types. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.78–89). Routledge. <https://doi.org/10.4324/9781351034784>
- Binti Hasrol, S., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), Article 100023. <https://doi.org/10.1016/j.rmal.2022.100023>
- Bowles, M. A. (2022). Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, 39(3), 355–376. <https://doi.org/10.1177/02655322221076033>
- Brown, J. D., Panahi, A., & Mohebbi, H. (2023). James Dean Brown's 50 years of work in language testing: A systematic review. *Language Teaching Research Quarterly*, 37, 4–75. <https://doi.org/10.32038/ltrq.2023.37.02>
- Browne, E. (2023). Making room for learning and assessing together in bilingual, bicultural teaching teams at a remote Warlpiri school in Central Australia. *Studies in Language Assessment*, 12(2), 151-167. <https://doi.org/10.58379/JKSA5503>
- Chang, Y.-T., Choe, A. T., Holden, D., & Isbell, D. R. (2024). Making each point count: Revising a local adaptation of the Jacobs et al.'s (1981) ESL COMPOSITION PROFILE rubric. *Language Testing*, 41(2), 443–455. <https://doi.org/10.1177/02655322231217979>
- Choi, Y., Vo, S., & Ockey, G. J. (2024). Investigation into the factor structure of a local English placement test of oral communication. *International Journal of Applied Linguistics*, 34(1), 63-85. <https://doi.org/10.1111/ijal.12478>
- Clark, T., & Yu, G. (2024). The pedagogical remit of test preparation: the case of writing acquisition on an IELTS course. *Applied Linguistics Review*, 15(3), 1009–1045. <https://doi.org/10.1515/applirev-2021-0164>
- Clark, T., Holland, M., & Spiby, R. (2023). Seeking empirical evidence to support online test validation: Building on the IELTS indicator assessment model. In K. Sadeghi (Ed), *Technology-assisted language assessment in diverse contexts* (pp. 16-33). Routledge.
- Clark, T., Spiby, R., & Tasviri, R. (2021). Crisis, collaboration, recovery: IELTS and COVID-19. *Language Assessment Quarterly*, 18(1), 17–25. <https://doi.org/10.1080/15434303.2020.1866575>
- Dai, D. W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test*. Peter Lang. <https://doi.org/10.3726/b21295>
- Davis, L., & Norris, J. (2021). Developing an innovative elicited imitation task for efficient English proficiency assessment. *ETS Research Report Series*, 2021(1), 1–30.

- <https://doi.org/10.1002/ets2.12338>
- De Jong, J. H.A.L. (2022) The Action-oriented approach and language testing: A critical view. In D. Little & N. Figueras (Eds.), *Reflecting on the Common European Framework of reference for languages and its companion volume* (pp. 53–63). Multilingual Matters. <https://doi.org/10.21832/9781800410206-009>
- De Silva, R. (2021). Assessment literacy and assessment practices of teachers of English in a South-Asian context: Issues and possible washback. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world* (pp. 433–446). Springer. https://doi.org/10.1007/978-981-33-4232-3_31
- Devos, N., Nizonkiza, D., & Lynch, S. (2024). Developing and validating a post-admission screening-diagnostic assessment procedure to offer language support in college diploma programs. *Canadian Journal of Applied Linguistics*, 27(1), 51-77. <https://doi.org/10.37213/cjal.2024.33259>
- Dimitrov, D. M., & Atanasov, D. V. (2021). An approach to test equating under the latent D-scoring method. *Measurement: Interdisciplinary Research and Perspectives*, 19(3), 153–162. <https://doi.org/10.1080/15366367.2020.1843107>
- Drackert, A., Zimmermann, S., Marks, D. & Korflür, D. (2023). New test format – new research agenda: An overview of the technology-related research at g.a.s.t. In: *Conference Proceedings of the ALTE 8th International Conference: Language Assessment Fit for the Future* (pp. 34–38).
- Drennan, L. (2022). Transitioning from undergraduate to postgraduate studies : A multistage evaluation of a discipline-specific writing intervention. *Journal for Language Teaching*, 56(2), 1-24. <https://doi.org/10.56285/jltVol56iss2a5293>
- Drennan, L., Joubert, M., Weideman, A. & Posthumus, R. (2021). Combined measures of identifying language risk for first-year students. *Journal for Language Teaching*, 55(2), 93-116. <https://doi.org/10.4314/jlt.v55i2.4>
- Du Plessis, C. (2021): Basic education and academic literacy: conflicting constructs in the South African National Senior Certificate (NSC) Language Examination. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 95-116). Multilingual Matters.
- Dunlea, J., O’Sullivan, B., Dunn, K., Felice, M., Lightfoot, A., Patel, M., Spiby, R., & Westbrook., C. (2024). Assessing English as an international language. In: Galloway, N. & Fuad Selvi, A. (Eds.). *Routledge Handbook of English as an International Language* (pp.184-198), Routledge. DOI: 10.4324/9781003203261-16
- Erickson, G., & Tholin, J. (2022). Overall, a good test, but...—Swedish lower secondary teachers’ perceptions and use of national test results of English. *Languages*, 7(1), Article 64. <https://doi.org/10.3390/languages7010064>
- Fan, J., Frost, K., & Jin, Y. (2022). Local English testing in China’s tertiary education: Contexts, policies, and practices. *Language Testing*, 39(3), 453–473. <https://doi.org/10.1177/02655322211070839>
- Fechter, T. M., & Yoon, H. (2024). Evaluating methodological enhancements to the Yes/No Angoff standard-setting method in language proficiency assessment. *Language Testing*, 41(3), 530–556. <https://doi.org/10.1177/02655322231222600>
- Feng, K.-Z. (2024). Exploration of vocational high school students experiencing difficulty in cloze test performances: A mixed-methods study in Taiwan. *Language Testing in Asia*, 14, Article 16. <https://doi.org/10.1186/s40468-024-00274-4>

- Fox, J., Abdulhamid, N. & Turner, C.E. (2022). Classroom based assessment. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing*, 2nd ed. (pp. 119-135). Routledge, Taylor & Francis Group.
<https://doi.org/10.4324/9781003220756-6>
- García Laborda, J., & Fernández Álvarez, M. (2021). Multilevel Language tests: Walking into the land of the unexplored. *Language Learning & Technology*, 25(2), 1–25.
<http://hdl.handle.net/10125/73428>
- Getman, E. P. (2024). Speaking test tasks for young language learners: Development and validation. In L. Davis & J. M. Norris (Eds.), *Challenges and Innovations in Speaking Assessment*. Routledge. <https://doi.org/10.4324/9781315201504-7>
- Ghamoushi, M., Zenouzagh, Z. M., & Hashamdar, M. (2022). Development and validation of a potential assessment inventory for assessing EFL teachers' ecological agency. *Language Testing in Asia*, 12, Article 37. <https://doi.org/10.1186/s40468-022-00190-5>
- Ghiasvand, F., Jahanbakhsh, A. A. & Sharifpour, P. (2023). Designing and validating an assessment agency questionnaire for EFL teachers: an ecological perspective. *Language Testing in Asia*, 13, Article 41.
<https://doi.org/10.1186/s40468-023-00255-z>
- Grant, R., Miller, G. A., Design, Y. C., & Cook, G. (2023). A culturally sustaining and valid Alaska Native Language (Yugtun) assessment for school children. *Studies in Language Assessment*, 12(2), 205–234. <https://doi.org/10.58379/LRBE1549>
- Green, B. A., & Lung, Y. S. M. (2021). English language placement testing at BYU-Hawaii in the time of COVID-19. *Language Assessment Quarterly*, 18(1), 6–11.
<https://doi.org/10.1080/15434303.2020.1863966>
- Green, J., Weideman, A., Davis, C., Judith, K., & Harmes, M. (2024). Using a five-phase applied linguistics design to develop a contextualized academic literacy placement test for pre-university pathway students. *Literacy Research and Instruction*, 1-27.
<https://doi.org/10.1080/19388071.2024.2340031>
- Hirai, A., Oka, H., Kato, T., & Maeda, H. (2022). Development and validation of an English test measuring EFL learners' critical thinking skills. *Language Testing in Asia*, 12, Article 45. <https://doi.org/10.1186/s40468-022-00193-2>
- Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101, Article 102569.
<https://doi.org/10.1016/j.system.2021.102569>
- Holmström, I., Schönström, K., & Ryttervik, M. (2023). Development of a Sign Repetition Task for Novice L2 Signers. *Language Assessment Quarterly*, 21(1), 33–59.
<https://doi.org/10.1080/15434303.2023.2256320>
- Holzknacht, F., & Harding, L. (2024). Repeating the listening text: effects on listener performance, metacognitive strategy use, and anxiety. *TESOL Quarterly*, 58(1), 451–478. <https://doi.org/10.1002/tesq.3249>
- Holzknacht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiiby, R., & Dunlea, J. (2021). The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test. *Language Testing*, 38(1), 41–61.
<https://doi.org/10.1177/0265532220917316>
- Hooft, H., Schiepers, M., & Vandommele, G. (2021). Developing and validating a multilingual literacy test for asylum seekers. *Language Assessment Quarterly*, 18(5), 530–546. <https://doi.org/10.1080/15434303.2021.1931230>

- Huang, H. T. D., Hung, S. T. A., Chao, H. Y., Chen, J. H., Lin, T. P., & Shih, C. L. (2022). Developing and validating a computerized adaptive testing system for measuring the English proficiency of Taiwanese EFL university students. *Language Assessment Quarterly*, 19(2), 162-188. <https://doi.org/10.1080/15434303.2021.1984490>
- Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 305–315). Routledge. <https://doi.org/10.4324/9781351034784-33>
- Isbell, D. R., Delaney, M., & Park, L. (2024). Developing measures of L2 Korean vocabulary knowledge. *Vocabulary Learning and Instruction*, 13(1), 1-10. <https://doi.org/10.29140/vli.v13n1.1227>
- Jin, Y. (2023). Decision making in large-scale language testing: Intersections of policy, practice and research. *Studies in Language Assessment*, 12(1), 64–92. <https://doi.org/10.58379/HKKX5020>
- Joo, S. H., Seong, Y., Suh, J., Jung, J.-Y., & Purpura, J. E. (2023). Assessing Korean writing ability through a scenario-based assessment approach. *Assessing Writing*, 58, Article 100766. <https://doi.org/10.1016/j.asw.2023.100766>
- Kara, Y., Kamata, A., Qiao, X., Potgieter, C. J., & Nese, J. F. (2024). Equating oral reading fluency scores: A model-based approach. *Educational and Psychological Measurement*, 84(1), 190-209. <https://doi.org/10.1177/00131644221148122>
- Karyolemou, M. (2022). Teaching endangered languages of oral tradition: how and what to assess?. *Applied Linguistics*, 43(2), 389-411. <https://doi.org/10.1093/applin/amab045>
- Khabbazzbashi, N., Nakatsuhara, F., Inoue, C., Kaplan, G., & Green, A. (2022). The design and validation of an online speaking test for young learners in Uruguay: challenges and innovations. *International Journal of TESOL Studies*, 4(1), 141-168. <https://doi.org/10.46451/ijts.2022.01.10>
- Knoch, U., & Fan, J. (2024). Test score comparison tables: How well are they serving test users? *Language Testing*, 41(3), 681–693. <https://doi.org/10.1177/02655322241239348>
- Knoch, U., Deygers, B., & Khamboonruang, A. (2021). Revisiting rating scale development for rater-mediated language performance assessments: Modelling construct and contextual choices made by scale developers. *Language Testing*, 38(4), 602–626. <https://doi.org/10.1177/0265532221994052>
- Kunnan, A. J., Qin, C. Y., & Zhao, C. G. (2022). Developing a scenario-based English language assessment in an Asian university. *Language Assessment Quarterly*, 19(4), 368–393. <https://doi.org/10.1080/15434303.2022.2073886>
- Liao, M., Patton, J., Yan, R., & Jiao, H. (2021). Mining process data to detect aberrant test takers. *Measurement: Interdisciplinary Research and Perspectives*, 19(2), 93–105. <https://doi.org/10.1080/15366367.2020.1827203>
- Liao, R. J. (2023). The use of generalizability theory in investigating the score dependability of classroom-based L2 reading assessment. *Language Testing*, 40(1), 86-106. <https://doi.org/10.1177/02655322211070840>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 123, 103344. <https://doi.org/10.1016/j.system.2024.103344>
- Lloyd, J., Seed, G., Wen, X. & Yue, H. (2024). Mapping the SMEEA Gaokao tests to the

- CEFR. CEFR Journal – Research and Practice, 6, 56-79.
https://cefrjapan.net/images/PDF/CEFRJournal/CEFRJournal-6-4_Lloyd_Seed_Wen_Yue_2024.pdf
- Lu, Y. (2022). *Assessing learners' competence in L2 Chinese*. Routledge.
- MacGregor, D., Yen, S. J., & Yu, X. (2022). Using multistage testing to enhance measurement of an english language proficiency test. *Language Assessment Quarterly*, 19(1), 54-75. <https://doi.org/10.1080/15434303.2021.1988953>
- Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *Modern Language Journal*, 107, 669–692. <https://doi.org/10.1111/modl.12866>
- Martín-Gascón, B. (2023). Developing L2 learners' metaphoric competence: a case study of figurative motion constructions. *International Review of Applied Linguistics in Language Teaching*, 61(1), 79–109. <https://doi.org/10.1515/iral-2022-0043>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68, Article 100963.
<https://doi.org/10.1016/j.stueduc.2020.100963>
- Min, S., & Bishop, K. (2024). A shortened test is feasible: Evaluating a large-scale multistage adaptive English language assessment. *Language Testing*, 41(3), 627–648. <https://doi.org/10.1177/02655322231225426>
- Morea, N., Kasproicz, R. E., Morrison, A., & Silvestri, C. (2024). Diverse population, homogenous ability: The development of a new receptive vocabulary size test for young language learners in England using Rasch analysis. *Research Methods in Applied Linguistics*, 3(3), 100166. <https://doi.org/10.1016/j.rmal.2024.100166>
- Myburgh-Smit J. & Weideman, A. (2021). How early should we measure academic literacy? The usefulness of an appropriate test of academic literacy for Grade 10 students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 117-131). Multilingual Matters. <https://doi.org/10.21832/9781788926218-007>
- Nakamura, S., Phung, L., & Reinders, H. (2021). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*, 43(2), 428–441.
<https://doi.org/10.1017/S0272263121000437>
- Nakatsuhara, F., Inoue, C., Nitta, R., & Aoyama, T. (2023). Towards reforming the university entrance exam system in Japan: a case study of Linguaskill General. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/685092-research-notes-84.pdf>
- Nishizawa, H. (2024). Authenticity of academic lecture passages in high-stakes tests: A temporal fluency perspective. *Language Testing*, 41(4), 792–816.
<https://doi.org/10.1177/02655322241262453>
- Nour, P., Esfandiari, R., & Zarei, A. A. (2021). Development and validation of a metamemory maturity questionnaire in the context of English as a foreign language. *Language Testing in Asia*, 11, Article 24.
<https://doi.org/10.1186/s40468-021-00141-6>
- O'Sullivan, B. (2023). Integrating a Remote Proctoring Solution: the case of Aptis Remote. In H. Işık, E. Kantarcıoğlu, A. Üner, & C. Thomas (Eds.), *Re-Vision and Re-Form in English for Academic Purposes Contexts after the Pandemic* (pp. 75-82). Cambridge Scholars.

- O'Sullivan, B. Saville, N. (2024). Foreword. In: Field, J. *Insights into Assessing Academic Listening: The Case of IELTS*. SiLT 53, Cambridge: Cambridge University Press and Assessment.
<https://www.cambridge.org/cf/cambridgeenglish/teacher-development/insights-assessing-academic-listening-case-ielts/insights-assessing-academic-listening-case-ielts-paperback>
- Paek, I., Liang, X., & Lin, Z. (2021). Regarding item parameter invariance for the Rasch and the 2-parameter logistic models: An investigation under finite non-representative sample calibrations. *Measurement: Interdisciplinary Research and Perspectives*, 19(1), 39–54. <https://doi.org/10.1080/15366367.2020.1754703>
- Pan, J., & Marsden, E. (2024). Developing internet-based Tests of Aptitude for Language Learning (TALL): An open research endeavour. *Language Testing*, 41(4), 817–827. <https://doi.org/10.1177/02655322241241849>
- Panahi, A., & Mohebbi, H. (2022). Cognitive diagnostic assessment of IELTS listening: Providing feedback from its internal structure. *Language Teaching Research Quarterly*, 29, 147–160. <https://doi.org/10.32038/ltrq.2022.29.10>
- Papageorgiou, S., & Manna, V. F. (2021). Maintaining access to a large-scale test of academic language proficiency during the pandemic: The launch of TOEFL iBT Home Edition. *Language Assessment Quarterly*, 18(1), 36–41. <https://doi.org/10.1080/15434303.2020.1864376>
- Park, Y., Lee, S., & Shin, S.-Y. (2022). Developing a local academic English listening test using authentic unscripted audio-visual texts. *Language Testing*, 39(3), 401–424. <https://doi.org/10.1177/02655322221076024>
- Peng, Y., Yan, W., & Cheng, L. (2021). Hanyu Shuiping Kaoshi (HSK): A multi-level, multi-purpose proficiency test. *Language Testing*, 38(2), 326–337. <https://doi.org/10.1177/0265532220957298>
- Prabowo, M. Y., & Rahmadian, S. (2022). Computer-based English competency assessment for scholarship selection: Challenges, strategies, and implementation in the Ministry of Finance. *Jurnal Sositelknologi*, 21(1), 84–96. <https://doi.org/10.5614/sostek.itbj.2022.21.1.9>
- Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of an online language placement exam. *Language Assessment Quarterly*, 18(1), 42–50. <https://doi.org/10.1080/15434303.2020.1867145>
- Purpura, J.E (2021). A rationale for using a scenario-based assessment to measure competency-based, situated second and foreign language proficiency. In M. Masperi, C Cervini, Y Bardière (Eds.), *Évaluation des acquisitions langagières: du formatif au certificatif* (pp. 54-96). MediAzioni. <http://www.mediazioni.sitlec.unibo.it>
- Qin, C. Y., & Baker, B. (2025). Review of the Canadian English Language Proficiency Index Program (CELPPI). *Language Testing*, 42(1), 100-113. <https://doi.org/10.1177/02655322241291764>
- Rafatbakhsh, E., Ahmadi, A., Moloodi, A., & Mehrpour, S. (2021). Development and validation of an automatic item generation system for English idioms. *Educational Measurement: Issues and Practice*, 40(2), 49–59. <https://doi.org/10.1111/emip.12401>
- Rafi, A. (2024). Creativity, criticality and translanguaging in assessment design: perspectives from Bangladeshi higher education. *Applied Linguistics Review*, 15(5), 1887-1909. <https://doi.org/10.1515/applirev-2023-0086>

- Rambiritch, A., Alston, L. & Graham, M. (2021). Diagnosing with care: The academic literacy needs of theology students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 170-195). Multilingual Matters. <https://doi.org/10.21832/9781788926218-012>
- Ranger, J., Kuhn, J., & Pohl, S. (2021). Effects of motivation on the accuracy and speed of responding in tests: The speed-accuracy tradeoff revisited. *Measurement: Interdisciplinary Research and Perspectives*, 19(1), 15–38. <https://doi.org/10.1080/15366367.2020.1750934>
- Riemenschneider, A., Weiss, Z., Schröter, P., & Meurers, D. (2024). The interplay of task characteristics, linguistic complexity, and language proficiency in High-Stakes English as a Foreign language writing. *TESOL Quarterly*, 58(2), 775–801. <https://doi.org/10.1002/tesq.3254>
- Rossi, O., & Brunfaut, T. (2021). Text authenticity in listening assessment: Can item writers be trained to produce authentic-sounding texts? *Language Assessment Quarterly*, 18(4), 398–418. <https://doi.org/10.1080/15434303.2021.1895162>
- Sadeghi, K. & Douglas, D. (Eds.). (2023). *Fundamental considerations in technology mediated language assessment*. Routledge. <https://doi.org/10.4324/9781003292395>
- Sadoughi, M., & Hejazi, S. Y. (2022). Development and validation of Foreign Language Teacher Support Scale (FLTSS). *Language Testing in Asia*, 12, Article 30. <https://doi.org/10.1186/s40468-022-00186-1>
- Salaberry, M. R. & Weideman, A. (2024). Context, construct and ethics. In M. R., Salaberry, A. Weideman, A. & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 3-27). Routledge. <https://doi.org/10.4324/9781003384922-2>
- Sasayama, S., Garcia Gomez, P., & Norris, J. M. (2021). Designing efficient L2 writing assessment tasks for low-proficiency learners of English. *ETS Research Report Series*, 2021(1), 1–31. <https://doi.org/10.1002/ets2.12341>
- Seed, G., Salamoura, A. & Saville, N. (2024). New approaches to the assessment of English as an additional language. In N. A. Nazari & A. Mehdi Riazi (Eds.), *Adaptable English Language Teaching: Advances and Frameworks for Responding to New Circumstances* (pp.79-95). Routledge.
- Selvi, A. F., Saracoğlu, E., & Çalışkan, E. (2023). When inclusivity means playing safe: Ideological discourses and representations in English testing materials. *RELC Journal*, 54(3), 773-787. <https://doi.org/10.1177/00336882211008662>
- Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts? *Language Learning & Technology*, 27(3), 27–40. <https://hdl.handle.net/10125/73530>
- Shin, D., & Lee, J. H. (2024). AI-powered automated item generation for language testing. *ELT Journal*, 78(4), 446-452. <https://doi.org/10.1093/elt/ccae016>
- Shin, S. Y., Lee, S., & Lidster, R. (2021). Examining the effects of different English speech varieties on an L2 academic listening comprehension test at the item level. *Language Testing*, 38(4), 580–601. <https://doi.org/10.1177/0265532220985432>
- Shin, S., & Lee, S. (2024). Effects of the type of audio-visual input and listenability on young L2 learners' listening comprehension. *TESOL Quarterly*, 58(4), 1580–1608. <https://doi.org/10.1002/tesq.3298>
- Shoahosseini, R., Baghaei, P., Khodabakhshzadeh, H., & Ashraf, H. (2024). C-Test construct validity: Evidence from nonparametric item response theory. *Language*

- Testing in Asia*, 14, Article 10. <https://doi.org/10.1186/s40468-024-00279-z>
- Silva, A. L. B. C. (2023). Language proficiency test design for a group of military pilots: Criteria, construct and purpose. *Journal of Teaching English for Academic and Specific Purposes*, 11(1), 1-9. <https://doi.org/10.22190/JTESAP230120002S>
- Skaggs, G. (2021). *Test development and validation*. SAGE Publications.
- Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition*, 43(2), 462–471. <https://doi.org/10.1017/S0272263121000437>
- Steyn, S. (2021). Pathways to parity between parallel tests of language ability: Lessons from a project. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 132-147). Multilingual Matters.
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>
- Suárez, M. M. (2022). A need for an aptitude test for young learners in Catalan: The case of the Modern Language Aptitude Test - Elementary in Catalan. *Language Teaching Research Quarterly*, 31, 101-118. <https://doi.org/10.32038/ltrq.2022.31.08>
- Suhan, M., Papageorgiou, S., & Wolf, M. K. (2024). Mapping the scores of the TOEFL Primary® Writing test to the Common European Framework of Reference (CEFR) levels (Research Memorandum ETS RM-24-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-03.pdf>
- Sun, T. & Kim, S. Y. (2021). Evaluating six approaches to handling zero-frequency scores under equipercentile equating. *Measurement: Interdisciplinary Research and Perspectives*, 19(4), 213–235. <https://doi.org/10.1080/15366367.2020.1855034>
- Taylor, L., & Banerjee, J. (2023). Accommodations in language testing and assessment: Safeguarding equity, access, and inclusion. *Language Testing*, 40(4), 847-855. <https://doi.org/10.1177/02655322231186221>
- Turner, C. E. (2023). Assessment and evaluation: mixed methods research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Wiley Online Library. <https://doi.org/10.1002/9781405198431.wbeal0040.pub2>
- Van Dyk, T. (2021). Postscript: What the data tell us: an overview of language assessment research in South Africa's multilingual context. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 217–235). Multilingual Matters.
- Van Dyk, T., & Weideman, A. (2022). Taaltoetsing en die assessering van taalvermoë [Language testing and the assessment of language ability]. In W. Carstens & T. van Dyk (Eds.), *Toegepaste Taalkunde in Afrikaans* [Applied linguistics in Afrikaans] (pp. 125-1420). Van Schaik. <https://toegepaste-taalkunde.co.za/hoofstuk-7/>
- Van Dyk, T., Murre, P., & Kotzé, H. (2021). Does one size fit all? Some considerations for test translation. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 52–74). Multilingual Matters.
- Wagner, E., & Krylova, A. (2021). Temple University's ITA placement test in times of COVID-19. *Language Assessment Quarterly*, 18(1), 12–16. <https://doi.org/10.1080/15434303.2020.1862849>

- Wagner, E., Liao, Y. F., & Wagner, S. (2021). Authenticated spoken texts for L2 listening tests. *Language Assessment Quarterly*, 18(3), 205–227.
<https://doi.org/10.1080/15434303.2020.1860057>
- Wang, W., Kingston, N. M., Davis, M. H., Tiemann, G. C., Tonks, S., & Hock, M. (2021). Applying evidence-centered design in the development of a multidimensional adaptive reading motivation measure. *Educational Measurement: Issues and Practice*, 40(4), 91–100. <https://doi.org/10.1111/emip.12468>
- Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, 53, Article 100641.
<https://doi.org/10.1016/j.asw.2022.100641>
- Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/S0272263121000449>
- Weideman, A. (2021). A skills-neutral approach to academic literacy assessment. In A. Weideman, J. Read, & T. du Plessis, T. (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 22-51). Multilingual Matters.
- Weideman, A. (2022). Context, construct, and validation: A perspective from South Africa. *Language Assessment Quarterly*, 19(2), 124-141.
<https://doi.org/10.1080/15434303.2020.1860991>
- Weideman, A. (2024). Yardsticks for the future of language assessment: Disclosing the meaning of measurement. In M. R. Salaberry, A. Weideman, A., & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 220-234). Routledge. <https://doi.org/10.4324/9781003384922 -12>
- Weideman, A. & Van Dyk, T. (2023). Achieving technical economy: A modification of cloze procedure. *Language Teaching Research Quarterly*, 37, 144-160.
<https://doi.org/10.32038/ltrq.2023.37.07>
- Weideman, A., & Deygers, B. (2024). Validity and validation: An alternative perspective. In M. R. Salaberry, A. Weideman, & W-L. Hsu.(Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 28-50). Routledge.
<https://doi.org/10.4324/9781003384922 -3>
- Weideman, A., Read, J. & Du Plessis, T. (Eds.). (2021). *Assessing academic literacy in a multilingual society: Transition and transformation* (Volume 84). Multilingual Matters.
<https://doi.org/10.21832/WEIDEM6201>.
- Weir, C. & Chan, S. (2019). *Research and practice in assessing academic English: The case of IELTS*. Cambridge University Press.
- Wilsenach, C., & Schaefer, M. (2022). Development and initial validation of productive vocabulary tests for isiZulu, Siswati and English in South Africa. *Language Testing*, 39(4), 567–592. <https://doi.org/10.1177/02655322211063785>
- Winke, P., & Brunfaut, T. (2021). Perspectives on “knowing” a second language: What are we seeking to measure? In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.1–8). Routledge.
<https://doi.org/10.4324/9781351034784-1>
- Winke, P., & Brunfaut, T. (Eds.). (2021). *The Routledge handbook of second language acquisition and language testing*. Routledge.
<https://doi.org/10.4324/9781351034784>
- Wolf, M. K., Bailey, A. L., & Ballard, L. (2023). Aligning English language proficiency assessments to standards: Conceptual and technical issues. *TESOL Quarterly*,

- 57(2), 670-685. <https://doi.org/10.1002/tesq.3199>
- Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). Design framework for the TOEFL Primary® Writing test (Research Memorandum ETS RM-24-02). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-02.pdf>
- Wu, S. L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, 44(1), 271-300. <https://doi.org/10.1017/S0272263121000103>
- Xie, Q. (2022). An online database of English assessment tasks for developing student teachers' assessment literacy. *International Journal of Mobile Learning and Organization*, 16, Article 4. <https://doi.org/10.1504/IJMLO.2022.10048804>
- Xu, Q., Sawaki, Y., & Zhu, Y. (2024). Linguistic complexity measures of short narrative writings for modeling overall Chinese proficiency. *System*, 121, 103220. <https://doi.org/10.1016/j.system.2024.103220>
- Yan, X., Dimova, S., & Ginther, A. (Eds.). (2023). *Local language testing: Practice across contexts*. Springer.
- Yang, C. (2023). Using eye-tracking and retrospective verbal reports to explore the cognitive processes of banked gap-filling: a case study featuring methodological triangulation. *Language Testing in Asia*, 13, Article 22 (2023). <https://doi.org/10.1186/s40468-023-00234-4>
- Yang, X., & Osborne, C. (2023). The development and validation of a C-test and a pseudo-character test for online CFL assessment. *Journal of China Computer-Assisted Language Learning*, 3(1), 101-131. <https://doi.org/10.1515/jccall-2022-0019>
- Yigiter, M. S., & Dogan, N. (2023). Computerized multistage testing: Principles, designs and practices with R. *Measurement: Interdisciplinary Research and Perspectives*, 21(4), 254–277, <https://doi.org/10.1080/15366367.2022.2158017>
- Zhang, H., Sun, P., Bianglae, Y., & Widiawati, W. (2024). The development of a Chinese vocabulary proficiency test (CVPT) for learners of Chinese as a second/foreign language. *Language Testing*, 41(2), 412–442. <https://doi.org/10.1177/02655322231219998>
- Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. *Educational Science*, 12(10), Article 723. <https://doi.org/10.3390/educsci12100723>
- Zhao, J.-H., Chen, Z.-W., & Yang, Q.-F. (2024). I do and I understand: A virtual reality-supported collaborative design-assessing activity for EFL students. *System*, 121, 103213. <https://doi.org/10.1016/j.system.2023.103213>

Rating and Rating Scales

- Ali, M. M. E. (2023). The foreign-accentedness, comprehensibility, and intelligibility of L2 Arabic speech. *Language Teaching Research*, 0(0). <https://doi-org.ezproxy2.library.arizona.edu/10.1177/13621688231158787>
- Anderson, R. C., Chaparro, E. A., Smolkowski, K., & Cameron, R. (2023). Visual thinking and argumentative writing: A social-cognitive pairing for student writing development. *Assessing Writing*, 55, Article 100694. <https://doi.org/10.1016/j.asw.2023.100694>

- Baker, B., Polikar, S., & Homayounzadeh, M. (2021). "I was not thoughtful enough before:" Exploiting the learning-oriented-assessment potential of a test taker-oriented rubric for summative assessment. In A. Gebriel. (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 162–181). Routledge.
<https://doi.org/10.4324/9781003014102-12>
- Barkaoui, K. (2024). Multi-faceted Rasch analysis for evaluating second language tests. In Kunnan, A. (Ed.), *The concise companion to language assessment* (pp. 602-618). Wiley-Blackwell.
- Barkaoui, K. & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26-63. 2._sila_12_1__barkaoui_and_woodworth.pdf (altaanz.org)
- Betts, J., Muntean, W., Kim, D., & Kao, S. (2022). Evaluating different scoring methods for multiple response items providing partial credit. *Educational and Psychological Measurement*, 82(1), 151–176. <https://doi.org/10.1177/0013164421994636>
- Bondar, N. A., Day, S. X., Mountain, L., Turchi, L. B., & Hutchison, L. F. (2023). Composition Organization and Development Analysis (CODA) Scale: Equipping high school students to evaluate argumentative essays. *Assessing Writing*, 57, Article 100724. <https://doi.org/10.1016/j.asw.2023.100724>
- Bouwer, R., Koster, M., & Van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*, 30(3-4), 302-319. <https://doi.org/10.1080/0969594X.2023.2241656>
- Burton, J. D. (2024). Evaluating the impact of nonverbal behavior on language ability ratings. *Language Testing*, 41(4), 729–758. <https://doi.org/10.1177/02655322241255709>
- Cai, H., & Yan, X. (2024). Triangulating natural language processing (NLP)-based analysis of rater comments and many-facet Rasch measurement (MFRM): An innovative approach to investigating raters' application of rating scales in writing assessment. *Language Testing*, 41(2), 384–411. <https://doi.org/10.1177/02655322231210231>
- Carey, M. D., & Szocs, S. (2024). Revisiting raters' accent familiarity in speaking tests : Evidence that presentation mode interacts with accent familiarity to variably affect comprehensibility ratings. *Language Testing*, 41(2), 290–315. <https://doi.org/10.1177/02655322231200808>
- Chambers, L., Vitello, S., & Vidal Rodeiro, C. (2024). Moderation of non-exam assessments: a novel approach using comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 31(1), 32-55. <https://doi.org/10.1080/0969594X.2024.2313237>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487-2512. <https://doi.org/10.1177/13621688211046740>

- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, Article 102408. <https://doi.org/10.1016/j.system.2020.102408>
- Dai, D. W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test*. Peter Lang. <https://doi.org/10.3726/b21295>
- Davis, L., & Papageorgiou, S. (2021). Complementary strengths? Evaluation of a hybrid human-machine scoring approach for a test of oral academic English. *Assessment in Education: Principles, Policy & Practice*, 28(4), 437–455. <https://doi.org/10.1080/0969594X.2021.1979466>
- De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476. <https://doi.org/10.1080/0969594X.2021.1951162>
- Deygers, B. (2021). The CEFR companion volume: Between research-based policy and policy-based research. *Applied Linguistics*, 42(1), 186–191. <https://doi.org/10.1093/applin/amz024>
- Dobrić, N., Sigott, G., Ilc, G., Lazović, V., Cesnik, H., & Stopar, A. (2021). Errors as indicators of writing task difficulty at the Slovene general matura in English. *International Journal of Applied Linguistics*, 31(3), 475–491. <https://doi.org/10.1111/ijal.12345>
- Doeniyas, C., Gül, Z. T., & Alci, B. (2023). A non-Western adaptation of the Situated Academic Writing Self-Efficacy Scale (SAWSES). *Assessing Writing*, 57, Article 100763. <https://doi.org/10.1016/j.asw.2023.100763>
- Eckes, T., & Jin, K. Y. (2021). Examining severity and centrality effects in TestDaF writing and speaking assessments: An extended Bayesian many-facet Rasch analysis. *International Journal of Testing*, 21(3–4), 131–153. <https://doi.org/10.1080/15305058.2021.1963260>
- Eerdemutu, L., Dewaele, J.-M., & Wang, J. (2024). Developing a short language classroom engagement scale (Lces) and linking it with needs satisfaction and achievement. *System*, 120, 103189. <https://doi.org/10.1016/j.system.2023.103189>
- Elosua, P. (2022). Validity evidences for scoring procedures of a writing assessment task. A case study on consistency, reliability, unidimensionality and prediction accuracy. *Assessing Writing*, 54, Article 100669. <https://doi.org/10.1016/j.asw.2022.100669>
- Eskin, D. (2023). Writing task performance and first language background on an ESL placement exam: A many-facets rasch analysis of facet main effects and differential facet functioning. *Studies in Applied Linguistics & TESOL*, 23(1), 37–65. <https://doi.org/10.52214/salt.v23i1.11805>
- Frisén, L. B. (2024). Eyes on the ball: Teachers' rating processes when assessing a national L2 speaking test through the lens of their scoring document. *Studies in Language Assessment (SILA)*, 13(1), 1–36. <https://doi.org/10.58379/DAGH1483>
- Gesuato, S., Pagliarini, E., & Sanfelici, E. (2023). YOUNG ADULTS' INFERENTIAL READING SKILLS - RECOGNISING IMPLICIT AND UNSTATED CONTENT IN ITALIAN L1 AND ENGLISH L2. *Italiano LinguaDue*, 15(2), 20–20. <https://doi.org/10.54103/2037-3597/22011>
- Ghaffarvand-Mokari, P. (2024). Effects of listeners' dialectal variation on comprehensibility and accentedness judgements of second language speech. *Journal of Second Language Pronunciation*, 10(1), 35–58. <https://doi.org/10.1075/jslp.23018.gha>

- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Golparvar, S. E., & Abolhasani, H. (2022). Unpacking the contribution of linguistic features to graph writing quality: An analytic scoring approach. *Assessing Writing*, 53, Article 100644. <https://doi.org/10.1016/j.asw.2022.100644>
- Goodwin, R., & Kirkpatrick, R. (2023). Using rubrics to improve writing skills: a study in Kuwait. *Language Testing in Asia*, 13, Article 17 (2023). <https://doi.org/10.1186/s40468-023-00224-6>
- Graves, B. C. (2023). Specifications grading to promote student engagement, motivation and learning: Possibilities and cautions. *Assessing Writing*, 57, Article 100754. <https://doi.org/10.1016/j.asw.2023.100754>
- Grocott, C., Moe, E., & Verhelst, N. (2024). Investigating level descriptors and their impact on validity. *Studies in Language Assessment (SILA)*, 13(1), 199-236. <https://doi.org/10.58379/LIHU1555>
- Guo, Y., Xu, J., & Chen, C. (2023). Measurement of engagement in the foreign language classroom and its effect on language achievement: the case of Chinese college EFL students. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1225–1270. <https://doi.org/10.1515/iral-2021-0118>
- Han, C., Zheng, B-H., Xie, M-Q., & Chen, S-R. (2024). Raters' scoring process in assessment of interpreting: an empirical study based on eye tracking and retrospective verbalisation. *Interpreter and Translator Trainer*, 18(3), 400-422. <https://doi.org/10.1080/1750399X.2024.2326400>
- Harsch, C., Koval, V., Delgado-Osorio, X. & Hartig, J. (2024). Usability of CEFR Companion Volume scales for the development of an analytic rating scale for academic integrated writing assessment. In *CEFR Journal - Research and Practice*. Vol. 6, 155-177, <https://doi.org/10.37546/JALTSIG.CEFR6-9>
- Heidari, N., Ghanbari, N., & Abbasi, A. (2022). Raters' perceptions of rating scales criteria and its effect on the process and outcome of their rating. *Language Testing in Asia*, 12, Article 20. <https://doi.org/10.1186/s40468-022-00168-3>
- Holzkecht, F., Tornay, S., Battisti, A., Batty, A. O., Tissi, K., Haug, T., & Ebling, S. (2024). Automated Sign Language Vocabulary Assessment: Comparing Human and Machine Ratings and Studying Learner Perceptions. *Language Assessment Quarterly*, 21(3), 245–265. <https://doi.org/10.1080/15434303.2024.2364877>
- Honko, M., Neittaanmäki, R., Jarvis, S., & Huhta, A. (2023). Beyond literacy and competency – The effects of raters' perceived uncertainty on assessment of writing. *Assessing Writing*, 57, Article 100768. <https://doi.org/10.1016/j.asw.2023.100768>
- Horák, T., & Gandini, E. A. M. (2021). Going off-grid: Multiple benefits of rethinking the marking criteria format. *Research Notes*, 80, 7–20, Cambridge Assessment English. <https://www.cambridgeenglish.org/Images/630216-research-notes-80.pdf>
- Huang, J., & Chen, G. (2022). Individualized feedback to raters in language assessment: Impacts on rater effects. *Assessing Writing*, 52, Article 100623. <https://doi.org/10.1016/j.asw.2022.100623>
- Huang, J., Shu, T., Dong, Y., & Zhu, D. (2023). Constructing and validating a self-assessment scale for Chinese college English-major students' feedback knowledge repertoire in EFL academic writing: Item response theory and factor analysis approaches. *Assessing Writing*, 56, Article 100716.

- <https://doi.org/10.1016/j.asw.2023.100716>
- Huang, J., Zhu, D., Xie, D., & Shu, T. (2023). Examining the reliability of an international Chinese proficiency standardized writing assessment: Implications for assessment policy makers. *Assessing Writing*, 55, Article 100693. <https://doi.org/10.1016/j.asw.2023.100693>
- Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, 55, Article 100676. <https://doi.org/10.1016/j.asw.2022.100676>
- Imperial, J. M., Forey, G., & Tayyar Madabushi, H. (2024). Standardize: Aligning Language Models with Expert-Defined Standards for Content Generation. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1573–1594). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.94>
- Isaacs, T. & Thomson, R. I. (2022). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. In J. M. Levis, T. M. Derwing, & M. J. Munro (Eds.), *The evolution of pronunciation teaching and research: 25 years of intelligibility, comprehensibility, and accentedness* (pp. 125-151). John Benjamins.
- Isaacs, T., & Rose, H. (2022). Redressing the balance in the native speaker debate: Assessment standards, standard language, and exposing double standards. *TESOL Quarterly*, 56(1), 401-412. <https://doi.org/10.1002/tesq.3041>
- Jin, K.-Y., & Eckes, T. (2022). Detecting Differential Rater Functioning in Severity and Centrality: The dual DRF facets model. *Educational and Psychological Measurement*, 82(4), 757–781. <https://doi.org/10.1177/00131644211043207>
- Jönsson, A., Balan, A., & Hartell, E. (2021). Analytic or holistic? A study about how to increase the agreement in teachers' grading. *Assessment in Education: Principles, Policy & Practice*, 28(3), 212–227. <https://doi.org/10.1080/0969594X.2021.1884041>
- Khodi, A. (2021). The affectability of writing assessment scores: A G-theory analysis of rater, task, and scoring method contribution. *Language Testing in Asia*, 11, Article 30. <https://doi.org/10.1186/s40468-021-00134-5>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73–95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Kim, M., Qiu, X., & Wang Y. (2024). Interrater agreement in genre analysis: A methodological review and a comparison of three measures, *Research Methods in Applied Linguistics*, 3(1), Article 100097. <https://doi.org/10.1016/j.rmal.2024.100097>
- Kim, S. (2021). Generalizability of CEFR criterial grammatical features in a Korean EFL corpus across A1, A2, B1, and B2 levels. *Language Assessment Quarterly*, 18(3), 273–295. <https://doi.org/10.1080/15434303.2020.1855647>
- Knoch, U., Deygers, B., & Khamboonruang, A. (2021). Revisiting rating scale development for rater-mediated language performance assessments: Modelling construct and contextual choices made by scale developers. *Language Testing*, 38(4), 602–626.

- <https://doi.org/10.1177/0265532221994052>
- Knoch, U., Fairbairn, J., & Jin, Y. (2021). *Scoring second language spoken and written performance: Issues, options and directions*. University of Toronto Press.
- Koizumi, R., & In'nami, Y. (2024). Predicting functional adequacy from complexity, accuracy, and fluency of second-language picture-prompted speaking. *System*, 120, 103208. <https://doi.org/10.1016/j.system.2023.103208>
- Koizumi, R., & Watanabe, A. (2021). Rater reliability in classroom speaking assessment in a Japanese senior high school. *Annual Review of English Language Education in Japan*, 32, 129-144. https://doi.org/10.20581/arele.32.0_129
- Koizumi, R., Fukazawa, M., & Inoue, C. (2024). Usability of a Speaking Assessment Portal for Japanese teachers of English. *SHIKEN: A Journal of Language Testing and Evaluation in Japan*, 28(1), 19-37. <https://doi.org/10.37546/JALTSIG.TEVAL28.1-2>
- Kvasova, O., Shevchenko, T., Hnapovska, L., Kalinichenko, V., & Budas, I. (2022). Implications of training university teachers in developing local writing rating scales. *Studies in Language Assessment*, 11(1), 37–52. https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4282559/SiLA-11.1-Kvasova-et-al.pdf
- Lamprianou, I., Tsagari, D., & Kyriakou, N. (2021). The longitudinal stability of rating characteristics in an EFL examination: Methodological and substantive considerations. *Language Testing*, 38(2), 273–301. <https://doi.org/10.1177/0265532220940960>
- Lamprianou, I., Tsagari, D., & Kyriakou, N. (2023). Experienced but detached from reality: Theorizing and operationalizing the relationship between experience and rater effects. *Assessing Writing*, 56, Article 100713. <https://doi.org/10.1016/j.asw.2023.100713>
- Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62–85. <https://doi.org/10.1177/0265532220929918>
- Lee, I., Karaca, M., & Inan, S. (2023). The development and validation of a scale on L2 writing teacher feedback literacy. *Assessing Writing*, 57, Article 100743. <https://doi.org/10.1016/j.asw.2023.100743>
- Lestari, S. B., & Brunfaut, T. (2023). Operationalizing the reading-into-writing construct in analytic rating scales: Effects of different approaches on rating. *Language Testing*, 40(3), 684–722. <https://doi.org/10.1177/02655322231155561>
- Li, J., & Huang, J. (2022). The impact of essay organization and overall quality on the holistic scoring of EFL writing: perspectives from classroom English teachers and national writing raters. *Assessing Writing*, 51, Article 100604. <https://doi.org/10.1016/j.asw.2021.100604>
- Liao, J. (2023). What skills are being assessed? Evaluating L2 Chinese essays written by hand and on a computer keyboard. *Assessing Writing*, 57, Article 100765. <https://doi.org/10.1016/j.asw.2023.100765>
- Lin, R. (2023). Examining the scoring of content integration in a listening-speaking Test: A G-Theory analysis. *Language Assessment Quarterly*, 20(3), 319–338. <https://doi.org/10.1080/15434303.2023.2242334>
- Liu, S., & Lin, D. (2022). Developing and validating an analytic rating scale for a paraphrase task. *Assessing Writing*, 53, Article 100646.

- <https://doi.org/10.1016/j.asw.2022.100646>
- Liu, Y., & Liu, J. (2023). Comparing computer-based and paper-based rating modes in an English writing test. *Assessing Writing*, 57, Article 100771. <https://doi.org/10.1016/j.asw.2023.100771>
- Lloyd, J., Seed, G., Wen, X. & Yue, H. (2024). Mapping the SMEEA Gaokao tests to the CEFR. *CEFR Journal – Research and Practice*, 6, 56-79. https://cefrjapan.net/images/PDF/CEFRJournal/CEFRJournal-6-4_Lloyd_Seed_Wen_Yue_2024.pdf
- Lukácsi, Z. (2021). Developing a level-specific checklist for assessing EFL writing. *Language Testing*, 38(1), 86–105. <https://doi.org/10.1177/0265532220916703>
- Ma, W. (2022). What the analytic versus holistic scoring of international teaching assistants can reveal: Lexical grammar matters. *Language Testing*, 39(2), 239–264. <https://doi.org/10.1177/02655322211040020>
- Matthews, J. (2023). Using inter-rater discourse to trace the origins of disagreement: towards collective reflective practice in L2 assessment. *RELC Journal*, 54(1), 99–113. <https://doi.org/10.1177/0033688220977373>
- Mitchell, K. M., McMillan, D. E., Lobchuk, M. M., Nickel, N. C., Rabbani, R., & Li, J. (2021). Development and validation of the situated academic writing self-efficacy scale (SAWSES). *Assessing Writing*, 48, Article 100524. <https://doi.org/10.1016/j.asw.2021.100524>
- Mohsen, M. A. (2024). L1 versus L2 writing processes: What insight can we obtain from a keystroke logging program? *Language Teaching Research*, 28(6), 2251-2275. <https://doi.org/10.1177/13621688211041292>
- Moradkhan, D., & Shahmirzadi, N. (2012). Utilizing measures of (CAF) complexity, accuracy and fluency for interpreter evaluation. *Translation Studies (Motale'at-e-Tarjome Quarterly)*, 10(39), 87–104, (ISSN: 1735–0212).
- Müller, A. & Han, W. (2022). *IELTS writing band scores 5.5–7.5: Grammatical error rates, stakeholder perceptions, and risk* (IELTS Research Reports Online Series, No. 1/22). International English Language Testing System. <https://www.ielts.org/-/media/research-reports/muller-and-han-2022.ashx>
- Mumford, S., & Atay, D. (2021). Teachers' perspectives on the causes of rater discrepancy in an English for Academic Purposes context. *Assessing Writing*, 48, Article 100527. <https://doi.org/10.1016/j.asw.2021.100527>
- Nagle, C., & Rehman, I. (2021). Doing L2 speech research online: Why and how to collect online ratings data. *Studies in Second Language Acquisition*, 43(4), 916–939. <https://doi.org/10.1017/S0272263121000292>
- Nakatsuhara, F., Inoue, C., & Taylor, L. (2021). Comparing rating modes: Analysing live, audio and video ratings of IELTS Speaking Test performance. *Language Assessment Quarterly*, 18(2), 83–106. <https://doi.org/10.1080/15434303.2020.1799222>
- Neittaanmäki, R., & Lamprianou, I. (2024). All types of experience are equal, but some are more equal: The effect of different types of experience on rater severity and rater consistency. *Language Testing*, 41(3), 606–626. <https://doi.org/10.1177/02655322241239362>
- Neittaanmäki, R., & Lamprianou, I. (2024). Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, 41(3), 584–605. <https://doi.org/10.1177/02655322241239363>

- Nikmard, F., & Tavassoli, K. (2023). The impact of test length on raters' mental processes during scoring test-takers' writing performance. *Journal of Language Horizons*, 7(1), 159-182. <https://doi.org/10.22051/lghor.2022.37340.1545>.
- Nikmard, F., Tavassoli, K., & Pourdana, N. (2023). Designing and validating a scale for evaluating the sources of unreliability of a high-stakes test. *Language Testing in Asia*, 13, Article 2 (2023). <https://doi.org/10.1186/s40468-023-00215-7>
- Nowruzi, M. (2021). A study of EFL teachers' classroom grading practices in secondary schools and private institutes: A mixed methods approach. *Language Testing in Asia*, 11, Article 29. <https://doi.org/10.1186/s40468-021-00145-2>
- Ogawa, C. (2022). CAF indices and human ratings of oral performances in an opinion-based monologue task. *Language Testing in Asia*, 12, Article 4. <https://doi.org/10.1186/s40468-022-00154-9>
- Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51, Article 100603. <https://doi.org/10.1016/j.asw.2021.100603>
- Palmour, L. (2023). Assessing oral presentations: An analysis of score-reaching dialogue between EAP practitioners. *Journal of English for Academic Purposes*, 62, Article 101210. <https://doi.org/10.1016/j.jeap.2022.101210>
- Phuong, H. Y., Phan, Q. T. & Le, T. T. (2023). The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: a Vietnamese contextual study. *Language Testing in Asia*, 13, Article 42 (2023). <https://doi.org/10.1186/s40468-023-00256-y>
- Pill J., & Smart, C. (2021). Raters: Behavior and training. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 135–144). Routledge. <https://doi.org/10.4324/9781351034784-15>
- Reed, D. K., Ma, J., & Gerde, H. K. (2023). Resiliency and vulnerability in early grades writing performance during the COVID-19 pandemic. *Assessing Writing*, 57, Article 100741. <https://doi.org/10.1016/j.asw.2023.100741>
- Rezaei, A., & Barkaoui, K. (2021). Peer and teacher assessment of L2 writing in high- and low-stakes conditions. *International Journal of Applied Linguistics*, 172(2), 199–228. <https://doi.org/10.1075/itl.20006.rez>
- Rock, K. (2022). Constructing a data-based analytic rubric for an academic blog post. *Assessing Writing*, 51, Article 100602. <https://doi.org/10.1016/j.asw.2021.100602>
- Safari, F., & Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *Journal of Second Language Writing*, 60, Article 101007. <https://doi.org/10.1016/j.jslw.2023.101007>
- Şahin, A. (2021). The Rater Performance Categorization System (RPCS) for intensive English programs. *Shanlax International Journal of Education*, 9(3), 225–241. <https://doi.org/10.34293/education.v9i3.3986>
- Saito, K., & Liu, Y. (2022). Roles of collocation in L2 oral proficiency revisited: Different tasks, L1 vs. L2 raters, and cross-sectional vs. longitudinal analyses. *Second Language Research*, 38(3), 531–554. <https://doi.org/10.1177/0267658320988055>
- Seedhouse, P., & Satar, M. (2021). VEO IELTS Project Report: Which specific features of candidate talk do examiners orient to when taking scoring decisions? *IELTS Research Reports Online Series*, 5. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/for-researchers/research-reports/online-series-2021-5>

- Shin, J. (2022). Investigating and optimizing score dependability of a local ITA speaking test across language groups: A generalizability theory approach. *Language Testing*, 39(2), 313–337. <https://doi.org/10.1177/02655322211052680>
- Shin, J., & Gierl, M. J. (2021). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 38(2), 247–272. <https://doi.org/10.1177/0265532220937830>
- Stansfield, C.W., Gao, J., & Rivers, W.P. (2010). A concurrent validity study of self-assessments and the federal interagency roundtable oral proficiency interview. *Russian Language Journal / Русский Язык*, 60, 299–315. <http://www.jstor.org/stable/43669189>
- Steiss, J., Krishnan, J., Kim, Y-S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, Article 100600. <https://doi.org/10.1016/j.asw.2021.100600>
- Su, Y., & Chen, X. (2024). Examining refusal and acceptance sequences employing a data-driven binary rating approach. *Research Methods in Applied Linguistics*, 3(3), Article 100161. <https://doi.org/10.1016/j.rmal.2024.100161>
- Su, Y., & Shin, S.-Y. (2024). Comparing two formats of data-driven rating scales for classroom assessment of pragmatic performance with roleplays. *Language Testing*, 41(2), 357–383. <https://doi.org/10.1177/02655322231210217>
- Sudina, E., Vernon, T., Foster, H., Del Villano, H., Hernandez, S., Beck, D., & Plonsky, L. (2021). Development and initial validation of the L2-teacher grit scale. *TESOL Quarterly*, 55(1), 156–184. <https://doi.org/10.1002/tesq.581>
- Sureeyatanapas, P., Sureeyatanapas, P., Panitanarak, U., Kraisiwattana, J., Sarootyanapat, P., & O'Connell, D. (2024). The analysis of marking reliability through the approach of gauge repeatability and reproducibility (GR&R) study: A case of English-speaking test. *Language Testing in Asia*, 14, Article 1. <https://doi.org/10.1186/s40468-023-00271-z>
- Syquia, J., & Leeming, P. (2024). Assessing interactional competence through group discussion: A mixed methods validation. *Research Methods in Applied Linguistics*, 3(3), Article 100144. <https://doi.org/10.1016/j.rmal.2024.100144>
- Tanaka, M., & Ross, S. J. (2023). Impact of self-construal on rater severity in peer assessments of oral presentations. *Assessment in Education: Principles, Policy & Practice*, 30(2), 203–220. <https://doi.org/10.1080/0969594X.2023.2189564>
- Tavassoli, K., Bashiri, L., & Pourdana, N. (2022). Experienced and novice L2 raters' cognitive processes while rating integrated and independent writing tasks. *Journal of Language and Education*, 8(4), 169–181. <https://doi.org/10.17323/jle.2022.13466>
- Thai, T., & Sheehan, S. (2022). The processes of rating L2 speaking performance using an analytic rating scale – A qualitative exploration. *Language Education and Assessment*, 5(1), 34–51. <https://doi.org/10.29140/lea.v5n1.777>
- Thomson, R. I., & Isaacs, T. (2022). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research* (pp. 24–44). Springer.
- Trofimovich, P., Isaacs, T., Kennedy, S., & Tsunemoto, A. (2022). Speech comprehensibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and L2 speaking* (pp. 174–187). Routledge.
- Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in

- an EAP context. *Assessing Writing*, 52, Article 100609.
<https://doi.org/10.1016/j.asw.2022.100609>
- Uludag, P., McDonough, K., & Trofimovich, P. (2022). Exploring shared and individual assessment of paired oral interactions. *Studies in Language Assessment*, 11(2), 1–24.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4258357/2022-Early-View_Uludag-et-al.pdf
- Vo, Y., Rickels, H., Welch, C., & Dunbar, S. (2023). Human scoring versus automated scoring for English learners in a statewide evidence-based writing assessment. *Assessing Writing*, 56, Article 100719. <https://doi.org/10.1016/j.asw.2023.100719>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, Article 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Wheeler, J. M., Engelhard, G., & Wang, J. (2022). Exploring rater accuracy using unfolding models combined with topic models: Incorporating supervised latent Dirichlet allocation. *Measurement: Interdisciplinary Research and Perspectives*, 20(1), 34–46.
<https://doi.org/10.1080/15366367.2021.1915094>
- Williams, K. (2024). Exploring English for academic purposes instructors' perceptions of speech fluency through developing and piloting a rating scale for a paired conversational task. *System*, 121, 103266.
<https://doi.org/10.1016/j.system.2024.103266>
- Wind, S. A. (2022). *Exploring rating scale functioning for survey research*. SAGE.
- Wind, S. A. (2023). A sequential approach to detecting differential rater functioning in sparse rater-mediated assessment networks. *Language Testing*, 40(2), 209–226.
<https://doi.org/10.1177/02655322221092388>
- Xie, Q. Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing*, 57, Article 100755. <https://doi.org/10.1016/j.asw.2023.100755>
- Yan, X., & Chuang, P.-L. (2023). How do raters learn to rate? Many-facet Rasch modeling of rater performance over the course of a rater certification program. *Language Testing*, 40(1), 153–179. <https://doi.org/10.1177/02655322221074913>
- Yin, Y., Yu, G., & Huang, D. (2024). Implementing the ICAO language proficiency requirements: The localization of aviation English testing. *Applied Linguistics*, 45(4), 699–717. <https://doi.org/10.1093/applin/amad057>
- Zhang, P., Fan, J., & Jia, W. (2022). Assessing L2 integrated writing self-efficacy: Development and validation of a scale. *Assessing Writing*, 54, Article 100665.
<https://doi.org/10.1016/j.asw.2022.100665>
- Zhang, T., & Weng, Z. (2024). Implementing rubric co-construction in ESL writing teaching. *ELT Journal*, 78(1), 32–41. <https://doi.org/10.1093/elt/ccad033>
- Zhang, X. (2022). The relationship between lexical use and L2 writing quality: A case of two genres. *International Journal of Applied Linguistics*, 32(3), 371–396.
<https://doi.org/10.1111/ijal.12420>
- Zhao, C. G., & Wu, J. (2022). Perceptions of authorial voice: Why discrepancies exist. *Assessing Writing*, 53, Article 100632. <https://doi.org/10.1016/j.asw.2022.100632>

Specific Purpose Testing

- Ahmadi Safa, M., & Sheykhmololuki, H. (2023). An impact study of the Iranian National University Entrance Exam from students and parents' perspectives. *Language Testing in Asia*, 13, Article 40. <https://doi.org/10.1186/s40468-023-00254-0>
- Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2023). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition*, 46(2), 307–329. <https://doi.org/10.1017/S0272263123000487>
- Appel, R., Arias, A., Baker, B., & Loignon, G. (2023). Insights from lexical and syntactic analyses of a French for academic purposes assessment. *Assessing Writing*, 58, Article 100789. <https://doi.org/10.1016/j.asw.2023.100789>
- Arias, A., Vessey, R., & Sheyholislami, J. (2022). Jumping through hoops in the Canadian immigration system: A critical view of the immigrant's journey to citizenship. *TESOL Quarterly*, 56(4), 1430–1444. <https://doi.org/10.1002/tesq.3184>
- Barnau, A. (2023). Testing English for medical purposes: The effects of traditional and distance education on learning outcomes. *Language Learning in Higher Education*, 13(2), 451–460. <https://doi.org/10.1515/cercles-2023-2027>
- Bell, D. E. (2024). EAP materials and EAP assessment. In D. E. Bell (Ed.), *English for academic purposes: Perspectives on the past, present and future* (pp. 105–123). Multilingual Matters. <https://doi.org/10.2307/jj.21995548.11>
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing*, 40(1), 133–152. <https://doi.org/10.1177/02655322211064629>
- Bullock, N., & Westbrook, C. (2021). Testing in ESP: Approaches in aviation and maritime English. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world*. (67–77). Springer. https://doi.org/10.1007/978-981-33-4232-3_7
- Camara, W. J., & Mattern, K. (2022). Inflection point: The role of testing in admissions decisions in a postpandemic environment. *Educational Measurement: Issues and Practice*, 41(1), 10–15. <https://doi.org/10.1111/emip.12493>
- Chang, T. (2023). A move analysis of Chinese L2 student essays from the sociocognitive perspective: Genres, languages, and writing quality. *Assessing Writing*, 57, Article 100750. <https://doi.org/10.1016/j.asw.2023.100750>
- Chovancová, B., Bilová, Š., & Hradilová, A. (2023). Turning the tables on online exam cheating via language mediation tasks. *Language Learning in Higher Education*, 13(2), 517–533. <https://doi.org/10.1515/cercles-2023-2033>
- Cliff, A. (2021). The use of mediation and feedback in a standardised test of academic literacy: Theoretical and design considerations. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 75–92). Multilingual Matters.
- Crosby, S. (2023). A fair go: Translanguaging and assessment practices in a New Zealand junior college. *Studies in Language Assessment*, 12(2), 117–150.
- Davidson, S. (2022). The domain expert perspective: A qualitative study into the views expressed in a standard-setting exercise on a language for specific purposes (LSP) test for health professionals. *Language Testing*, 39(1), 117–141. <https://doi.org/10.1177/02655322211010737>
- Delgado-Osorio, X., Koval, V., Hartig, J., & Harsch, C. (2023). Strategic processing of source text in reading-into-writing tasks: A comparison between summary and argumentative tasks. *Journal of English for Academic Purposes*, 62, Article 101227.

- <https://doi.org/10.1016/j.jeap.2023.101227>
- Dissake, E. M. K. (2022). Assessing litigant's language proficiency: the case of the Bafoussam Court of First Instance. *Language Policy*, 21(2), 217-234. <https://doi.org/10.1007/s10993-021-09603->
- Doeniyas, C., Gül, Z. T., & Alçı, B. (2023). A non-Western adaptation of the Situated Academic Writing Self-Efficacy Scale (SAWSES). *Assessing Writing*, 57, Article 100763. <https://doi.org/10.1016/j.asw.2023.100763>
- Douglas, D. (2023). Specificity, authenticity, and inseparability: Assessing integrated oral skills for pilots and air traffic controllers. *Applied Linguistics Papers*, 27(1), 4–13. <https://doi.org/10.32612/uw.25449354>
- Drennan, L. (2022). Transitioning from undergraduate to postgraduate studies : A multistage evaluation of a discipline-specific writing intervention. *Journal for Language Teaching*, 56(2), 1-24. <https://doi.org/10.56285/jltVol56iss2a5293>
- Drennan, L., Joubert, M., Weideman, A. & Posthumus, R. (2021). Combined measures of identifying language risk for first-year students. *Journal for Language Teaching*, 55(2), 93-116. <https://doi.org/10.4314/jlt.v55i2.4>
- Du Plessis, C. (2021): Basic education and academic literacy: conflicting constructs in the South African National Senior Certificate (NSC) Language Examination. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 95-116). Multilingual Matters.
- Du Plessis, T. (2021). Institutional language policy and academic literacy in South African higher education. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 3-21). Multilingual Matters.
- Ducasse, A.M. (2022) Oral reflection tasks: Advanced Spanish L2 learner insights on emergency remote teaching assessment practices in a higher education context. *Languages*, 7(1), Article 26. <https://doi.org/10.3390/languages7010026>
- Gaillat, T., Lafontaine, A., & Knefati, A. (2023). Visualizing linguistic complexity and proficiency in learner English writings. *CALICO Journal*, 40(2), 178-197. <https://doi.org/10.1558/cj.19487>
- Gui, M., Cao, X., & Zhao, J. (2023). EAP reading ability development during the process of generating research questions: A CDST perspective. *English for Academic Purposes*, 62, Article 101249. <https://doi.org/10.1016/j.jeap.2023.101249>
- Haug, T., Mann, W., & Knoch, U. (2022). *The handbook of language assessment across modalities*. Oxford University Press.
- Ho, S.K., & Gan, Z. (2023). Instructional practices and students' reading performance: a comparative study of 10 top performing regions in PISA 2018. *Language Testing in Asia*, 13, Article 48 (2023). <https://doi.org/10.1186/s40468-023-00261-1>
- Horák, T., & Gandini, E. A. M. (2021). Going off-grid: Multiple benefits of rethinking the marking criteria format. *Research Notes*, 80, 7–20, Cambridge Assessment English. <https://www.cambridgeenglish.org/Images/630216-research-notes-80.pdf>
- Hu, T & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486-1501, DOI: 10.1080/13670050.2019.1691498
- Hyung-Jo, Y., & Mahmoud, A.T. (2023). Authorial voice in source-based and opinion-based argumentative writing: Patterns of voice across task types and proficiency levels.

- English for Academic Purposes*, 62, Article 101228.
<https://doi.org/10.1016/j.jeap.2023.101228>
- Ji, J. (2021). College-level placement for heritage language learners. *Foreign Language Annals*, 54(3), 690–713. <https://doi.org/10.1111/flan.12523>
- Katagiri, W. N. (2023). How sociocultural factors mediate washback effect: a study of the Hong Kong Diploma of Secondary Education Examination English Language Paper 4. *Language Testing in Asia*, 13, Article 38 (2023).
<https://doi.org/10.1186/s40468-023-00250-4>
- Khan, K. (2022). The securitisation of language borders and the (re) production of inequalities. *TESOL Quarterly*, 56(4), 1458-1470. <https://doi.org/10.1002/tesq.3186>
- Kim, M., Qiu, X., & Wang Y. (2024). Interrater agreement in genre analysis: A methodological review and a comparison of three measures, *Research Methods in Applied Linguistics*, 3(1), Article 100097.
<https://doi.org/10.1016/j.rmal.2024.100097>
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247.
<https://doi.org/10.1016/j.jeap.2023.101247>
- Knoch, U., & Fan, J. (2024). Assessing clinician language skills. In Ortega, P., Martínez, G., Lor, M., & Ramírez, A. S. *The Handbook of Language in Public Health and Healthcare*. Wiley. <https://doi.org/10.1002/9781119853855.ch12>
- Kunnan, A. J., Qin, C. Y., & Zhao, C. G. (2022). Developing a scenario-based English language assessment in an Asian university. *Language Assessment Quarterly*, 19(4), 368–393. <https://doi.org/10.1080/15434303.2022.2073886>
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781–812. <https://doi.org/10.1017/S0272263120000546>
- Liu, Y., Cui, Y., & Yu, H. (2023). Test review: a general academic English summative test. *Language Testing in Asia*, 13, Article 49 (2023).
<https://doi.org/10.1186/s40468-023-00263-z>
- MacGregor, D., Yen, S. J., & Yu, X. (2022). Using multistage testing to enhance measurement of an english language proficiency test. *Language Assessment Quarterly*, 19(1), 54-75. <https://doi.org/10.1080/15434303.2021.1988953>
- McDonough, K., Uludag, P. & Neumann, H. (2021). Assessing case study exams in a business course: What role does language play? *Business and Professional Communication Quarterly*, 82(2), 116–134.
<https://doi.org/10.1177/2F23294906211012398>
- Monfils, L. F., & Manna, V. F. (2021). Time to achieving a designated criterion score level: A survival analysis study of test taker performance on the TOEFL iBT® test. *Language Testing*, 38(1), 154–176. <https://doi.org/10.1177/0265532220940709>
- Myburgh-Smit J.& Weideman, A. (2021). How early should we measure academic literacy? The usefulness of an appropriate test of academic literacy for Grade 10 students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 117-131). Multilingual Matters. <https://doi.org/10.21832/9781788926218-007>
- Nakamura, K. (2022). Investigating the factor structure of the Test of English for Academic Purposes (TEAP) and its relation to test takers' perceived test task value. *Language*

- Testing in Asia*, 12, Article 35. <https://doi.org/10.1186/s40468-022-00183-4>
- Nakatsuhara, F., Inoue, C., Nitta, R., & Aoyama, T. (2023). Towards reforming the university entrance exam system in Japan: a case study of Linguaskill General. Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/685092-research-notes-84.pdf>
- Neumann, H., *Kozak, S., & *Gil, L. (2023). The impact of academic acculturation and language proficiency on international students' university experience and academic success: A longitudinal study. *Canadian Modern Language Review*, 79, 121-140. <https://doi.org/10.3138/cmlr-2021-0055>
- Palmour, L. (2023). Assessing oral presentations: An analysis of score-reaching dialogue between EAP practitioners. *Journal of English for Academic Purposes*, 62, Article 101210. <https://doi.org/10.1016/j.jeap.2022.101210>
- Phuong, H. Y., Phan, Q. T. & Le, T. T. (2023). The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: a Vietnamese contextual study. *Language Testing in Asia*, 13, Article 42 (2023). <https://doi.org/10.1186/s40468-023-00256-y>
- Rambiritch, A., Alston, L. & Graham, M. (2021). Diagnosing with care: The academic literacy needs of theology students. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 170-195). Multilingual Matters. <https://doi.org/10.21832/9781788926218-012>
- Sánchez-Pérez, M. D. M. (2021). Predicting content proficiency through disciplinary-literacy variables in English-medium writing. *System*, 97, Article 102463. <https://doi.org/10.1016/j.system.2021.102463>
- Schmidgall, J., & Powers, D. E. (2021). Predicting communicative effectiveness in the international workplace: Support for TOEIC® Speaking test scores from linguistic laypersons. *Language Testing*, 38(2), 302–325. <https://doi.org/10.1177/0265532220941803>
- Schnoor, B., Hartig, J., Klinger, T., Naumann, A., & Usanova, I. (2023). Measuring the development of general language skills in English as a foreign language—Longitudinal invariance of the C-test. *Language Testing*, 40(3), 796-819. <https://doi.org/10.1177/02655322231159829>
- Sebolai, K. (2021). Generic academic literacy testing: A logical precursor for faculty-specific language teaching and assessment. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 151-169). Multilingual Matters.
- Sebolai, K. (2022). Determining the faculty specific academic literacies needs of first year university students: a mixed methods approach. *Journal for Language Teaching*, 56(2), 1-20. <https://doi.org/10.56285/jltVol56iss2a5702>
- Sebolai, K. (2022). The Academic Development approach to academic literacy in higher education South Africa: a disconnect between teaching and assessment. *Journal for Language Teaching*, 56(2), 1-16. <https://doi.org/10.56285/jltVol56iss2a5387>
- Shamshiri, F., Esfahani, F. R., & Hosseini, S. E. (2023). Models of assessment in the classroom: A comparative research of CALL-based vs. traditional assessment on vocabulary learning among Iranian EFL learners. *Language Testing in Asia*, 13(1), Article 43. <https://doi.org/10.1186/s40468-023-00259-9>
- Shi, Z., & Li, S. (2024). From mock interviews to real job interviews: Assessing the efficacy of business Chinese instruction. *Foreign Language Annals*, 57(4), 962-980.

- <https://doi.org/10.1111/flan.12759>
- Silva, A. L. B. C. (2023). Aviation English assessment for military and civil pilots in Brazil: Similarities and differences. *Applied Linguistics Papers*, 27(2), 53-62.
<https://doi.org/10.32612/uw.25449354.2023.1.pp.53-62>
- Silva, A. L. B. C. & Raymundo, N. A. (2023). Test preparation issues in the aeronautical context in Brazil. In P. Tosqui-Lucks & J. C. Santana (Eds.), *Aviation English – A global perspective: Analysis, teaching, assessment* (pp. 172-182). Bookerfield Editors. <https://bookerfield.com/catalogo/10.53268/BKF22080500>
- Steyn, S. (2021). Pathways to parity between parallel tests of language ability: Lessons from a project. In A. Weideman, J. Read, & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 132-147). Multilingual Matters.
- Tarlani-Aliabadi, H., Tazik, K., & Azizi, Z. (2022). Exploring the role of language knowledge and background knowledge in reading comprehension of specific-purpose tests in higher education. *Language Testing in Asia*, 12, Article 48.
<https://doi.org/10.1186/s40468-022-00198-x>
- Tavakoli, P., Kendon, G., Mazhurnaya, S., & Ziomek, A. (2023). Assessment of fluency in the test of English for educational purposes. *Language Testing*, 40(3), 607-629.
<https://doi.org/10.1177/02655322231151384>
- Taylor, L., & Pill, J. (2024). Assessing health and other professionals. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 371–382). Wiley-Blackwell.
- Toprak, T. E., & Cakir, A. (2021). Examining the L2 reading comprehension ability of adult ELLs: Developing a diagnostic test within the cognitive diagnostic assessment framework. *Language Testing*, 38(1), 106–131.
<https://doi.org/10.1177/0265532220941470>
- Treadaway, M., & Read, J. (2024). Setting standards for a diagnostic test of aviation English for student pilots. *Language Testing*, 41(3), 557–583.
<https://doi.org/10.1177/02655322231224051>
- Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, 5, 1-15. <https://doi.org/10.22599/jesla.74>
- Tsagari, D., & Demetriou, T. (2022). Investigating the parameters accounting for language proficiency: The case of an English language test used for academic purposes. *Research Papers in Language Teaching and Learning*, 12(1), 89-110.
https://rplt.eap.gr/images/2022/01_loannou_Tsagari.pdf
- Van Dyk, T. (2021). Postscript: What the data tell us: an overview of language assessment research in South Africa's multilingual context. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 217–235). Multilingual Matters.
- Van Dyk, T., Murre, P., & Kotzé, H. (2021). Does one size fit all? Some considerations for test translation. In A. Weideman, J. Read & T. du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 52–74). Multilingual Matters.
- Wang, L., & Fan, J. (2021). *Working towards a proficiency scale of Business English writing: A mixed-methods approach*. Springer. <https://doi.org/10.1007/978-981-16-5449-7>
- Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programmes. *Journal of English for Academic Purposes*, 61, Article

101203. <https://doi.org/10.1016/j.jeap.2022.101203>
- Weideman, A. (2021). A skills-neutral approach to academic literacy assessment. In A. Weideman, J. Read, & T. du Plessis, T. (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 22-51). Multilingual Matters.
- Xia, T., Chen, X., Parsaei, H. R., & Qiu, F. (2023). An intelligent vocabulary size measurement method for second language learner. *Language Testing in Asia*, 13, Article 45 (2023). <https://doi.org/10.1186/s40468-023-00258-w>
- Yan, X., Dimova, S., & Ginther, A. (Eds.). (2023). *Local language testing: Practice across contexts*. Springer.
- Yang, C. (2023). Using eye-tracking and retrospective verbal reports to explore the cognitive processes of banked gap-filling: a case study featuring methodological triangulation. *Language Testing in Asia*, 13, Article 22 (2023). <https://doi.org/10.1186/s40468-023-00234-4>
- Yeager, R. R., Park, G., & Liao, R. J. T. (2024). Notetaking as Validity Evidence: A Mixed-Methods Investigation of Question Preview in Listening Assessment. *Journal of English for Academic Purposes*, 68(3), 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Yilmaz, Y., Arroyo, D., Carver, C., Choi, J., & Dibartolomeo, M. (2023). The effects of distributed versus massed corrective feedback on the acquisition of Spanish differential object marking. *Modern Language Journal*, 107, 756–781. <https://doi.org/10.1111/modl.12862>
- Yin, Y., Yu, G., & Huang, D. (2024). Implementing the ICAO language proficiency requirements: The localization of aviation English testing. *Applied Linguistics*, 45(4), 699-717. <https://doi.org/10.1093/applin/amad057>
- Yuksel, D., Soruç, A., Altay, M., & Curle, S. (2023). A longitudinal study at an English medium instruction university in Turkey: the interplay between English language improvement and academic success. *Applied Linguistics Review*, 14(3), 533-552. <https://doi.org/10.1515/applirev-2020-0097>
- Zhang, L., & Zhang, L. (2023). Development and validation of the questionnaire on EFL students' perceptions of authorial stance in academic writing. *Applied Linguistics Review*, 14(5), 1219–1251. <https://doi.org/10.1515/applirev-2021-0003>

Technology and Language Assessment

- Abdellatif, M. S., Alshehri, M. A., Alshehri, H. A., Hafez, W. E., Gafar, M. G., & Lamouchi, A. (2024). I am all ears: Listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement. *Language Testing in Asia*, 14, Article 54. <https://doi.org/10.1186/s40468-024-00329-6>
- Aladini, A., Mahmud, R., & Ali, A. A. H. (2024). The importance of needs satisfaction, teacher support, and L2 learning experience in Intelligent Computer-Assisted Language Assessment (ICALA): A probe into the state of willingness to communicate as well as academic motivation in EFL settings. *Language Testing in Asia*, 14, Article 58. <https://doi.org/10.1186/s40468-024-00334-9>
- Alam, S., Vadivel, B., Banu, S., & Jamalyar, R. (2024). Reflecting the voices of EFL teachers in the world of Intelligent Computer-Assisted Language Assessment (ICALA): An insight into teacher immunity, reflective teaching, job satisfaction, and

- L2-teacher grit. *Language Testing in Asia*, 14, Article 43.
<https://doi.org/10.1186/s40468-024-00314-z>
- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Albrecht, S. (2022). What do teachers do in the age of Zoom? The Covid-19 crisis and the role of the language teacher (Research Notes Issue 82). Cambridge University Press and Assessment.
<https://www.cambridgeenglish.org/Images/654641-research-notes-82.pdf>
- Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing*, 52, Article 100610.
<https://doi.org/10.1016/j.asw.2022.100610>
- An, N., Wright, C., & Wang, J. (2024). How does background knowledge affect second language reading? An eye movement study. *International Journal of Applied Linguistics*, 34(4), 1767–1789. <https://doi.org/10.1111/ijal.12601>
- Appel, R., Arias, A., Baker, B., & Loignon, G. (2023). Insights from lexical and syntactic analyses of a French for academic purposes assessment. *Assessing Writing*, 58, Article 100789. <https://doi.org/10.1016/j.asw.2023.100789>
- Aryadoust, V. (2024). Topic and accent coverage in a commercialized L2 listening test: Implications for test-takers' identity. *Applied Linguistics*, 45(5), 765-785.
<https://doi.org/10.1093/applin/amad062>
- Aryadoust, V. (2025). Generative Artificial Intelligence in assessment of listening. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from
<https://meteor.springer.com/container/contribute.jsf?id=263244>
- Aryadoust, V., & Yichen, J. (2025). ChatGPT in assessment of writing. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from
<https://meteor.springer.com/container/contribute.jsf?id=263243>
- Aryadoust, V., & Zakaria, A. (2025). ChatGPT in assessment of speaking. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from
<https://meteor.springer.com/container/contribute.jsf?id=263242>
- Aryadoust, V., Ng, L. Y., Foo, S., & Esposito, G. (2022). A neurocognitive investigation of test methods and gender effects in listening assessment. *Computer Assisted Language Learning*, 35(4), 743–763. <https://doi.org/10.1080/09588221.2020.1744667>
- Aryadoust, V., Zakaria, A., & Jia, Y. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence*. Advance online publication.
<https://doi.org/10.1016/j.caeai.2024.100204>
- Barkaoui, K., & Woodworth, J. (2023). An exploratory study of the construct measured by automated writing scores across task types and test occasions. *Studies in Language Assessment*, 12(1), 26–63. <https://doi.org/10.58379/QCFS2805>
- Barnau, A. (2023). Testing English for medical purposes: The effects of traditional and distance education on learning outcomes. *Language Learning in Higher Education*, 13(2), 451-460. <https://doi.org/10.1515/cercles-2023-2027>

- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Bennett, R. E., Zhang, M., Sinharay, S., Guo, H., & Deane, P. (2022). Are there distinctive profiles in examinee essay-writing processes? *Educational Measurement: Issues and Practice*, 41(2), 55–69. <https://doi.org/10.1111/emip.12469>
- Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: an exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14, Article 45. <https://doi.org/10.1186/s40468-024-00322-z>
- Bolender, B., Foster, C., & Vispoel, S. (2023). The criticality of implementing principled design when using AI technologies in test development. *Language Assessment Quarterly*, 20(4-5), 512-519. <https://doi.org/10.1080/15434303.2023.2288266>
- Bruce, E., & Clark, T. (2023). IELTS Online: Overview. *Research Reports*. <https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/ielts-online-overview-bruce-et-al-2023.pdf>
- Burton, J. D. (2024). Evaluating the impact of nonverbal behavior on language ability ratings. *Language Testing*, 41(4), 729–758. <https://doi.org/10.1177/02655322241255709>
- Cai, H., & Yan, X. (2024). Triangulating natural language processing (NLP)-based analysis of rater comments and many-facet Rasch measurement (MFRM): An innovative approach to investigating raters' application of rating scales in writing assessment. *Language Testing*, 41(2), 384–411. <https://doi.org/10.1177/02655322231210231>
- Çakmak, F., Ismail, S. M., & Karami, S. (2023). Advancing learning-oriented assessment (LOA): Mapping the role of self-assessment, academic resilience, academic motivation in students' test-taking skills, and test anxiety management in Telegram-assisted-language learning. *Language Testing in Asia*, 13, Article 20. <https://doi.org/10.1186/s40468-023-00230-8>
- Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, Article 102464. <https://doi.org/10.1016/j.system.2021.102464>
- Castillo, P. A. P., Soto, C., Asún, R. A., & Gutiérrez, F. (2023). Profiling support in literacy development: Use of natural language processing to identify learning needs in higher education. *Assessing Writing*, 58, Article 100787. <https://doi.org/10.1016/j.asw.2023.100787>
- Chapman, M., Kemp, J. A., Kim, A. A., MacGregor, D., & MacMillan, F. Testing young multilingual learners remotely: Prioritizing equity, fairness, and social justice. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 153-170). Routledge. <https://doi.org/10.4324/9781003292395-13>
- Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*, 57, Article 100915. <https://doi.org/10.1016/j.jslw.2022.100915>
- Cheng, Y. (2023). Concerto Software for Computerized Adaptive Testing - Free Version. *Measurement: Interdisciplinary Research and Perspectives*, 21(3), 194–202. <https://doi.org/10.1080/15366367.2023.2187274>

- Choi, I., & Deane, P. (2021). Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly*, 18(2), 107–132. <https://doi.org/10.1080/15434303.2020.1804913>
- Choi, I., Hao, J., Deane, P., & Zhang, M. (2021). Benchmark keystroke biometrics accuracy from high-stakes writing tasks. *ETS Research Report Series*, 2021(1), 1–13. <https://doi.org/10.1002/ets2.12326>
- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, Article 102408. <https://doi.org/10.1016/j.system.2020.102408>
- Chong, J. J. Q., & Aryadoust, V. (2023). Investigating the effect of multimodality and sentiments on speaking assessments: A facial emotional analysis. *Education and Information Technologies*, 28, 7413–7436. <https://doi.org/10.1007/s10639-022-11478-7>
- Chukharev-Hudilainen, E., & Ockey, G. J. (2021). The development and evaluation of interactional competence elicitor for oral language assessments. *ETS Research Report Series*, 2021(1), 1–20. <https://doi.org/10.1002/ets2.12319>
- Clark, T., & Endres, H. (2021). Online diagnostic grammar testing for high school students—An international trial. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 602–632. <https://doi.org/10.1080/0969594X.2021.1970513>
- Cohen, A. D., Rahmati, T., & Sadeghi, K. (2023). Test-taking strategies in technology-assisted language assessment. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated language assessment* (pp. 235–254). Routledge. <https://doi.org/10.4324/9781003292395-19>
- Davis, L., & Papageorgiou, S. (2021). Complementary strengths? Evaluation of a hybrid human-machine scoring approach for a test of oral academic English. *Assessment in Education: Principles, Policy & Practice*, 28(4), 437–455. <https://doi.org/10.1080/0969594X.2021.1979466>
- De Jong, N. H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456–476. <https://doi.org/10.1080/0969594X.2021.1951162>
- Drennan, L., Joubert, M., Weideman, A. & Posthumus, R. (2021). Combined measures of identifying language risk for first-year students. *Journal for Language Teaching*, 55(2), 93–116. <https://doi.org/10.4314/jlt.v55i2.4>
- Drown, L., Giovannone, N., Pisoni, D., & Theodore, R. (2023). Validation of two measures for assessing English vocabulary knowledge on web-based testing platforms: brief assessments. *Linguistics Vanguard*, 9(1), 99–111. <https://doi.org/10.1515/lingvan-2022-0116>
- Dunn, K. J. (2024). Random-item Rasch models and explanatory extensions: A worked example using L2 vocabulary test item responses. *Research Methods in Applied Linguistics*, 3(3), Article 100143. <https://doi.org/10.1016/j.rmal.2024.100143>
- Ebadi, S., Karimi, E., & Vakili, S. (2023). An exploration into EFL learners' perspectives on online computerized listening comprehension dynamic assessment. *Language Testing in Asia*, 13, Article 5 (2023). <https://doi.org/10.1186/s40468-023-00221-9>
- Eberharter, K., Kormos, J., Guggenbichler, E., Ebner, V. S., Suzuki, S., Moser-Frötscher, D., Konrad, E., & Kremmel, B. (2023). Investigating the impact of self-pacing on the L2 listening performance of young learner candidates with differing L1 literacy skills.

- Language Testing*, 40(4), 960-983. <https://doi.org/10.1177/02655322221149642>
- Ebrahimi, B., & Ebadi, S. (2024). Exploring the impact of screencast-based ipsative assessment on EFL students' writing development. *Language Testing in Asia*, 14, Article 41. <https://doi.org/10.1186/s40468-024-00315-y>
- El Ebyary, K., Shabara, R., & Boraie, D. (2024). The differential role of AI-operated WCF in L2 students' noticing of errors and its impact on writing scores. *Language Testing in Asia*, 14, Article 59. <https://doi.org/10.1186/s40468-024-00312-1>
- Elsayed, A. M., Kholikov, A., Abdullayeva, I., Al-Farouni, M., & Wodajo, M. R. (2024). Teacher support in AI-assisted exams: An experimental study to inspect the effects on demotivation, anxiety management in exams, L2 learning experience, and academic success. *Language Testing in Asia*, 14, Article 53. <https://doi.org/10.1186/s40468-024-00328-7>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14, Article 51. <https://doi.org/10.1186/s40468-024-00326-9>
- Falk, C. F., & Feuerstahler, L. M. (2022). On the performance of semi- and nonparametric item response functions in computer adaptive tests. *Educational and Psychological Measurement*, 82(1), 57–75. <https://doi.org/10.1177/00131644211014261>
- Fan, J., Jin, T., & Chen, I. (2022). *Applying the Lexile Framework to evaluate text complexity in the Chinese context: A corpus-based study* (Report 1/2022). https://www.britishcouncil.org/sites/default/files/rirg_fan_jin_chen_layout_1.pdf
- Feng, H.-H., & Chukharev-Hudilainen, E. (2022). Genre-based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology*, 26(2), 58–77. <https://doi.org/10.1257/73479>
- Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(1-2), 179-221. <https://doi.org/10.1080/09588221.2022.2033787>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- García Laborda, J., & Fernández Álvarez, M. (2021). Multilevel Language tests: Walking into the land of the unexplored. *Language Learning & Technology*, 25(2), 1–25. <https://hdl.handle.net/10125/73428>
- Geranpayeh, a. (2023). Leveraging computational psychometrics for language testing. *Language Teaching Research Quarterly*, 37, 204-212, <https://doi.org/10.32038/ltrq.2023.37.11>
- Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, 11, Article 27. <https://doi.org/10.1186/s40468-021-00143-4>
- Gong, T., Zhang, M., & Li, C. (2022). Association of keyboarding fluency and writing performance in online-delivered assessment. *Assessing Writing*, 51, Article 100575. <https://doi.org/10.1016/j.asw.2021.100575>
- Gràcia, M., Adam-Alcocer, A., & Castillo Mardones, P. (2023). Exploring the impact of a teacher development programme using a digital application on linguistic interactions in the classroom: a multiple case study. *Applied Linguistics Review*, 14(5),

- 1305–1343. <https://doi.org/10.1515/applirev-2020-0132>
- Graves, B. C. (2023). Specifications grading to promote student engagement, motivation and learning: Possibilities and cautions. *Assessing Writing*, 57, Article 100754. <https://doi.org/10.1016/j.asw.2023.100754>
- Green, B. A., & Lung, Y. S. M. (2021). English language placement testing at BYU-Hawaii in the time of COVID-19. *Language Assessment Quarterly*, 18(1), 6–11. <https://doi.org/10.1080/15434303.2020.1863966>
- Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58–76. <https://doi.org/10.1080/0969594X.2020.1735995>
- Ha, H. T., Nguyen, D. T. B., & Stoeckel, T. (2024). What is the best predictor of word difficulty? A case of data mining using random forest. *Language Testing*, 41(4), 828–844. <https://doi.org/10.1177/02655322241263628>
- Hannah, L., Kim, H., & Jang, E. E. (2022). Investigating the effects of task type and linguistic background on accuracy in automated speech recognition systems: Implications for use in language assessment of young learners. *Language Assessment Quarterly*, 19(3), 289–313. <https://doi.org/10.1080/15434303.2022.2038172>
- Hayashi, Y., Kondo, Y., & Ishii, Y. (2023). Automated speech scoring of dialogue response by Japanese learners of English as a foreign language. *Innovation in Language Learning and Teaching*, 18(1), 32–46. <https://doi.org/10.1080/17501229.2023.2217181>
- Huang, H. T. D., Hung, S. T. A., Chao, H. Y., Chen, J. H., Lin, T. P., & Shih, C. L. (2022). Developing and validating a computerized adaptive testing system for measuring the English proficiency of Taiwanese EFL university students. *Language Assessment Quarterly*, 19(2), 162–188. <https://doi.org/10.1080/15434303.2021.1984490>
- Hunte, M. R., McCormick, S., Shah, M., Lau, C., & Jang, E. E. (2021). Investigating the potential of NLP-driven linguistic and acoustic features for predicting human scores of children's oral language proficiency. *Assessment in Education: Principles, Policy & Practice*, 28(4), 477–505. <https://doi.org/10.1080/0969594X.2021.1999209>
- Hwang, H., & Kim, H. (2024). Korean Syntactic Complexity Analyzer (KOSCA): An NLP application for the analysis of syntactic complexity in second language production. *Language Testing*, 41(3), 506–529. <https://doi.org/10.1177/02655322231222596>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines”. *Language Testing in Asia*, 13, Article 46 (2023). <https://doi.org/10.1186/s40468-023-00260-2>
- Ibrahim, K. A. A.-A., Kassem, M. A. M., & Lami, D. (2024). Intelligent Computer-Assisted Language Assessment (ICALA) in philosophy-based language instruction: Unraveling the effects on critical thinking, self-evaluation, academic resilience, and speaking development. *Language Testing in Asia*, 14, Article 46. <https://doi.org/10.1186/s40468-024-00320-1>
- Imperial, J. M., Forey, G., & Tayyar Madabushi, H. (2024). Standardize: Aligning Language Models with Expert-Defined Standards for Content Generation. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1573–1594). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.94>

- Isbell, D. R., Kim, K. M., & Chen, X. (2023). Exploring the potential of automated speech recognition for scoring the Korean Elicited Imitation Test. *Research Methods in Applied Linguistics*, 2(3), Article 100076. <https://doi.org/10.1016/j.rmal.2023.100076>
- Jalilzadeh, K., Rashtchi, M., & Mirzapour, F. (2024). Cheating in online assessment: A qualitative study on reasons and coping strategies focusing on EFL teachers' perceptions. *Language Testing in Asia*, 14, Article 29. <https://doi.org/10.1186/s40468-024-00304-1>
- Jin, Y., & Fan, J. (2023). Test-taker engagement in AI technology-mediated language assessment. *Language Assessment Quarterly*, 20(4-5), 488–500. <https://doi.org/10.1080/15434303.2023.2291731>
- Johnson, C., Cardoso, W., Zuercher, B., Brannen, K., & Springer, S. (2024). Assessing pronunciation using dictation tools: The use of Google Voice Typing to score a pronunciation placement test. *Journal of Second Language Pronunciation*, 10(1), 10–34. <https://doi.org/10.1075/jslp.23033.joh>
- Joo, S. H., Seong, Y., Suh, J., Jung, J.-Y., & Purpura, J. E. (2023). Assessing Korean writing ability through a scenario-based assessment approach. *Assessing Writing*, 58, Article 100766. <https://doi.org/10.1016/j.asw.2023.100766>
- Kang, O., Hirschi, K., Miao, Y., Ahn, H., & Won, Y. (2022). Test-takers' IELTS preparations and their attitudes towards IELTS practices and the use of technologies in the global pandemic. IELTS Research Reports Online Series, No. 2/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/kang-et-al-2022.ashx>
- Karakolidis, A., O'Leary, M., & Scully, D. (2021). Animated videos in assessment: comparing validity evidence from and test-takers' reactions to an animated and a text-based situational judgment test. *International Journal of Testing*, 21(2), 57–79. <https://doi.org/10.1080/15305058.2021.1916505>
- Karatay, Y. (2023). Using spoken dialog systems to assess L2 learners' oral skills in a local language testing context. In Y. Xun, D. Slobodanka & J. G. April (Eds.), *Local language testing: Practice across contexts* (pp. 231–252). Springer. https://link.springer.com/chapter/10.1007/978-3-031-33541-9_12
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, 103332. <https://doi.org/10.1016/j.system.2024.103332>
- Karatay, Y., & Karatay, L. (2024). The impact of automated writing evaluation on language learning and teaching: A research synthesis. *System*, 123. <https://doi.org/10.1016/j.system.2024.103332>
- Kazemian, M. & Khonamri, F. (2024). The online assessment in the Middle East. A conversation with Dr. Christine Coombe, *MEXTESOL Journal*, 48(1), 1-6.
- Khabbazzbashi, N., Chan, S. and Clark, T. (2023). Towards the new construct of academic English in the digital age, *ELT Journal*, 77(2), 207–216 <https://doi.org/10.1093/elt/ccac010>
- Khabbazzbashi, N., Xu, J., & Galaczi, E. (2021). Opening the black box: Exploring automated speaking assessment. In B. Lantaigne, C. Coombe & J.D. Brown (Eds.), *Issues in language testing around the world* (pp.333–344). Springer. https://doi.org/10.1007/978-981-33-4232-3_25
- Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support,

- immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, 14, Article 48. <https://doi.org/10.1186/s40468-024-00324-x>
- Khodi, A., Khezerlou, H., & Sahraei, H. (2024). Dependability and utility of using e-portfolios in assessing EFL learners' speaking proficiency. *Computer Assisted Language Learning*, 37(7), 1579-1601. <https://doi.org/10.1080/09588221.2022.2093379>
- Kim, A. A., Tywoniu, R. L., & Chapman, M. (2022). Technology-enhanced items in Grades 1–12 English language proficiency assessments. *Language Assessment Quarterly*, 19(4), 343–367. <https://doi.org/10.1080/15434303.2022.2039659>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73–95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Kim, K. M., Chen, X., & Liu, X. (2024). Accuracy scoring of elicited imitation: A tutorial of automating speech data with commercial NLP support. *Research Methods in Applied Linguistics*, 3(3), Article 100127. <https://doi.org/10.1016/j.rmal.2024.100127>
- Kim, P., Beltrán, J., & Joo, S. H. (2022). An interview with APPLE Lecture speaker Dr. Paul D. Deane. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University*, 22(1), 47–55. <https://doi.org/10.52214/salt.v22i1.9901>
- Kılıçkaya, F. (2022). Pre-service EFL teacher candidates' concerns over digital language assessment and test anxiety. In W. M. El-Henawy & M. del Mar Suárez (Eds.), *English as a foreign language in a new-found post-pandemic world* (pp. 164-181). IGI Global. <https://doi.org/10.4018/978-1-6684-4205-0>
- Kollias, C. (2023). *Virtual standard setting: Setting cut scores*. Peter Lang. <https://doi.org/10.3726/b20407>
- Koster, T., & Morgan, G. (2023). Investigating students' perceptions of effective engagement methods for online and multi-modal ELICOS classes. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 43–60). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Kunnan, A. J., Qin, C. Y., & Zhao, C. G. (2022). Developing a scenario-based English language assessment in an Asian university. *Language Assessment Quarterly*, 19(4), 368–393. <https://doi.org/10.1080/15434303.2022.2073886>
- Kwon, S. K., & Yu, G. (2024). The effect of viewing visual cues in a listening comprehension test on second language learners' test-taking process and performance: An eye-tracking study. *Language Testing*, 41(3), 649–680. <https://doi.org/10.1177/02655322241239356>
- Kyle, K., Choe, A. T., Eguchi, M., LaFlair, G., & Ziegler, N. (2021). A comparison of spoken and written language use in traditional and technology-mediated learning environments. *ETS Research Report Series*, 2021(1), 1–29. <https://doi.org/10.1002/ets2.12329>
- Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62–85. <https://doi.org/10.1177/0265532220929918>
- Latimer, N., & Chan, S. H. C. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: benefits and challenges. *International Journal of TESOL Studies*, 4(1),

- 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Lee, H.-W., & Galaczi, E. (Eds.). (2023). *Exploring video-call paired speaking tests*. Research Notes, 86. Retrieved from: <https://www.cambridgeenglish.org/english-research-group/published-research/research-notes/>
- Lee, H.-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring interaction in video-call paired speaking tests: A look at scores, language, and perceptions. *Applied Linguistics*, 45(3), 533-552. <https://doi.org/10.1093/applin/amad046>
- Li, A. W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57, Article 100746. <https://doi.org/10.1016/j.asw.2023.100746>
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, Article 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Lim, J., Whitehead, G. E. K., & Choi, Y. (2021). Interactive e-book reading vs. paper-based reading: Comparing the effects of different mediums on middle school students' reading comprehension. *System*, 97, Article 102434. <https://doi.org/10.1016/j.system.2020.102434>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 123, 103344. <https://doi.org/10.1016/j.system.2024.103344>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liou, G., Bonner, C. V., & Tay, L. (2022). A psychometric view of technology-based assessments. *International Journal of Testing*, 22(3-4), 216-242. <https://doi.org/10.1080/15305058.2022.2070757>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78–105. <https://doi.org/10.1257/73480>
- Liu, T. & Aryadoust, V. (2022). The effect of in-class and one-on-one video feedback on EFL learners' English public speaking competency and anxiety. *Studies in Language Assessment*, 11(2), 25–57. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4396264/SiLA_11.2_Liu_and_Aryadoust.pdf
- Lopez, A. A., Guzman-Orth, D., Zapata-Rivera, D., Forsyth, C. M., & Luce, C. (2021). Examining the accuracy of a conversation-based assessment in interpreting English learners' written responses. *ETS Research Report Series*, 2021(1), 1–15. <https://doi.org/10.1002/ets2.12315>
- Low, A. R. L., & Aryadoust, V. (2023). Investigating test-taking strategies in listening assessment: A comparative study of eye-tracking and self-report questionnaires. *International Journal of Listening*, 37(2), 93-112. DOI: 10.1080/10904018.2021.1883433
- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49, Article 100529. <https://doi.org/10.1016/j.asw.2021.100529>
- Ma, H., Wang, J., & He, L. (2024). Linguistic features distinguishing students' writing ability

- aligned with CEFR levels. *Applied Linguistics*, 45(4), 637–657.
<https://doi.org/10.1093/applin/amad054>
- Mäkipää, T., Hahl, K., & Luodonpää-Manni, M. (2021). Teachers' perceptions of assessment and feedback practices in Finland's foreign language classes during the Covid-19 pandemic. *CEPS Journal*, 11, 219–240. <https://doi.org/10.26529/cepsj.1108>
- Martin, J. L., & Wright, K. E. (2023). Bias in automatic speech recognition: The case of African American language. *Applied Linguistics*, 44(4), 613–630.
<https://doi.org/10.1093/applin/amac066>
- McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, 52, Article 100608.
<https://doi.org/10.1016/j.asw.2022.100608>
- McNamara, D. S., & Kendeou, P. (2022). The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*, 29(2), 150–182.
<https://doi.org/10.1080/0969594X.2022.2037509>
- Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2021). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 34(7), 868–897.
<https://doi.org/10.1080/09588221.2019.1645181>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.
- Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), Article 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2021). Video-conferencing speaking tests: Do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice*, 28(4), 369–388.
<https://doi.org/10.1080/0969594X.2021.1951163>
- Nakatsuhara, F., May, L., Inoue, C., Willcox-Ficzere, E., Westbrook, C., & Spiby, R. (2021). *Exploring the potential for assessing interactional and pragmatic competence in semi-direct speaking tests* (Report No. VS/2021/001). British Council.
https://www.britishcouncil.org/sites/default/files/nakatsuhara_et_al_2021_-_assessing_interactional_and_pragmatic_competence_in_semi-direct_speaking_tests_v2.pdf
- Namkung, Y., & Kim, Y. (2024). Learner engagement in collaborative writing: The effects of SCMC mode, interlocutor familiarity, L2 proficiency, and task repetition. *System*, 121, 103251. <https://doi.org/10.1016/j.system.2024.103251>
- Nguyen L., Yuan Z., & Seed G. (2022). Building educational technologies for code-switching: Current practices, difficulties and future directions. *Languages*, 7, Article 220. <https://doi.org/10.3390/languages7030220>
- Nguyen, T. H. H., Nguyen, B. T. T., Hoang, G. T. L., Pham, N. T. H., & Dang, T. T. C. (2024). Computer-delivered vs. face-to-face score comparability and test takers' perceptions: The case of the two English speaking proficiency tests for Vietnamese EFL learners. *Language Testing in Asia*, 14, Article 6.
<https://doi.org/10.1186/s40468-024-00277-1>
- O'Sullivan, B. (2023). Integrating a Remote Proctoring Solution: the case of Aptis Remote. In H. Işık, E. Kantarcıoğlu, A. Üner, & C. Thomas (Eds.), *Re-Vision and Re-Form in*

- English for Academic Purposes Contexts after the Pandemic* (pp. 75-82). Cambridge Scholars.
- O'Sullivan, B. (2023). Reflections on the application and validation of technology in language testing. *Language Assessment Quarterly*, 20(4-5), 501-511. <https://doi.org/10.1080/15434303.2023.2291486>
- O'Sullivan, B., Breakspear, T., & Bayliss, W. (2023). Validating an AI-Driven Scoring System: The Model Card Approach. In Karim Sadeghi & Dan Douglas (eds), *Fundamental Considerations in Technology Mediated Language Assessment*, (pp.115-134). New York: Routledge. DOI: 10.4324/9781003292395-10
- Ockey, G. J., & Chukharev-Hudilainen, E. (2021). Human versus computer partner in the paired oral discussion test. *Applied Linguistics*, 42(5), 924–944. <https://doi.org/10.1093/applin/amaa067>
- Ockey, G. J., & Neiriz, R. (2021). Evaluating technology-mediated second language oral communication assessment delivery models. *Assessment in Education: Principles, Policy & Practice*, 28(4), 350–368. <https://doi.org/10.1080/0969594X.2021.1976106>
- Ockey, G. J., Muhammad, A. A., Prasetyo, A. H., Elnegahy, S., Kochem, T., Neiriz, R., Kim, H., & Beck, J. (2021). Iowa State University's English placement test of oral communication in times of COVID-19. *Language Assessment Quarterly*, 18(1), 26–35. <https://doi.org/10.1080/15434303.2020.1862122>
- Oh, S. (2022). The use of spelling and reference tools in second language writing: Their impact on students' writing performance and process. *Journal of Second Language Writing*, 57, Article 100916. <https://doi.org/10.1016/j.jslw.2022.100916>
- Osidak, V., Kvasova, O., Kavytska, T., & Drobotun, V. (2024). Book Creator as a tool of formative assessment in the TEFL classroom: Emergency remote teaching during wartime. *Studies in Language Assessment (SILA)*, 13(2), 1-37. <https://doi.org/10.58379/RBXT4093>
- Ozer, O. (2024). AI language models: A breach of academic integrity in online language learning? *Studies in Language Assessment (SILA)*, 13(1), 237-262. <https://doi.org/10.58379/EZYS5680>
- Papageorgiou, S., & Manna, V. F. (2021). Maintaining access to a large-scale test of academic language proficiency during the pandemic: The launch of TOEFL iBT Home Edition. *Language Assessment Quarterly*, 18(1), 36–41. <https://doi.org/10.1080/15434303.2020.1864376>
- Peng, Y., Sun, J., Quan, J., Wang, Y., Lv, C., & Zhang H. (2023). Predicting Chinese EFL Learners' Human-rated Writing Quality in Argumentative Writing Through Multidimensional Computational Indices of Lexical Complexity. *Assessing Writing*, 56, Article 100722. <https://doi.org/10.1016/j.asw.2023.100722>
- Perry, S. (2024). Podcasting as a form of assessment: Increasing student motivation in academic English-speaking assessment. *Language Testing in Asia*, 14, Article 15. <https://doi.org/10.1186/s40468-024-00288-y>
- Piyumi Udeshinee, W. A., Knutsson, O., Barbutiu, S. M., & Jayathilake, C. (2024). Re-designing a regulatory scale for dynamic assessment in the synchronous text chat environment in collaboration with teachers. *Computer Assisted Language Learning*, 37(7), 1527-1553. <https://doi.org/10.1080/09588221.2022.2092153>
- Prabowo, M. Y. (2022). Assessing equivalency between paper-and-pencil and computer-based English competency test. In R. E. Capel & Q. S. Hoon (Eds.), *Conference proceedings iELT-CON2021* (pp. 42–51). Penang English Language

- Learning and Teaching Association.
https://pellta.org/uploads/8/1/8/9/81899300/ielc-con_2021_proceedings.pdf
- Prabowo, M. Y., & Rahmadian, S. (2022). Computer-based English competency assessment for scholarship selection: Challenges, strategies, and implementation in the Ministry of Finance. *Jurnal Siositeknologi*, 21(1), 84–96.
<https://doi.org/10.5614/sostek.itbj.2022.21.1.9>
- Prasad, K., Aladini, A., Normurodova, N. Z., & Belton, B. (2023). Take language assessment easy: the mediator impacts of self-assessment, test-taking skills in predicting student evaluation apprehension, foreign language learning self-esteem, and language achievement in online classes. *Language Testing in Asia*, 13, Article 34 (2023). <https://doi.org/10.1186/s40468-023-00246-0>
- Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of an online language placement exam. *Language Assessment Quarterly*, 18(1), 42–50.
<https://doi.org/10.1080/15434303.2020.1867145>
- Qin, T., & van Compernelle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, 25(2), 55–74.
<http://hdl.handle.net/10125/73433>
- Qiu, Y., & Aryadoust, V. (2024). The predictive value of gaze behavior and mouse-clicking in testing listening proficiency: A sensor technology study. *System*, 126, 103440.
<https://doi.org/10.1016/j.system.2024.103440>
- QUEVEDO-CAMARGO, G., & TONELLI, J. A. R. (2021) Contribuições de um curso EAD para o letramento em avaliação de professores de línguas adicionais para crianças [Contributions of a distance-mode course on assessment literacy for teachers of additional languages to young learners]. *Revista do GEL*, v. 18, p. 230-248. DOI: <https://doi.org/10.21165/gel.v18i1.3094>
- Rafatbakhsh, E., Ahmadi, A., Moloodi, A., & Mehrpour, S. (2021). Development and validation of an automatic item generation system for English idioms. *Educational Measurement: Issues and Practice*, 40(2), 49–59.
<https://doi.org/10.1111/emip.12401>
- Rafi, F., & Pourdana, N. (2023). E-diagnostic assessment of collaborative and individual oral tiered task performance in differentiated second language instruction framework. *Language Testing in Asia*, 13, Article 6 (2023).
<https://doi.org/10.1186/s40468-023-00223-7>
- Rahmati, T., & Nushi, M. (2023). EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography. *Iranian Journal of Language Teaching Research*, 11(3), 31-52.
<https://doi.org/10.30466/ijltr.2023.121405>
- Rahmati, T., & Sadeghi, K. (2024). Technological innovations used in researching language assessment: A scoping review. In K. Sadeghi (Ed.), *Routledge handbook of technological advances in researching language learning* (pp. 118-129). <https://doi.org/10.4324/9781003459088-11>
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, Article 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Randall, T. S., & Urbanski, K. (2023). Development of a Computerized Dynamic Assessment Program for Second Language Grammar Instruction and Assessment. *Language and Sociocultural Theory*, 10(1), 50-81. <https://doi.org/10.1558/ist.21006>

- Revesz, A., Michel, M., Lu, X., Kourtali., Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, 58, Article 100927. <https://doi.org/10.1016/j.jslw.2022.100927>
- Rezadoust Siah Khaleh Sar, H., Khalili Sabet, M., Mahdavi Zafarghandi, A., & Hassaskhah, J. (2024). Iranian teachers' and learners' attitudes towards the use of e-portfolio in second language speaking assessment. *Language Testing in Asia*, 14, Article 61. <https://doi.org/10.1186/s40468-024-00335-8>
- Ritonga, M., Shaban, A. A., Al-Rashidi, A. H., & Chilani, N. (2023). Engagement in on-line language assessment: Are test-taking skills, self-assessment, resilience, and autonomy critical?. *Language Testing in Asia*, 13, Article 25 (2023). <https://doi.org/10.1186/s40468-023-00236-2>
- Ruegg, R., & Prommas, P. (2024). Comparing automated analysis and human analysis of hedging in academic texts. *Studies in Language Assessment (SILA)*, 13(1), 263-287. <https://doi.org/10.58379/SLVL8621>
- Sadeghi, K. (Ed.). (2022). *Technology-assisted language assessment in diverse contexts: Lessons from the transition to online testing during COVID-19*. Routledge. <https://doi.org/10.4324/9781003221463>
- Sadeghi, K. & Douglas, D. (Eds.). (2023). *Fundamental considerations in technology mediated language assessment*. Routledge. <https://doi.org/10.4324/9781003292395>
- Saeed, M. A., & Alharbi, M. A. (2023). Towards fostering Saudi EFL learners' collaborative engagement and feedback literacy in writing. *Assessing Writing*, 56, Article 100721. <https://doi.org/10.1016/j.asw.2023.100721>
- Sawaki, Y., Ishii, Y., & Oi, Y. (2024). The efficacy of automated feedback on summary content in an online formative assessment module for undergraduate L2 writing instruction in Japan. *Studies in Language Assessment (SILA)*, 13(2), 38–72. <https://doi.org/10.58379/mjfp3942>
- Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, 14, Article 50. <https://doi.org/10.1186/s40468-024-00325-w>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14, Article 49. <https://doi.org/10.1186/s40468-024-00321-0>
- Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to Generate Reading Comprehension Items. *Educational Measurement: Issues and Practice*, 43(1), 5-18. <https://doi.org/10.1111/emip.12590>
- Screti, F. (2023). Using google docs for guided academic writing assessments: Students' perspectives. *Language Learning in Higher Education*, 13(1), 153-174. <https://doi.org/10.1515/cercles-2023-2007>
- Seedhouse, P., & Satar, M. (2021). VEO IELTS Project Report: Which specific features of candidate talk do examiners orient to when taking scoring decisions? *IELTS Research Reports Online Series*, 5. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/for-researchers/research-reports/online-series-2021-5>

- Shi, X., Wang, X., & Zhang, W. (2024). Exploring the relationships between ASS indices and CAF and the impact on Chinese college students' oral English performance. *Language Testing in Asia*, 14, Article 30. <https://doi.org/10.1186/s40468-024-00306-z>
- Shi, Z., Liu, F., Lai, C., & Jin, T. (2022). Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback. *Language Learning & Technology*, 26(2), 106–128. <https://doi.org/10.125/73481>
- Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts? *Language Learning & Technology*, 27(3), 27–40. <https://hdl.handle.net/10125/73530>
- Shin, D., & Lee, J. H. (2024). AI-powered automated item generation for language testing. *ELT Journal*, 78(4), 446–452. <https://doi.org/10.1093/elt/ccae016>
- Song, Y., Sparks, J. R., & Guzman-Orth, D. (2021). Investigating student performance on an assessment tool to support the development of English learners' argumentation skills. *Computer Assisted Language Learning*, 35(9), 2291–2311. <https://doi.org/10.1080/09588221.2021.1879160>
- Spino, L. L., Echevarría, M. M., & Wu, Y. (2022). Assisted self-assessment to optimize the OPIc test experience. *Foreign Language Annals*, 55(3), 853–876. <https://doi.org/10.1111/flan.12645>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Suvorov, R., Karatay, Y., & Hegelheimer, V. (2024). Computer-Assisted Language Testing. In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment*. Wiley.
- Suzumura, N. (2022). Content analysis of test taker responses on an AP Japanese computer-simulated conversation test: A mixed methods approach for a validity argument. *Language Assessment Quarterly*, 19(5), 480–502. <https://doi.org/10.1080/15434303.2022.2130326>
- Tabari, M. A., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing*, 57, Article 100749. <https://doi.org/10.1016/j.asw.2023.100749>
- Tabari, M. A., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing*, 58, Article 100783. <https://doi.org/10.1016/j.asw.2023.100783>
- Taguchi, N. (2022). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, 56(1), 308–335. <https://doi.org/10.1002/tesq.3070>
- Taguchi, N. (2023). Using immersive virtual reality for the assessment of intercultural conflict mediation. *Language Learning & Technology*, 27(3), 108–128. <https://hdl.handle.net/10125/73534>
- Thi, N. K., Nikolov, M., & Simon, K. (2023). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 17(3), 690–705. <https://doi.org/10.1080/17501229.2022.2122476>
- Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2024). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*, 37(1-2), 149–178.

- <https://doi.org/10.1080/09588221.2022.2032184>
- Tsagari, D. & H. Bøhn (eds.) (2022) Recent Developments in Language Testing and Assessment (Special Issue) Issue 7. Languages (open access journal) https://www.mdpi.com/journal/languages/special_issues/Language_Testing_Assessment?fbclid=IwAR2_A6SIGEKehpZ95o5AQvPmM3kzi6O3rqypHjoTPA22LIdl6F_dHJLpnas
- Valkova, R. B. (2024). Remote Proctoring in language assessment: Exploring the impact on test-takers' scores and perceptions. *Studies in Language Assessment (SILA)*, 13(1), 141-198. <https://doi.org/10.58379/FCZM2855>
- Vo, Y., Rickels, H., Welch, C., & Dunbar, S. (2023). Human scoring versus automated scoring for English learners in a statewide evidence-based writing assessment. *Assessing Writing*, 56, Article 100719. <https://doi.org/10.1016/j.asw.2023.100719>
- Voss, E. (2023). Proctoring remote language assessments. In K. Sadeghi (Eds.), *Fundamental considerations in technology mediated language assessment*. Routledge. <https://doi.org/10.4324/9781003292395-15>
- Voss, E. (2024). Language assessment and artificial intelligence. In A. Kunnan (Ed.) *The Concise Companion to Language Assessment* (pp. 112–125). John Wiley & Sons.
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4-5), 520-532. <https://doi.org/10.1080/15434303.2023.2288256>
- Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 17(1), 47–71. <https://doi.org/10.1080/17501229.2021.1914062>
- Wagner, E., & Krylova, A. (2021). Temple University's ITA placement test in times of COVID-19. *Language Assessment Quarterly*, 18(1), 12–16. <https://doi.org/10.1080/15434303.2020.1862849>
- Wang, T-H., Sun, Y., & Huang, N-W. (2023). Implementation of web-based dynamic assessment in improving low English achiever's learning effectiveness. *Computer Assisted Language Learning*, 36(8), 1450–1476. <https://doi.org/10.1080/09588221.2021.1998129>
- Wang, Y., Xu, M., Jiang, Y., & Jin, T. (2023). Connecting form with function: Model texts for bilingual learners' narrative writing. *Assessing Writing*, 57, Article 100753. <https://doi.org/10.1016/j.asw.2023.100753>
- Wilson, J., Potter, A., Cordero, T. C., & Myers, M. C. (2023). Integrating goal-setting and automated feedback to improve writing outcomes: A pilot study. *Innovation in Language Learning and Teaching*, 17(3), 518–534. <https://doi.org/10.1080/17501229.2022.2077348>
- Wolf, M. K., & Lopez, A. (2022). Developing a technology-based classroom assessment of academic reading skills for English language learners and teachers: validity evidence for formative use. *Languages*, 7(2), Article 71. <https://doi.org/10.3390/languages7020071>
- Wolf, M. K., & Oh, S. (2024). Investigating the types and use of feedback in middle-school English language learners' academic writing. *Language Learning & Technology*, 28(1), 1-25. <https://hdl.handle.net/10125/73583>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). Exploring the provision of AI-based feedback for TOEFL Junior Writing practice tasks (Research Memorandum No. RM

- 24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>
- Wu, C., & Gao, F. (2023). Comparing the Intelligibility of Different Varieties of English Through Predicting Learners Comprehension: A Phonetic Experimental Approach. *Chinese Journal of Applied Linguistics*, 46(1), 19-30. <https://doi.org/10.1515/CJAL-2023-0102>
- Wu, Y-j. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology*, 25(2), 192–214. <http://hdl.handle.net/10125/73440>
- Xi, X. (2023). Advancing Language Assessment with AI and ML-Leaning into AI is Inevitable, but Can Theory Keep Up? *Language Assessment Quarterly*, 20(4-5), 357-376. <https://doi.org/10.1080/15434303.2023.2291488>
- Xu, J., Jones, E., Laxton, V., & Galaczi, E. (2021). Assessing L2 English speaking using automated scoring technology: Examining automarker reliability. *Assessment in Education: Principles, Policy & Practice*, 28(4), 411–436. <https://doi.org/10.1080/0969594X.2021.1979467>
- Xu, J., Schmidt, E., Galaczi, E., & Somers, A. (2024). Automarking in language assessment: Key considerations for best practice. Cambridge Papers in English Language Education. Cambridge University Press & Assessment. <https://cambunipress.prod.acquia-sites.com/sites/default/files/media/documents/cambridge-research-papers-in-english-language-education.pdf>
- Yamashita, T. (2024). An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, 3(3), Article 100133. <https://doi.org/10.1016/j.rmal.2024.100133>
- Yang, X., & Osborne, C. (2023). The development and validation of a C-test and a pseudo-character test for online CFL assessment. *Journal of China Computer-Assisted Language Learning*, 3(1), 101-131. <https://doi.org/10.1515/jccall-2022-0019>
- Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, 50, Article 100568. <https://doi.org/10.1016/j.asw.2021.100568>
- Yeldham, M., & Chen, R. T.-H. (2024). Conducting L2 listener verbal reports: Common missteps by novice researchers and ways to address them. *Research Methods in Applied Linguistics*, 3(2), Article 100113. <https://doi.org/10.1016/j.rmal.2024.100113>
- Yu, G., & Xu, J. (Eds.). (2024). Language test validation in a digital age. Cambridge University Press & Assessment.
- Yu, W., & Iwashita, N. (2021). Comparison of test performance on paper-based testing. (PBT). and computer-based testing. (CBT). by English-majored undergraduate students in China. *Language Testing in Asia*, 11, Article 32. <https://doi.org/10.1186/s40468-021-00147-0>
- Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, 35(9), 2817-2842. <https://doi.org/10.1080/09588221.2021.1897019>
- Zhai, X., & Aryadoust, V. (2022). The metacognitive and neurocognitive signature of test methods in academic listening assessment. *Frontiers in Psychology*, 13. Article 930075. <https://doi.org/10.3389/fpsyg.2022.930075>
- Zhang, E. D., & Yu, S. (2024). Understanding L2 student writers' self-assessment in digital multimodal composing: A process-oriented approach. *System*, 121, 103219.

- <https://doi.org/10.1016/j.system.2024.103219>
- Zhang, L., & Jin, Y. (2021). Assessing interactional competence in the computer-based CET-SET: An investigation of the use of communication strategies. *Assessment in Education: Principles, Policy & Practice*, 28(4), 389–410. <https://doi.org/10.1080/0969594X.2021.1976107>
- Zhang, L., Zheng, Y., O'Sullivan, B., & Stanley, G. (2023). Teacher perceptions of learner engagement with a new online Llearning environment. In H. Işıl Işık, E. Kantarcıoğlu, A. Üner & C. Thomas (Eds.), *Re-vision and re-form in English for academic purposes contexts after the Pandemic* (pp. 124 –132). Cambridge Scholars.
- Zhang, S., & Isaacs, T. (2023). Can interactions happen across the screens? The use of videoconferencing technology in assessing second language pragmatic competence. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts* (pp. 196–211). Routledge.
- Zhang, Y., & MacWhinney, B. (2023). Using diagnostic feedback to enhance the development of phonetic knowledge of an L2: A CALL design based on the unified competition model and the implementation with the Pinyin Tutor. *Language Testing in Asia*, 13, Article 35 (2023). <https://doi.org/10.1186/s40468-023-00232-6>
- Zhang, Y., & Wu, W. (2021). How effective are lexical richness measures for differentiations of vocabulary proficiency? A comprehensive examination with clustering analysis. *Language Testing in Asia*, 11, Article 15. <https://doi.org/10.1186/s40468-021-00133-6>
- Zhao, J.-H., Chen, Z.-W., & Yang, Q.-F. (2024). I do and I understand: A virtual reality-supported collaborative design-assessing activity for EFL students. *System*, 121, 103213. <https://doi.org/10.1016/j.system.2023.103213>
- Zheng, C., Chen, X., Zhang, H., & Chai, C. S (2024). Automated versus peer assessment: Effects on learners' English public speaking. *Language Learning & Technology*, 28(2), 210–228. <https://hdl.handle.net/10125/73577>
- Zhu, X., & Aryadoust, V. (2022). An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test. *Computer Assisted Language Learning*, 35(3), 412-436. <https://doi.org/10.1080/09588221.2019.1704788>

Artificial Intelligence and Language Assessment

- Abdellatif, M. S., Alshehri, M. A., Alshehri, H. A., Hafez, W. E., Gafar, M. G., & Lamouchi, A. (2024). I am all ears: Listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement. *Language Testing in Asia*, 14, Article 54. <https://doi.org/10.1186/s40468-024-00329-6>
- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Aryadoust, V. (2025). Generative Artificial Intelligence in assessment of listening. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*. Palgrave-Springer. Retrieved from

- <https://meteor.springer.com/container/contribute.jsf?id=263244>
- Aryadoust, V., Zakaria, A., & Jia, Y. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence*. Advance online publication. <https://doi.org/10.1016/j.caeai.2024.100204>
- Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: an exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14, Article 45. <https://doi.org/10.1186/s40468-024-00322-z>
- Bolender, B., Foster, C., & Vispoel, S. (2023). The criticality of implementing principled design when using AI technologies in test development. *Language Assessment Quarterly*, 20(4-5), 512-519. <https://doi.org/10.1080/15434303.2023.2288266>
- Elsayed, A. M., Kholikov, A., Abdullayeva, I., Al-Farouni, M., & Wodajo, M. R. (2024). Teacher support in AI-assisted exams: An experimental study to inspect the effects on demotivation, anxiety management in exams, L2 learning experience, and academic success. *Language Testing in Asia*, 14, Article 53. <https://doi.org/10.1186/s40468-024-00328-7>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The Terminator Versus the Machines". *Language Testing in Asia*, 13, Article 46 (2023). <https://doi.org/10.1186/s40468-023-00260-2>
- Jin, Y., & Fan, J. (2023). Test-taker engagement in AI technology-mediated language assessment. *Language Assessment Quarterly*, 20(4-5), 488-500. <https://doi.org/10.1080/15434303.2023.2291731>
- Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, 14, Article 48. <https://doi.org/10.1186/s40468-024-00324-x>
- Kim, H., Baghestani, Sh., Yin, Sh., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 73-95). Iowa State University Digital Press. <https://www.iastatedigitalpress.com/plugins/books/154/chapter/1230>
- Li, A. W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57, Article 100746. <https://doi.org/10.1016/j.asw.2023.100746>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.
- O'Sullivan, B., Breakspear, T., & Bayliss, W. (2023). Validating an AI-Driven Scoring System: The Model Card Approach. In Karim Sadeghi & Dan Douglas (eds), *Fundamental Considerations in Technology Mediated Language Assessment*,

- (pp.115-134). New York: Routledge. DOI: 10.4324/9781003292395-10
- Ozer, O. (2024). AI language models: A breach of academic integrity in online language learning? *Studies in Language Assessment (SILA)*, 13(1), 237-262.
<https://doi.org/10.58379/EZYS5680>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14, Article 49.
<https://doi.org/10.1186/s40468-024-00321-0>
- Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to Generate Reading Comprehension Items. *Educational Measurement: Issues and Practice*, 43(1), 5-18.
<https://doi.org/10.1111/emip.12590>
- Shin, D., & Lee, J. H. (2024). AI-powered automated item generation for language testing. *ELT Journal*, 78(4), 446-452. <https://doi.org/10.1093/elt/ccae016>
- Voss, E. (2024). Language assessment and artificial intelligence. In A. Kunnan (Ed.) *The Concise Companion to Language Assessment* (pp. 112–125). John Wiley & Sons.
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4-5), 520-532.
<https://doi.org/10.1080/15434303.2023.2288256>
- Wolf, M. K., Suhan, M., & Alderman, J. (2024). Exploring the provision of AI-based feedback for TOEFL Junior Writing practice tasks (Research Memorandum No. RM 24-05). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-05.pdf>
- Xi, X. (2023). Advancing Language Assessment with AI and ML-Leaning into AI is Inevitable, but Can Theory Keep Up? *Language Assessment Quarterly*, 20(4-5), 357-376. <https://doi.org/10.1080/15434303.2023.2291488>

Social Aspects of Language Assessment

Test Impact, Washback, and Consequences

- Ahmadi Safa, M., & Sheykhmolmolu, H. (2023). An impact study of the Iranian National University Entrance Exam from students and parents' perspectives. *Language Testing in Asia*, 13, Article 40. <https://doi.org/10.1186/s40468-023-00254-0>
- Ahmed Abdel-Al Ibrahim, K., Cuba Carbajal, N., Zuta, M. E. C., & Bayat, S. (2023). Collaborative learning, scaffolding-based instruction, and self-assessment: Impacts on intermediate EFL learners' reading comprehension, motivation, and anxiety. *Language Testing in Asia*, 13, Article 16. <https://doi.org/10.1186/s40468-023-00229-1>
- Al-Hawamdeh, B. O. S., Hussien, N., & Abdelrasheed, N. S. G. (2023). Portfolio vs. summative assessment: Impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13, Article 12. <https://doi.org/10.1186/s40468-023-00225-5>
- AL-TAMIMI, Y. A., QUEVEDO-CAMARGO, G., & MELO, R. S. R. (2022). Washback of a High-stakes Language Exam on Spanish Teachers' Practice in Brazil. *Journal of Linguistics and Language Teaching*, v. 13, p. 1-20.
- Alazemi, A. F. T., Jember, B., & Al-Rashidi, A. H. (2023). How to decrease test anxiety: A focus on academic emotion regulation, L2 grit, resilience, and self-assessment. *Language Testing in Asia*, 13, Article 28. <https://doi.org/10.1186/s40468-023-00241-5>
- Ali, M. M., & Hamid, M. O. (2023). Teacher agency and washback: Insights from classrooms in rural Bangladesh. *ELT Journal*, 77(1), 52-61. <https://doi.org/10.1093/elt/ccac005>
- Allahyari, R., Abbasabady, M. M., Akhter, S., & Alibakhshi, G. (2023). EFL teachers' cognition of social and psychological consequences of high-stake national language tests: Role of teacher training workshops. *Language Testing in Asia*, 13, Article 54. <https://doi.org/10.1186/s40468-023-00262-0>
- Allen, D., & Tahara, T. (2021). A review of washback research in Japan. *Japanese Language Testing Association Journal*, 24, 3-22. https://doi.org/10.20622/jltajournal.24.0_3
- Allen, D., Yamamura, K., Meguro, S., & Nakamura, T. (2024). Washback from Four-Skills English Exams in the Japanese secondary school context. *Language Assessment Quarterly*, 21(4-5), 473-505. <https://doi.org/10.1080/15434303.2024.2440895>
- Antia, B. (2021). Multilingual examinations: Towards a schema of politicization of language in end of high school examinations in sub-Saharan Africa. *International Journal of Bilingual Education and Bilingualism*, 24(1), 138-153. <https://doi.org/10.1080/13670050.2018.1450354>
- Arques, A. C., & Ferrero, C. L. (2023). Peer-feedback of an occluded genre in the Spanish language classroom: A case study. *Assessing Writing*, 57, Article 100756. <https://doi.org/10.1016/j.asw.2023.100756>
- Barnawi, O. Z., Alharbi, M. S., & Alzahrani, A. A. (Eds.). (2022). *Transnational English language assessment practices in the age of metrics*. Routledge. <https://doi.org/10.4324/9781003252382>
- Brunsting, N., Yu, Q., Smart, J., and Bingham, W. P. (2022). Investigating linkages between international students' English language proficiency, social-contextual outcomes, and

wellbeing in U.S. universities. IELTS Research Reports Online Series, No. 4/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/brunsting-yu-smart-bingham-2022.ashx>

- Cardwell, R. L., Nydick, S. W., Lockwood, J. R., & von Davier, A. A. (2024). Practical considerations when building concordances between English tests. *Language Testing*, 41(1), 192–202. <https://doi.org/10.1177/02655322231195027>
- Carlsen, C. H., & Rocca, L. (2021). Language test misuse. *Language Assessment Quarterly*, 18(5), 477–491. <https://doi.org/10.1080/15434303.2021.1947288>
- Carlsen, C. H., & Rocca, L. (2022). Language test activism. *Language Policy*, 21(4), 597–616. <https://doi.org/10.1007/s10993-022-09614-7>
- Cheng, L. (2024). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–1146). Chichester, UK: John Wiley & Sons. doi:10.1002/9781118411360.wbcla071. <https://onlinelibrary.wiley.com/doi/10.1002/9781118411360.wbcla071>
- Cheng, L. & Sultana, N. (2022). Washback: Looking backward and forward. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp 136–152). <https://doi.org/10.4324/9781003220756-12>
- Curtis, A. & Cheng, L. (2024). The Complexities of Diversity: A Personal Journey Across Time and Space. *The Language Teacher*, 48(5), 14-17 <https://doi.org/10.37546/JALTTLT48.5-3>
- Dao, P., Nguyen, M.X.N.C, & Nguyen, H.V. (2024). ‘IELTS juniors’ in Vietnam: Perceptions of learners, parents and IELTS preparation course providers. IELTS Research Reports Online Series, No. 3/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at <https://ielts.org/researchers/our-research/research-reports>
- De Silva, R. (2021). Assessment literacy and assessment practices of teachers of English in a South-Asian context: Issues and possible washback. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world* (pp. 433–446). Springer. https://doi.org/10.1007/978-981-33-4232-3_31
- Dee, T. S., & Domingue, B. W. (2021). Assessing the impact of a test question: Evidence from the “Underground Railroad” controversy. *Educational Measurement: Issues and Practice*, 40(2), 81–88. <https://doi.org/10.1111/emip.12411>
- Devine, A., Lopes, S., & Harrison, G. (2021). Young learners’ positive attitudes to learning in South America. *Research Notes*, 80, 21–28. Cambridge Assessment English.
- Ducasse, A.M. (2022) Oral reflection tasks: Advanced Spanish L2 learner insights on emergency remote teaching assessment practices in a higher education context. *Languages*, 7(1), Article 26. <https://doi.org/10.3390/languages7010026>
- Ekiert, M., & di Gennaro, K. (2021). Focused written corrective feedback and linguistic target mastery: Conceptual replication of Bitchener and Knoch (2010). *Language Teaching*, 54(1), 71–89. <http://doi.org/10.1017/S0261444819000120>
- Elder, C. (2021). Towards ‘policy responsible’ language assessment: Framing the Language Testing Research Centre’s contribution in languages other than English. *Papers in Language Testing and Assessment*, 10(1), 4–29. http://www.altanz.org/uploads/5/9/0/8/5908292/2._elder.pdf
- Erickson, G., Borger, L., & Olsson, E. (2022). National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language Testing*, 39(3),

- 474–493. <https://doi.org/10.1177/02655322221075067>
- François, J., & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. *Language Learning & Technology*, 25(2), 26–41. <http://hdl.handle.net/10125/73429>
- Frisén, L. B. (2024). Eyes on the ball: Teachers' rating processes when assessing a national L2 speaking test through the lens of their scoring document. *Studies in Language Assessment (SILA)*, 13(1), 1-36. <https://doi.org/10.58379/DAGH1483>
- Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, 11, Article 27. <https://doi.org/10.1186/s40468-021-00143-4>
- Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58–76. <https://doi.org/10.1080/0969594X.2020.1735995>
- Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486–1501. <https://doi.org/10.1080/13670050.2019.1691498>
- Hu, T & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1486-1501, DOI: 10.1080/13670050.2019.1691498
- Inbar-Lourie, O., & Shohamy, E. (2021). The Conflict and Consequences of Two Assessment Measures in Israel: Global PISA vs. the National MEITZAV. In B. Lantaigne (Ed.). *Challenges in Language Testing around the World* (pp. 191-200). Springer International Publishing
- Islam, M. S., Hasan, M. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, 11, Article 1. <https://doi.org/10.1186/s40468-020-00116-z>
- Ismail, S. M., Nikpoo, I., & Prasad, K. D. V. (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. *Language Testing in Asia*, 13, Article 27 (2023). <https://doi.org/10.1186/s40468-023-00239-z>
- Ismail, S.M., & Heydarnejad, T. (2023). Probing into the influence of EFL learners' self-assessment and evaluation apprehension in predicting their personal best goals and self-efficacy skills: A structural equation modeling. *Language Testing in Asia*, 13, Article 8 (2023). <https://doi.org/10.1186/s40468-023-00219-3>
- Iwashita, N., Sasaki, M., Stell, S., & Yucel, M. (2021). Japanese stakeholders' perceptions of IELTS writing and speaking tests and their impact on communication and achievement. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/teaching-and-research/research-reports>
- Kang, O., Ahn, H., Yaw, K., & Chung, S-Y. (2021). Investigation of relationships between learner background, linguistic progression, and score gain on IELTS. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia. https://www.ielts.org/-/media/research-reports/ielts-rr_2021-1_kang-et-al.ashx

- Kang, O., Hirschi, K., Miao, Y., Ahn, H., & Won, Y. (2022). Test-takers' IELTS preparations and their attitudes towards IELTS practices and the use of technologies in the global pandemic. IELTS Research Reports Online Series, No. 2/22. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/-/media/research-reports/kang-et-al-2022.ashx>
- Karatay, Y., & Karatay, L. (2024). The impact of automated writing evaluation on language learning and teaching: A research synthesis. *System*, 123.
<https://doi.org/10.1016/j.system.2024.103332>
- Katagiri, W. N. (2023). How sociocultural factors mediate washback effect: a study of the Hong Kong Diploma of Secondary Education Examination English Language Paper 4. *Language Testing in Asia*, 13, Article 38 (2023).
<https://doi.org/10.1186/s40468-023-00250-4>
- Khodi, A., Ponniah, L. S., Farrokhi, A. H., & Sadeghi, F. (2024). Test review of Iranian English language proficiency test: MSRT test. *Language Testing in Asia*, 14, Article 4. <https://doi.org/10.1186/s40468-023-00270-0>
- Kılıçkaya, F. (2022). Pre-service EFL teacher candidates' concerns over digital language assessment and test anxiety. In W. M. El-Henawy & M. del Mar Suárez (Eds.), *English as a foreign language in a new-found post-pandemic world* (pp. 164-181). IGI Global. <https://doi.org/10.4018/978-1-6684-4205-0>
- Lee, E. (2022). Examining the social consequences of a locally-developed placement test using test takers' attitudes. *Assessing Writing*, 53, Article 100626.
<https://doi.org/10.1016/j.asw.2022.100626>
- Li, J. (2021). Perceived effects of CET4 test preparation, language ability, and test performance: An exploratory study of Chinese EFL learners. *Language Education & Assessment*, 4(2), 38–58. <https://doi.org/10.29140/lea.v4n2.480>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
- López-Gopar, M. E. (2021). Citizenship in language testing: A call for respectful collaborations. *Language Testing*, 38(4), 649–653.
<https://doi.org/10.1177/02655322211022061>
- Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing*, 56, Article 100725.
<https://doi.org/10.1016/j.asw.2023.100725>
- Lydster, C. (2024). Who says “Ps get Degrees”? Examining the Profile of Undergraduate Students Maintaining High Achievement at University. *Student Success*, 15(2), 23–33. <https://doi.org/10.5204/ssj.3402>
- Macqueen, S, Pill, J., & Knoch, U. (2021). Trust the test: Score-user perspectives on the roles of language tests in professional registration and skilled migration. *Papers in Language Testing and Assessment*, 10(1), 49–69.
http://www.altanz.org/uploads/5/9/0/8/5908292/4._macqueen_pill_knoch.pdf
- Macqueen, S. (2024). [Review of the book *Reconsidering Context in Language Assessment: Transdisciplinary Perspectives, Social Theories, and Validity* by J. Fox and N. Artemeva]. *Language Testing*, 41(1), 220–226.
<https://doi.org/10.1177/02655322231199501>

- Mäkipää, T. (2021). Students' and teachers' perceptions of self-assessment and teacher feedback in foreign language teaching in general upper secondary education – A case study in Finland. *Cogent Education*, 8(1), 1978622. <https://doi.org/10.1080/2331186X.2021.1978622>
- Mäkipää, T. (2023). Students' perceptions of teacher assessment practices in foreign language emergency remote teaching during Covid-19 in Finland. *Studies in Language Assessment*, 12(1), 1–25. <https://doi.org/10.58379/AHRY2151>
- McLeod, M. (2023). English language proficiency requirements for migration to Canada, Australia, the United Kingdom, and the United States, and the implications for language testing research. *Canadian Journal of Applied Linguistics*, 26(1), 1–27. <https://doi.org/10.37213/cjal.2023.32825>
- Müller, A. & Han, W. (2022). *IELTS writing band scores 5.5–7.5: Grammatical error rates, stakeholder perceptions, and risk* (IELTS Research Reports Online Series, No. 1/22). International English Language Testing System. <https://www.ielts.org/-/media/research-reports/muller-and-han-2022.ashx>
- Müller, A., & Brenner, A. (2024). Benchmarking English standards across professions and professional university degrees IELTS Research Reports Online Series, No. 4/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at: <https://ielts.org/researchers/our-research/research-reports>
- Nguyen, B. T. T., Nguyen, T. T. B., Phan, Q. N. (2022). Vietnamese IELTS learners' perceptions of IELTS Writing Task 2 and their performance-based lexical resource. IELTS Research Reports Online Series, No. 5/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/nguyen-nguyen--phan-oct-2022.ashx>
- Nieto, D., & Nguyen, A. (2023). Strategizing assessment to support emerging bilingual students: a focus on the home language. *Multicultural Learning and Teaching*, 18(2), 239–250. <https://doi.org/10.1515/mlt-2021-0005>
- Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of IELTS preparation materials from a critical perspective. *Language Assessment Quarterly*, 18(4), 382–397. <https://doi.org/10.1080/15434303.2021.1883618>
- Nurjain, A., Salazar-Espinoza, D. E., Saenko, N., & Bina, E. (2023). Learner-oriented assessment matters: testing the effects of academic buoyancy, reflective thinking, and learner enjoyment in self-assessment and test-taking anxiety management of the EFL learners. *Language Testing in Asia*, 13, Article 30 (2023). <https://doi.org/10.1186/s40468-023-00247-z>
- O'Sullivan, B., Dunn, K., & Berry, V. (2021). Test preparation: An international comparison of test takers' preferences. *Assessment in Education: Principles, Policy & Practice*, 28(1), 13–36. <https://doi.org/10.1080/0969594X.2019.1637820>
- Ortiz, M., and Díaz, C. (2022). *The effect of screencast feedback on the performance of the IELTS essay writing component among pre-service teachers of English: An intervention study in Chile*. IELTS Research Reports Online Series, No. 3/22. British Council, Cambridge Assessment English and IDP: IELTS Australia. <https://www.ielts.org/-/media/research-reports/ortiz-and-diaz-2022-june.ashx>
- Ozer, O. (2024). AI language models: A breach of academic integrity in online language learning? *Studies in Language Assessment (SILA)*, 13(1), 237–262.

- <https://doi.org/10.58379/EZYS5680>
- Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly*, 39, 372–425. <https://doi.org/10.32038/ltrq.2024.39.23>
- Panahi, A., Mohebbi, H., & Oxford, R. L. (2024). Rebecca Oxford's 50 years of contributions to language education and related fields: A systematic review. *Language Teaching Research Quarterly*, 41, 169–230.
- Pearson, W. (2021). The sale of IELTS testing in India: Motivations and implications. *ELT Journal*, 75(4), 515–517. <https://doi.org/10.1093/elt/ccab053>
- Pearson, W. S. (2022). The mediating effects of student beliefs on engagement with written feedback in preparation for high-stakes English writing assessment. *Assessing Writing*, 52, Article 100611. <https://doi.org/10.1016/j.asw.2022.100611>
- Pill, J., Bandman, J., Bottini, R., Brunfaut, T., Davidson, N., Dávila Pérez, G., Harding, L., Jung, Y., Lestari, S. B., Ramos Gálvez, C., & Rossi, O. (2024). Shaping a language testing curriculum: Insights from an oral history of a Master's programme. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence. Volume 1: Research and reflections from the field* (pp. 168–191). Cambridge University Press. <https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- QUEVEDO-CAMARGO, G., & SCARAMUCCI, M. V. R. (2016) Methodology in washback studies. In: GITSAKI, Christine; COOMBE, Christine. (Eds.). *Current Trends in Language Evaluation, Assessment & Testing: Research Perspectives*. 1ed. Cambridge: Cambridge Scholars Publishing, v. 1, p. 287-304.
- Rahman, K. A., Ibna Seraj, P.M., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11, Article 12. <https://doi.org/10.1186/s40468-021-00129-2>
- Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education: Principles, Policy & Practice*, 29(3), 383-394. <https://doi.org/10.1080/0969594X.2022.2103516>
- Rios, J. A., & Ihlenfeldt, S. D. (2021). State assessment score reporting practices for English learner parents. *Educational Measurement: Issues and Practice*, 40(3), 31–41. <https://doi.org/10.1111/emip.12424>
- Ruegg, R., Hoang, H., & Petersen, N. (2024). The influence of English language proficiency test scores on the academic success of ESL undergraduate students. *Educational Research and Evaluation*, 29(3-4), 111-129. <https://doi.org/10.1080/13803611.2024.2314533>
- Russell, B., Barron, C., Kim, H., & Jang, E. E. (2022). A mixed-method investigation into international university students' experience with academic language demands. *Frontiers in Education*. <https://doi.org/10.3389/educ.2022.934692>
- Saif, S., Ma, J., May, L., & Cheng, L. (2021). Complexity of test preparation across three contexts: Case studies from Australia, Iran and China. *Assessment in Education: Principles, Policy & Practice*, 28(1), 37–57. <https://doi.org/10.1080/0969594X.2019.1700211>
- Sayyadi, A., & Rezvani, R. (2021). Questioning in TOEFL iBT speaking test: A case of washback and construct underrepresentation. *Language Testing in Asia*, 11, Article 21. <https://doi.org/10.1186/s40468-021-00137-2>

- Schissel, J. L., & Khan, K. (2021). Responsibilities and opportunities in language testing with respect to historicized forms of socio-political discrimination: A matter of academic citizenship. *Language Testing*, 38(4), 640–648. <https://doi.org/10.1177/02655322211028590>
- Shamsaldeen, F., Wang, J., & Ahn, S. (2024). Evaluating the fairness of a high-stakes college entrance exam in Kuwait. *Language Testing in Asia*, 14, Article 27. <https://doi.org/10.1186/s40468-024-00301-4>
- Sherkuziyeva, N., Imamutdinovna Gabidullina, F., Ahmed Abdel-Al Ibrahim, K., & Bayat, S. (2023). The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety. *Language Testing in Asia*, 13, Article 15. <https://doi.org/10.1186/s40468-023-00227-3>
- Shin, D. (2024). Critical discursive approaches to evaluating policy-driven testing: Social impact as a target for validation. *Language Testing*, 41(1), 162–180. <https://doi.org/10.1177/02655322231163863>
- Shohamy, E. (2022). Critical language testing, multilingualism and social justice. *TESOL Quarterly*, 56(4), 1445–1457. <https://doi.org/10.1002/tesq.3185>
- Suvorov, R., Karatay, Y., & Hegelheimer, V. (2024). Computer-Assisted Language Testing. In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment*. Wiley.
- Tashmuradova, B., Prasad, K. D. V., Ahmed Abdel-Al Ibrahim, K., & Sarabani, L. (2023). The comparative effects of consciousness raising tasks and dynamic assessment on morphological awareness: The case of pre-intermediate EFL learners. *Language Testing in Asia*, 13, Article 19. <https://doi.org/10.1186/s40468-023-00231-7>
- Taylor, L. (2023). Reframing the discourse and rhetoric of language testing and assessment for the public square. *Language Testing*, 40(1), 47–53. <https://doi.org/10.1177/02655322221127421>
- Tinkle, T., Godfrey, J., Menon, A. R., Moos, A., Romaine, L., & Sprouse, M. (2022). (In)Equities in directed self-placement. *Assessing Writing*, 54, Article 100671. <https://doi.org/10.1016/j.asw.2022.100671>
- Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, 5, 1–15. <https://doi.org/10.22599/jesla.74>
- Tsagari, D. & H. Bøhn (eds.) (2022) Recent Developments in Language Testing and Assessment (Special Issue) Issue 7. *Languages* (open access journal) https://www.mdpi.com/journal/languages/special_issues/Language_Testing_Assessment?fbclid=IwAR2_A6SIGEKHpZ95o5AQvPmM3kzi6O3rqypHjoTPA22LIdl6F_dHJLpnas
- Tsagari, D., & Giannikas, C. (2021). The impact of cert-mania of English language learning and teaching: the Cypriot case. *The European Journal of Applied Linguistics and TEFL*, 10(1), 193–215.
- Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238. <https://doi.org/10.1177/02655322211050600>
- Valkova, R. B. (2024). Remote Proctoring in language assessment: Exploring the impact on test-takers' scores and perceptions. *Studies in Language Assessment (SILA)*, 13(1), 141–198. <https://doi.org/10.58379/FCZM2855>

- Wang, J., Zheng, Y., & Zou, Y. (2024). Face validity and washback effects of the shortened PTE Academic: Insights from teachers in Mainland China. *Language Testing in Asia*, 14, Article 32. <https://doi.org/10.1186/s40468-024-00302-3>
- Wang, P. & Cheng, L. (2024). The Impact of TOEFL iBT Test Preparation on Chinese Test-Takers' Perceived Test Value. *Critical Inquiry in Language Studies*.
- Wang, S. (2021). *Consequential validity of the Chinese Proficiency Test (HSK) from macro and micro perspectives*. Cambridge Scholars Publishing.
- Wolf, M. K., Lopez, A. A., & Lee, J. (2023). An investigation of the use of standardized and local assessments for young EAL students. In G. Brooks, J. Clenton, & S. Fraser (Eds.), *EAL research for the classroom: Practical and pedagogical implications* (pp. 164-184). Routledge. <https://doi.org/10.4324/9781003274889-13>
- Xu, J., & Wang, Y. (2023). Individual differences in L2 writing feedback-seeking behaviors: The predictive roles of various motivational constructs. *Assessing Writing*, 56, Article 100698. <https://doi.org/10.1016/j.asw.2023.100698>
- Yan, X., Chuang, P.-L., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment*, 13(1), 108-140. <https://doi.org/10.58379/DLDI3148>
- Yan, X., Chuang, P., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment (SILA)*, 13(1), 108-140. <https://doi.org/10.58379/DLDI3148>
- Yan, Z., King, R. B., & Haw, J. Y. (2021). Formative assessment, growth mindset, and achievement: Examining their relations in the East and the West. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 676–702. <https://doi.org/10.1080/0969594X.2021.1988510>
- Yen, P. H., & Thao, L. T. (2024). Exploring the implementation and perception of competency-based assessment practices among Vietnamese EFL instructors. *Language Testing in Asia*, 14, Article 26. <https://doi.org/10.1186/s40468-024-00300-5>
- Yeo, M. (2023). Peer Review Circles: The Value of Working in Triadic Groups and Oral Observational Peer Feedback in the Development of Student Feedback Literacy. *Chinese Journal of Applied Linguistics*, 46(2), 270-288. <https://doi.org/10.1515/CJAL-2023-0208>
- Yi, Y.-S. (2021). On the usefulness of CDA-based score reporting: Implications for self-regulated learning. *Language Testing in Asia*, 11, Article 13. <https://doi.org/10.1186/s40468-021-00127-4>
- Yu, G., & Green, A. (2021). Preparing for admissions tests in English. *Assessment in Education: Principles, Policy & Practice*, 28(1), 1–12. <https://doi.org/10.1080/0969594X.2021.1880120>
- Yu, S., & Jiang, L. (2023). L2 university students' motivational self system in English writing: a sociocultural inquiry. *Applied Linguistics Review*, 14(3), 553-578. <https://doi.org/10.1515/applirev-2020-0110>
- Yumsek, M. (2023). Educational L2 constructs and diagnostic measurement. *Language Testing in Asia*, 13, Article 3 (2023). <https://doi.org/10.1186/s40468-022-00214-0>
- Zarrinabadi, N., Mahmoudi-Gahrouei, V., & Mohammadabadi, A. M. (2023). Your writing could have been better: Examining the effects of upward and downward counterfactual communication on the motivational aspects of L2 writing. *Assessing*

- Writing*, 56, Article 100714. <https://doi.org/10.1016/j.asw.2023.100714>
- Zhan, Y., Lee, J. C. K., & Wan, Z. H. (2022). Who perceives more value of teacher feedback? Exploring the roles of college students' possible second language selves and language learning strategies. *Language Testing in Asia*, 12, Article 62. <https://doi.org/10.1186/s40468-022-00212-2>
- Zhang, H. (2024). The washback of the National Matriculation English Test on senior high school English learning outcomes: Do test takers from different provinces think alike? *Language Testing in Asia*, 14, Article 14. <https://doi.org/10.1186/s40468-024-00286-0>
- Zhang, J. (2021). A moderated mediation analysis of the relationship between a high-stakes English test and test takers' extracurricular English learning activities. *Language Testing in Asia*, 11, Article 5. <https://doi.org/10.1186/s40468-021-00120-x>
- Zhang, L., Zheng, Y., O'Sullivan, B., & Stanley, G. (2023). Teacher perceptions of learner engagement with a new online Llearning environment. In H. Işıl Işık, E. Kantarcioğlu, A. Üner & C. Thomas (Eds.), *Re-vision and re-form in English for academic purposes contexts after the Pandemic* (pp. 124 –132). Cambridge Scholars.
- Zhang, Z. V., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, 58, Article 100784. <https://doi.org/10.1016/j.asw.2023.100784>

Standards and Frameworks

- Ali, M. M. E. (2023). The foreign-accentedness, comprehensibility, and intelligibility of L2 Arabic speech. *Language Teaching Research*, 0(0). <https://doi-org.ezproxy2.library.arizona.edu/10.1177/13621688231158787>
- Allen, D., & Koizumi, R. (2024). The English Speaking Achievement Test for Japanese Junior High School Students (ESAT-J): An analysis of public criticisms using the socio-cognitive framework. *Language Assessment Quarterly*, 21(4–5), 404–426. <https://doi.org/10.1080/15434303.2024.2438982>
- Arce, A. J., & Young, M. J. (2022). Test efficacy: Refocusing validation from college exams to candidates. *International Journal of Testing*, 22(1), 100–119. <https://doi.org/10.1080/15305058.2021.2019752>
- Chapelle, C. A., Voss, E., & Kim, H. (2024). Designing evaluations for validation of language assessments. In A. J. Kunnan (Ed.). *The Concise Companion to Language Assessment*. (pp. 67–79). Wiley-Blackwell.
- Colman, A., Segers, W., & Verplaetse, H. (2023). Extending Preselected Items Evaluation (PIE) with the ATA (American Translators Association) error categories for objective and targeted translation evaluation. *Journal of Applied Linguistics and Professional Practice*, 17(3), 274-293. <https://doi.org/10.1558/jalpp.21681>
- De Jong, J. H.A.L. (2022) The Action-oriented approach and language testing: A critical view. In D. Little & N. Figueras (Eds.), *Reflecting on the Common European Framework of reference for languages and its companion volume* (pp. 53–63). *Multilingual Matters*. <https://doi.org/10.21832/9781800410206-009>
- Deygers, B. (2021). The CEFR companion volume: Between research-based policy and policy-based research. *Applied Linguistics*, 42(1), 186–191. <https://doi.org/10.1093/applin/amz024>

- Fan, J., Knoch, U., & Chen, I. (2021). *Linking the GEPT writing subtest (Part 1) to the Common European Framework of Reference (CEFR)*. Language Testing Research Centre.
https://www.lttc.ntu.edu.tw/lttc-grants/doc/108report/GEPT_linking_study_Final_report.pdf
- Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices. *Language Testing in Asia*, 11, Article 2. <https://doi.org/10.1186/s40468-020-00117-y>
- Fox, J., Abdulhamid, N. & Turner, C.E. (2022). Classroom based assessment. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing*, 2nd ed. (pp. 119-135). Routledge, Taylor & Francis Group.
<https://doi.org/10.4324/9781003220756-6>
- Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56. <https://doi.org/10.32038/ltrq.2022.29.03>
- Green, A. and Chan, S. (2022). Report to the Nursing and Midwifery Council on language test policy. *Nursing and Midwifery Council*, 1–27.
<https://www.nmc.org.uk/globalassets/sitedocuments/english-language-consultation/crella-report-to-the-nmc-june-22.pdf>
- Harsch, C., Koval, V., Delgado-Osorio, X. & Hartig, J. (2024). Usability of CEFR Companion Volume scales for the development of an analytic rating scale for academic integrated writing assessment. In *CEFR Journal - Research and Practice*. Vol. 6, 155-177, <https://doi.org/10.37546/JALTSIG.CEFR6-9>
- He, L., Jiang, Z., & Min, S. (2021). Diagnosing writing ability using China's Standards of English Language Ability: Application of cognitive diagnosis models. *Assessing Writing*, 50, Article 100565. <https://doi.org/10.1016/j.asw.2021.100565>
- Heydarnejad, T., Ibrahim, K. A. A., Abdelrasheed, N. S. G., & Rezvani, E. (2022). The effect of academic emotion regulation on EFL learners' core of self-assessment and academic buoyancy: a structural equation modeling. *Language Testing in Asia*, 12, Article 57. <https://doi.org/10.1186/s40468-022-00207-z>
- Hoeve, K. B. (2022). A validity framework for accountability: educational measurement and language testing. *Language Testing in Asia*, 12, Article 3.
<https://doi.org/10.1186/s40468-021-00153-2>
- Huhta, A., Harsch, C., Leontjev, D., & Nieminen, L. (2024). *The diagnosis of writing in a second or foreign language*. Routledge.
- Imperial, J. M., Forey, G., & Tayyar Madabushi, H. (2024). Standardize: Aligning Language Models with Expert-Defined Standards for Content Generation. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1573–1594). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.94>
- Isaacs, T., & Rose, H. (2022). Redressing the balance in the native speaker debate: Assessment standards, standard language, and exposing double standards. *TESOL Quarterly*, 56(1), 401-412. <https://doi.org/10.1002/tesq.3041>
- Jin, Y. (2022). Consequential research of accountability testing: the case of the CET. *Language Testing in Asia*, 12, Article 15.
<https://doi.org/10.1186/s40468-022-00165-6>
- Khosravani, M., Rostamian, M., Ashraf, H., & Motallebzade, K. (2022). A structural equation

- modeling of English tests' social and educational consequences: exploring target, leverage, risk, and critical variables. *Language Testing in Asia*, 12, Article 46.
<https://doi.org/10.1186/s40468-022-00177-2>
- Kollias, C. (2023). *Virtual standard setting: Setting cut scores*. Peter Lang.
<https://doi.org/10.3726/b20407>
- Latif, M. M. A. (2021). Remodeling writers' composing processes: Implications for writing assessment. *Assessing Writing*, 50, Article 100547.
<https://doi.org/10.1016/j.asw.2021.100547>
- Lee, I. (2021). The development of feedback literacy for writing teachers. *TESOL Quarterly*, 55(3), 1048–1059. <https://doi.org/10.1002/tesq.3012>
- Leung, C., Evans, M., & Liu, Y. (2021). English as an additional language assessment framework: Filling a void in policy and provision in school education in England. *Language Assessment Quarterly*, 18(3), 296–315.
<https://doi.org/10.1080/15434303.2020.1869745>
- Little, D., & Figueras, N. (2022). *Reflecting on the common European framework of reference for languages and its companion volume*. Multilingual Matters.
- Liu, C., & Yu, S. (2022). Reconceptualizing the impact of feedback in second language writing: A multidimensional perspective. *Assessing Writing*, 53, Article 100630.
<https://doi.org/10.1016/j.asw.2022.100630>
- McNamara, D. S., & Kendeou, P. (2022). The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*, 29(2), 150–182.
<https://doi.org/10.1080/0969594X.2022.2037509>
- Moradkhan, D., & Shahmirzadi, N. (2012). Utilizing measures of (CAF) complexity, accuracy and fluency for interpreter evaluation. *Translation Studies (Motale'at-e-Tarjome Quarterly)*, 10(39), 87–104, (ISSN: 1735–0212).
- North, B., & Piccardo, E. (2023). Aligning Language Frameworks: An Example with the CLB and CEFR. *Language Assessment Quarterly*, 20(2), 143-165.
<https://doi.org/10.1080/15434303.2023.21842666>
- O'Reilly, D., & Marsden, E. (2023). Elicited metaphoric competence in a second language: a construct associated with vocabulary knowledge and general proficiency?. *International Review of Applied Linguistics in Language Teaching*, 61(2), 287-327.
<https://doi.org/10.1515/iral-2020-0054>
- Papageorgiou, S., Davis, L., Ohta, R., & Garcia Gomez, P. (2022), Mapping TOEFL® Essentials™ test scores to the Canadian Language Benchmarks. *ETS Research Report Series*, 2022(1), 1–42. <https://doi.org/10.1002/ets2.12357>
- Park, H.I., Solon, M., Dehghan-Chaleshtori, M., & Ghanbar, H. (2022). Proficiency reporting practices in research on second language acquisition: Have we made any progress?. *Language Learning*, 72, 198-236. <https://doi.org/10.1111/lang.12475>
- Poehner, M. E., & Lantolf, J. P. (2023). Advancing L2 dynamic assessment: Innovations in Chinese contexts. *Language Assessment Quarterly*, 20(1), 1-19.
<https://doi.org/10.1080/15434303.2022.2158465>
- Rahmati, T., Kazemian, M., Khonamri, F., & Daryabak, A. (2023). A chaos complexity perspective on EFL learners' sensitivity dependence on initial conditions in vocabulary tests. *The Journal of Asia TEFL*, 20(4), 895-903.
<http://dx.doi.org/10.18823/asiatefl.2023.20.4.9.895>
- Razavipour, K., & Raji, B. (2022). Reliability of measuring constructs in applied linguistics research: a comparative study of domestic and international graduate theses.

- Language Testing in Asia*, 12, Article 16. <https://doi.org/10.1186/s40468-022-00166-5>
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50–61. <https://doi.org/10.1016/j.asw.2017.09.003>
- Saeed, M. A., AbuSa'aleek, A., & Alharbi, M. A. (2024). Examining teacher's evaluative language in written, audio and screencast feedback on EFL learners' writing from the appraisal framework: A linguistic perspective. *Assessing Writing*, 61, 100871. <https://doi.org/10.1016/j.asw.2024.100871>
- Şahin, A. (2021). The Rater Performance Categorization System (RPCS) for intensive English programs. *Shanlax International Journal of Education*, 9(3), 225–241. <https://doi.org/10.34293/education.v9i3.3986>
- Saville, N. (2021). Learning-oriented assessment: Basic concepts and frameworks in using assessment to support language learning. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 13–33). Routledge. <https://doi.org/10.4324/9781003014102-3>
- Savski, K. (2023). CEFR and the ELT practitioner: Empowerment or enforcement?. *ELT Journal*, 77(1), 62–71. <https://doi.org/10.1093/elt/ccac013>
- Steiss, J., Krishnan, J., Kim, Y.-S. G., & Olson, C. B. (2022). Dimensions of text-based analytical writing of secondary students. *Assessing Writing*, 51, Article 100600. <https://doi.org/10.1016/j.asw.2021.100600>
- Sun, Z., Xu, P., & Wang, J. (2023). Dynamic assessment of the learning potential of Chinese as a second language. *Language Assessment Quarterly*, 20(1), 127–142. <https://doi.org/10.1080/15434303.2022.2151911>
- Taylor, L., & Banerjee, J. (2023). Language assessment accommodations: Issues and challenges for the future. *Language Testing*, 40(4), 1000–1006. <https://doi.org/10.1177/02655322231186222>
- Treadaway, M., & Read, J. (2024). Setting standards for a diagnostic test of aviation English for student pilots. *Language Testing*, 41(3), 557–583. <https://doi.org/10.1177/02655322231224051>
- Weideman, A. 2023. The practicality of theory: Reciprocity, assessment and applied linguistics. *Stellenbosch Papers in Linguistics Plus*, 66, 77–196. <https://doi.org/10.5842/66-1-935>
- Wolf, M. K., Bailey, A. L., & Ballard, L. (2023). Aligning English language proficiency assessments to standards: Conceptual and technical issues. *TESOL Quarterly*, 57(2), 670–685. <https://doi.org/10.1002/tesq.3199>
- Xie, Q., & Lei, Y. (2022). Diagnostic assessment of L2 academic writing product, process and self-regulatory strategy use with a comparative dimension. *Language Assessment Quarterly*, 19(3), 231–263. <https://doi.org/10.1080/15434303.2021.1903470>
- Yang, Y. (2023). The Correlation Between English and French Proficiency of Chinese Trilingual Learners. *Chinese Journal of Applied Linguistics*, 46(3), 404–423. <https://doi.org/10.1515/CJAL-2023-0305>

Language Assessment Literacy

- Abrar-ul-Hassan, S., & Nassaji, H. (2024). Rescoping language assessment literacy: An expanded perspective. *System*, 120, Article 103195.

- <https://doi.org/10.1016/j.system.2023.103195>
- Acar, A. S. (2023). Genre pedagogy: A writing pedagogy to help L2 writing instructors enact their classroom writing assessment literacy and feedback literacy. *Assessing Writing*, 56, Article 100717. <https://doi.org/10.1016/j.asw.2023.100717>
- Alicia Kim, A., Yumsek, M., Kemp, J. A., Chapman, M., & Gary Cook, H. (2023). Universal tools activation in English language proficiency assessments: A comparison of Grades 1–12 English learners with and without disabilities. *Language Testing*, 40(4), 877–903. <https://doi.org/10.1177/02655322221149009>
- Allahyari, R., Abbasabady, M. M., Akhter, S., & Alibakhshi, G. (2023). EFL teachers' cognition of social and psychological consequences of high-stake national language tests: Role of teacher training workshops. *Language Testing in Asia*, 13, Article 54. <https://doi.org/10.1186/s40468-023-00262-0>
- Baker, B. (2024). Introduction. In B. Baker and L. Taylor (Eds.), *Language assessment literacy and Competence volume 2: Case studies from around the world*. Cambridge.
- Baker, B. & Taylor, L. (2024). *Language assessment literacy and competence volume 2: Case studies from around the world*. Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/717471-silt-volume-56.pdf>
- Baker, B. & Taylor, L. (2024). *Language assessment literacy and competence volume 1: Research and reflections from the field*. Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- Balbino, T. R. T. & Moraes, I. T. (2023). Letramento em avaliação em contexto de línguas para pais e responsáveis por crianças [Language assessment literacy for children's parents and guardians]. *Revista X*, 18(1), 5–38. <https://doi.org/10.5380/rvx.v18i1.88026>
- Banitz, B. (2022). Language assessment in Mexico: Exploring university language teachers' backgrounds, practices, and opinions. *System*, 110, Article 102898. <https://doi.org/10.1016/j.system.2022.102898>
- Barkaoui, K., Shi, L., & Riaz, M. (2021). "Follow your interests because those will motivate you to excel": An interview with Alister Cumming. *Language Assessment Quarterly*, 18(1), 56–75. <https://doi.org/10.1080/15434303.2020.1846190>
- Batista, E. G. & Moraes, I. T. (2020). Avaliação orientada para a aprendizagem no ensino de línguas para crianças [Learning-oriented language assessment in the context of young learners]. *REVELLI*, 12, 1–19. <https://doi.org/10.51913/revelli.v12i0.10234>
- Bazvand, A. D. (2023). Exploring student teachers' perceptions of assessment ethics across university-based teacher education programs in Iran. *Language Testing in Asia*, 13, Article 9. <https://doi.org/10.1186/s40468-022-00205-1>
- Belda-Medina, J., & Kokošková, V. (2024). Assessing Pre-Service Teachers' Knowledge and Attitudes Towards Language Testing for Citizenship Through LAMAS (Language Accreditation for Migration Attitudes Scale). *Language Assessment Quarterly*, 21(4–5), 427–451. <https://doi.org/10.1080/15434303.2024.2438995>
- Berger, A. (2023). A Difficulty-Informed Approach to developing Language Assessment Literacy for classroom purposes. *Chinese Journal of Applied Linguistics*, 46(2), 289–307. <https://doi.org/10.1515/CJAL-2023-0209>
- Boggs, J. A., & Manchón, R. M. (2023). Feedback literacy in writing research and teaching: Advancing L2 WCF research agendas. *Assessing Writing*, 58, Article 100786. <https://doi.org/10.1016/j.asw.2023.100786>

- Bøhn, H. & Tsagari, D. (2022) Language assessment literacy: Understanding the construct from Norwegian EFL teachers' perspective. *Studies in Language Assessment*, 11(1), 119–148.
- Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>
- Britton, E. R. (2023). Developing teacher feedback literacy through self-study: Exploring written commentary in a critical language writing curriculum. *Assessing Writing*, 56, Article 100709. <https://doi.org/10.1016/j.asw.2023.100709>
- Britton, E. R., & Kraemer, A. Assessing Languages Across the Curriculum: Critical reflection and student learning perspectives. *Foreign Language Annals*, 57(4), 1051-1071. <https://doi.org/10.1111/flan.12775>
- Brookhart, S. M. (2023). Assessment Literacy in a better assessment future. *Chinese Journal of Applied Linguistics*, 46(2), 162-179. <https://doi.org/10.1515/CJAL-2023-0202>
- Brown, J. D., Panahi, A., & Mohebbi, H. (2023). James Dean Brown's 50 years of work in language testing: A systematic review. *Language Teaching Research Quarterly*, 37, 4–75. <https://doi.org/10.32038/ltrq.2023.37.02>
- Bui, H. P. (2023). Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment. *Language Testing in Asia*, 13, Article 10 (2023). <https://doi.org/10.1186/s40468-023-00220-w>
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455. <https://doi.org/10.1177/0265532221992274>
- Chappuis, S., Brookhart, S., & Chappuis, J. (2021). *Ten assessment literacy goals for school leaders*. SAGE.
- Chong, S. W., & Isaacs, T. (2023). An ecological perspective on classroom-based assessment. *TESOL Quarterly*, 57(4), 1558-1570. <http://doi.org/10.1002/tesq.3201>
- Crusan, D., Masjedlou, A. P., & Mohebbi, H. (2023). Writing teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.) *Insights into teaching learning writing: A practical guide for teachers* (pp. 45-55). Castledown. <http://doi.org/10.29140/9781914291159>.
- Crusan, D., Panahi, A., & Mohebbi, H. (2023). Writing teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 45-55). Castledown. <https://castledown.online/reference/9781914291159/>
- Cui, Y., Jin, H., & Gao, Y. (2023). Developing EFL teachers' feedback literacy for research and publication purposes through intra- and inter-disciplinary collaborations: A multiple-case study. *Assessing Writing*, 57, Article 100751. <https://doi.org/10.1016/j.asw.2023.100751>
- Cui, Y., Liu, Y., Yu, H., & Gao, Y. (2022). Developing English teachers' language assessment literacy in an EAP reform context through test design: A case study. *System*, 109, 102866. <https://doi.org/10.1016/j.system.2022.102866>
- Cushing, S. T., Ren, H., & Tan, Y. (2024). The use of TOEFL iBT® in admissions decisions: Stakeholder perceptions of policies and practices. ETS Research Report Series, 2024(1), 1–16. <https://doi.org/10.1002/ets2.12375>

- DAMACENA, M., & QUEVEDO-CAMARGO, G. (2021) Avaliação e formação de professores de línguas [Assessment and language teacher education]. *Revista Olhares e Trilhas*, v. 23, p. 1054-1073. DOI: <https://doi.org/10.14393/OT2021v23.n.3.62193>
- Davison, C. (2023). Assessment Literacy: Changing Cultures, Enculturing Change in Hong Kong. *Chinese Journal of Applied Linguistics*, 46(2), 180-197. <https://doi.org/10.1515/CJAL-2023-0203>
- Davison, C. (2023). Assessment Literacy: Changing Cultures, Enculturing Change in Hong Kong. *Chinese Journal of Applied Linguistics*, 46(2), 180-197. <https://doi.org/10.1515/CJAL-2023-0203>
- De Silva, R. (2021). Assessment literacy and assessment practices of teachers of English in a South-Asian context: Issues and possible washback. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world* (pp. 433–446). Springer. https://doi.org/10.1007/978-981-33-4232-3_31
- Dunn, K. J. (2024). Random-item Rasch models and explanatory extensions: A worked example using L2 vocabulary test item responses. *Research Methods in Applied Linguistics*, 3(3), Article 100143. <https://doi.org/10.1016/j.rmal.2024.100143>
- Erlam, R., & Rau, W. T. (2023). Fairer for all: Teachers identifying and solving problems with the assessment of English language learners in the mainstream classroom. *Studies in Language Assessment*, 12(2), 94-116.
- Estaji, M. (2024). Unveiling the impact of language assessment teacher education on EFL teachers' assessment agency and professional development trajectory: Perceptions, opportunities, and challenges. *Language Testing in Asia*, 14, Article 20. <https://doi.org/10.1186/s40468-024-00292-2>
- Farhady, H., & Tavassoli, K. (2021). EFL teachers' perceptions and practices of their language assessment knowledge. *Language Testing in Asia*, 11, Article 17, 1-19. <https://doi.org/10.1186/s40468-021-00135-4>.
- Farhady, H., & Tavassoli, K. (2021). EFL teachers' perceptions and practices of their language assessment knowledge. *Language Testing in Asia*, 11, Article 17. <https://doi.org/10.1186/s40468-021-00135-4>
- Fong, C. J., Schallert, D. L., Williamson, Z. H., Lin, S., Williams, K. M., & Kim, Y. W. (2023). Are self-compassionate writers more feedback literate? Exploring undergraduates' perceptions of feedback constructiveness. *Assessing Writing*, 57, Article 100761. <https://doi.org/10.1016/j.asw.2023.100761>
- Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56. <https://doi.org/10.32038/ltrq.2022.29.03>
- Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503-525. <https://doi.org/10.1080/15434303.2022.2128802>
- Gan, L., & Lam, R. (2023). A tale of two teachers: Construction of novice university English teachers' assessment identity. *Chinese Journal of Applied Linguistics*, 46(2), 219-236. <https://doi.org/10.1515/CJAL-2023-0205>
- Gebril, A. (2021). Nation-wide assessment reform in Egypt: Challenges and potentials of learning-oriented assessment in an examinations-based instructional setting. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 109–122). Routledge. <https://doi.org/10.4324/9781003014102-9>

- Glasson, N., & Halley, K. (2024). Less talk, more action? Exploring proficiency scores and embodied resources in online L2 interactions. *Classroom Discourse*, 1–22. <https://doi.org/10.1080/19463014.2024.2408421>
- Goh, R., & Kiat Tan, K. (2023). Teachers' qualitatively different ways of experiencing assessment feedback: Implications for teacher assessment literacy. *Chinese Journal of Applied Linguistics*, 46(2), 253–269. <https://doi.org/10.1515/CJAL-2023-0207>
- Gottlieb, M. (2021). *Assessment in multiple languages: A handbook for school and district leaders*. SAGE.
- Gottlieb, M. (2021). *Classroom assessment in multiple languages: A handbook for teachers*. Corwin.
- Greenier, V., Liu, X. & Xiao, Y. (2024). Creative translanguaging in formative assessment: Chinese teachers' perceptions and practices in the primary EFL classroom. *Applied Linguistics Review*, 15(5), 1861–1885. <https://doi.org/10.1515/applirev-2023-0085>
- Haddadian, G., Radmanesh, S., & Haddadian, N. (2024). Construction and validation of a Computerized Formative Assessment Literacy (CFAL) questionnaire for language teachers: An exploratory sequential mixed-methods investigation. *Language Testing in Asia*, 14, Article 33. <https://doi.org/10.1186/s40468-024-00303-2>
- Haghi, I., Clark, T. (2024). Assessment in the EAP Classroom. In E. Zehir Topkaya & H. Çelik (Eds.), *Teaching English for Academic Purposes*. Palgrave Macmillan/Cham. https://doi.org/10.1007/978-3-031-72545-6_13
- Harsch, C., Seyferth, S., & Villa Larenas, S. (2021). Evaluating a collaborative and responsive project to develop language assessment literacy. *Language Learning in Higher Education*, 11(2), 311–342. <https://doi.org/10.1515/cercles-2021-2020>
- Hill, K. & Ducasse, A. M. (2022). Contextual variables in written assessment feedback in a university-level Spanish program. *Studies in Language Assessment*, 11(1), 16–32. https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/4282558/SiLA-11.1-Hill_and_Ducasse.pdf
- Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101, Article 102569. <https://doi.org/10.1016/j.system.2021.102569>
- Inbar-Lourie, O. (2022). EMI programs and formative assessment: Implications for the assessment literacy of content lecturers. *Journal of English Medium Instruction* 1(2), 204–231. <https://doi.org/10.1075/jemi.21014.inb>
- Jalilzadeh, K., & Coombe, C. (2023). Constraints in employing learning-oriented assessment in EFL classrooms: teachers' perceptions. *Language Testing in Asia*, 13, Article 7 (2023). <https://doi.org/10.1186/s40468-023-00222-8>
- Janssen, G. (2022). Local placement test retrofit and building language assessment literacy with teacher stakeholders: A case study from Colombia. *Language Testing*, 39(3), 377–400. <https://doi.org/10.1177/02655322221076153>
- Kang, M., & Lam, R. (2024). Understanding university English instructors' assessment literacy: A formative assessment perspective. *Language Testing in Asia*, 14, Article 52. <https://doi.org/10.1186/s40468-024-00323-y>
- Kazemian, M. & Khonamri, F. (2022). Past, present, and future of language assessment: An interview with Dr. Hossein Farhady, *MEXTESOL Journal*, 46(4), 1–8. https://www.mextesol.net/journal/index.php?page=journal&id_article=46395
- Kazemian, M. & Khonamri, F. (2024). The online assessment in the Middle East. A

- conversation with Dr. Christine Coombe, *MEXTESOL Journal*, 48(1), 1-6.
- Kazemian, M., Khodareza, M.R., Khonamri, F., & Rahimy, R. (2023). ELT scholars' attitudes towards inclusion of intercultural competence assessment in language proficiency tests. *TESL-EJ Journal*, 26(4), 1–21.
<https://tesl-ej.org/wordpress/issues/volume26/ej104/ej104a6/>
- Kılıçkaya, F. (2021). Reconsidering language assessment courses in teacher training programs during and after the COVID-19 pandemic. *Lublin Studies in Modern Languages and Literature*, 45(3), 45–55.
<https://doi.org/10.17951/lsmll.2021.45.3.45-55>
- Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142-161.
<https://doi.org/10.1080/15434303.2021.2023542>
- Koizumi, R., Fukazawa, M., & Inoue, C. (2024). Usability of a Speaking Assessment Portal for Japanese teachers of English. *SHIKEN: A Journal of Language Testing and Evaluation in Japan*, 28(1), 19-37.
<https://doi.org/10.37546/JALTSIG.TEVAL28.1-2>
- Lam, D. M. K., Green, A., Murray, N., & Gayton, A. (2021). How are IELTS scores set and used for university admissions selection: A cross-institutional case study. *IELTS Research Reports Online Series*, 3. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/for-researchers/research-reports/online-series-2021-3>
- Latif, M. W. (2021). Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: Implications for teacher professional development. *Language Testing in Asia*, 11, Article 19. <https://doi.org/10.1186/s40468-021-00130-9>
- Latif, M. W., & Wasim, A. (2022). Teacher beliefs, personal theories and conceptions of assessment literacy—a tertiary EFL perspective. *Language Testing in Asia*, 12(1), Article 11. <https://doi.org/10.1186/s40468-022-00158-5>
- Latif, M. W., & Wasim, A. (2024). Investigating EFL instructors' approaches to classroom-based assessment culture: An explanatory sequential mixed-method approach. *Language Testing in Asia*, 14, Article 57.
<https://doi.org/10.1186/s40468-024-00332-x>
- Lee, I. (2021). The development of feedback literacy for writing teachers. *TESOL Quarterly*, 55(3), 1048–1059. <https://doi.org/10.1002/tesq.3012>
- Lee, I., Karaca, M., & Inan, S. (2023). The development and validation of a scale on L2 writing teacher feedback literacy. *Assessing Writing*, 57, Article 100743.
<https://doi.org/10.1016/j.asw.2023.100743>
- Lee, J. (2024). Analysis of ESL preservice teachers' collaboration with ESL learners through a language assessment development project. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence: Case studies from around the world* (Studies in Language Testing, Vol. 56, pp. 43–55). Cambridge University Press.
- Lee, J., Goldenberg, N., & Williams, D. P. (2023). Three stakeholders' reflections on language assessment literacy. *ELT Journal*, 77(4), 383–392.
<https://doi.org/10.1093/elt/ccad026>
- Li, J., & Gu, P. Y. (2023). Developing classroom-based formative assessment literacy: An EFL teacher's journey. *Chinese Journal of Applied Linguistics*, 46(2), 198–218.
<https://doi.org/10.1515/CJAL-2023-0204>

- Li, J., & Yongqi Gu, P. (2024). Formative assessment for self-regulated learning: Evidence from a teacher continuing professional development programme. *System*, 125, 103414. <https://doi.org/10.1016/j.system.2024.103414>
- Mäkipää, T. (2021). Students' and teachers' perceptions of self-assessment and teacher feedback in foreign language teaching in general upper secondary education – A case study in Finland. *Cogent Education*, 8(1), 1978622. <https://doi.org/10.1080/2331186X.2021.1978622>
- Mäkipää, T., Hahl, K., & Luodonpää-Manni, M. (2021). Teachers' perceptions of assessment and feedback practices in Finland's foreign language classes during the Covid-19 pandemic. *CEPS Journal*, 11, 219–240. <https://doi.org/10.26529/cepsj.1108>
- Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103, Article 102674. <https://doi.org/10.1016/j.system.2021.102674>
- Mohammadi, M., & Sanavi, R. V. (2021). Language assessment literacy: Ontogenetic and phylogenetic perspectives. In S. Hidri (Eds.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 52-65). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003016083-4/language-assessment-literacy-mojtaba-mohammadi-reza-vahdani-sanavi>
- Mohebbi, H. (2021). 25 years on, the correction debate continues: An interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), Article 3. <https://doi.org/10.1186/s40862-021-00110-9>
- Moraes, I. T. (2024). Concepções e práticas de avaliação em diálogo em cursos de formação de professores/as de inglês em instituições de ensino superior do Paraná [Assessment Concepts and Practices in Dialogue in English Teacher Education Programs in Higher Education Institutions in the State of Paraná]. *Revista X*, 19(1), 327-352. <https://doi.org/10.5380/rvx.v19i2.94796>
- Moraes, I. T. & Batista, E. G. (2020). Letramento em avaliação para professores de línguas estrangeiras para crianças: orientações teórico-práticas [Assessment literacy for foreign language teachers of children: theoretical and practical guidelines]. *Revista Horizontes de Linguística Aplicada*, 19(2), 15-42. <https://doi.org/10.26512/rhla.v19i2.26804>
- Morrison, A., & Sepulveda-Escobar, P. (2024). Chilean English as a Foreign Language Teacher Educators' Conceptions and Practices of Online Assessment. *RELJ Journal*, 55(2), 394–407. <https://doi.org/10.1177/00336882221079059>
- Nayernia, A., & Mohebbi, H. (2023). Teachers' online language assessment literacy. In Z. Tajeddin & C. Griffiths (Eds.), *Language Education Programs: Perspectives on Policies and Practices (Volume 34)* (pp. 121-140). Springer. https://doi.org/10.1007/978-3-031-38754-8_8
- Nayernia, A., Nosrati, R., & Mohebbi, H. (2022). The factors contributing to language teachers' effectiveness in an EFL learning context: A questionnaire validation study. *Profile: Issues in Teachers' Professional Development*, 24(2), 63–79. <https://doi.org/10.15446/profile.v24n2.93571>
- Noori, M. (2022). Road not taken" in language testing: Sociocultural implications of test and teaching contents. *TESOL Quarterly*, 56(4), 1486-1503. <https://doi.org/10.1002/tesq.3189>

- Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of IELTS preparation materials from a critical perspective. *Language Assessment Quarterly*, 18(4), 382–397. <https://doi.org/10.1080/15434303.2021.1883618>
- Pill, J., & Harding, L. (2019). A most engaging scholar: Tim McNamara and the role of language testing expertise. In C. Roever & G. Wigglesworth (Eds.), *Social perspectives on language testing: Papers in honour of Tim McNamara* (pp. 217–228). Peter Lang. <https://doi.org/10.3726/b15199>
- Pill, J., Bandman, J., Bottini, R., Brunfaut, T., Davidson, N., Dávila Pérez, G., Harding, L., Jung, Y., Lestari, S. B., Ramos Gálvez, C., & Rossi, O. (2024). Shaping a language testing curriculum: Insights from an oral history of a Master's programme. In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence. Volume 1: Research and reflections from the field* (pp. 168–191). Cambridge University Press. <https://www.cambridgeenglish.org/Images/717470-silt-volume-55.pdf>
- Purpura, J. E. (2024). An Interview with Timothy Francis McNamara. *Language Assessment Quarterly*, 21(2). 123-124.
- Purpura, J. E. (2024). In Memoriam: Timothy Francis McNamara. *Language Assessment Quarterly*. 21(2). 125-140.
- Purpura, J. E., & Kunnan, A. J. (Eds.) (2024). *The writings of L. F. Bachman: "Assuring that what we count counts."* New York: Routledge.
- QUEVEDO-CAMARGO, G, TONELLI, J. A.R., FURTOSO, V. A. B. (2024). Avaliação e formação de professores de línguas: contextos, necessidades, desafios e avanços [Assessment and language teacher education: needs, challenges and advances]. Special volume. *REVISTA X*, v. 19, p. 217-221. <https://revistas.ufpr.br/revistax/issue/view/3629>
- QUEVEDO-CAMARGO, G. (2020) Formação de professores de línguas adicionais e letramento em avaliação: breve panorama e desafios para os cursos de licenciatura em LEM no Brasil [Additional language teacher education and assessment literacy: a brief overview and challenges for undergraduate courses in Brazil]. *Calidoscópio (UNISINOS)*, v. 18, p. 435-459. DOI: <https://doi.org/10.4013/cld.2020.182.10>
- QUEVEDO-CAMARGO, G. & SCARAMUCCI, M. V. R. (2018) O conceito de letramento em avaliação de línguas: origem de relevância para o contexto brasileiro [The concept of language assessment literacy: origin and relevance for the Brazilian context]. *Linguagem: estudos e pesquisas*, v. 22, p. 225-245. DOI: <https://doi.org/10.5216/lep.v22i1.54474>
- QUEVEDO-CAMARGO, G., & SOUSA, A. A. C. P. (2022). Letramento em avaliação para professores de inglês e a perspectiva glocal: um microestudo [Language assessment literacy for language teachers and the glocal perspective: a micro-study]. *Revista Do GEL*, 19(1), 223–245. <https://doi.org/10.21165/gel.v19i1.3371>
- QUEVEDO-CAMARGO, G., & TONELLI, J. A. R. (2021) Contribuições de um curso EAD para o letramento em avaliação de professores de línguas adicionais para crianças [Contributions of a distance-mode course on assessment literacy for teachers of additional languages to young learners]. *Revista do GEL*, v. 18, p. 230-248. DOI: <https://doi.org/10.21165/gel.v18i1.3094>
- Rahmati, T., & Sadeghi, K. (2024). Technological innovations used in researching language assessment: A scoping review. In K. Sadeghi (Ed.), *Routledge handbook of*

- technological advances in researching language learning (pp. 118-129). <https://doi.org/10.4324/9781003459088-11>
- Rasooli, A., DeLuca, C. & Cheng, L. (2022). Beginning teacher candidates' approaches to grading and assessment conceptions—Implications for teacher education in assessment. *Educational Research for Policy and Practice*, 22, 63–90. <https://doi.org/10.1007/s10671-022-09320-5>
- Razavipour, K. (2023). Classroom writing assessment and feedback practices: A new materialist encounter. *Assessing Writing*, 57, Article 100760. <https://doi.org/10.1016/j.asw.2023.100760>
- Rezagah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Language Testing Journal*, 12(2), 76-94, <https://doi.org/10.22034/ijlt.2022.157126>
- Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: Hearing from Iranian university teachers. *Language Testing in Asia*, 11, Article 26. <https://doi.org/10.1186/s40468-021-00142-5>
- Roslan, R., Nishio, Y., & Jawawi, R. (2022). Analyzing English language teacher candidates' assessment literacy: A case of Bruneian and Japanese universities. *System*, 111, Article 102934. <https://doi.org/10.1016/j.system.2022.102934>
- Saeed, M. A., & Alharbi, M. A. (2023). Towards fostering Saudi EFL learners' collaborative engagement and feedback literacy in writing. *Assessing Writing*, 56, Article 100721. <https://doi.org/10.1016/j.asw.2023.100721>
- Saglam, A. L. G., & Tsagari, D. (2022). Perceptions towards the consequential validity of integrated language proficiency assessment'. *Languages*, 7(1), 65. <https://doi.org/10.3390/languages7010065>
- Seed, G., Salamoura, A. & Saville, N. (2024). New approaches to the assessment of English as an additional language. In N. A. Nazari & A. Mehdi Riazzi (Eds.), *Adaptable English Language Teaching: Advances and Frameworks for Responding to New Circumstances* (pp.79-95). Routledge.
- Shahzadi, A., & Ducasse, A. M. (2022). Language assessment literacy of teachers in an English medium of instruction university: Implications for ELT training in Pakistan. *Studies in Language Assessment*, 11(1), 92–118.
- Shan, Z., Yang, H., & Xu, H. (2023). The mediating role of curriculum configuration on teacher's L2 writing assessment literacy and practices in embedded French writing. *Assessing Writing*, 57, Article 100742. <https://doi.org/10.1016/j.asw.2023.100742>
- Sohrabi, Z., Ghanbari, N., & Abbasi, A. (2022). Iranian EFL teachers' perceptions of writing assessment literacy: a countrywide study. *Language Testing in Asia*, 12, Article 23. <https://doi.org/10.1186/s40468-022-00172-7>
- Su, X., & Lee, I. (2024). Is the magic in the mix? The development and validation of the L2 teachers' blended assessment literacy scale. *System*, 123, 103354. <https://doi.org/10.1016/j.system.2024.103354>
- Sun, Y., Lan, G., & Zhang, L. (2023). Pedagogical values of translingual practices in improving student feedback literacy in academic writing. *Assessing Writing*, 56, Article 100715. <https://doi.org/10.1016/j.asw.2023.100715>
- Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, 39(4), 649–678. <https://doi.org/10.1177/02655322211057040>

- Tavassoli, K., & Sorat, Z. (2023). Iranian EFL teachers' oral/aural skills language assessment literacy: Instrument development and validation. *International Journal of Language Testing*, 13(2), 56-76. <https://doi.org/10.22034/IJLT.2023.376466.1219>.
- Taylor, L. (2023). Reframing the discourse and rhetoric of language testing and assessment for the public square. *Language Testing*, 40(1), 47-53. <https://doi.org/10.1177/02655322221127421>
- Tayyebi, M., Abbasabady, M. M., & Abbassian, G. R. (2022). Examining classroom writing assessment literacy: a focus on in-service EFL teachers in Iran. *Language Testing in Asia*, 12, Article 12. <https://doi.org/10.1186/s40468-022-00161-w>
- Tenca, E., Hartle, S., & Piccinin, S. (2023). Teaching oral skills in Higher Education (HE) language learning: Piloting a model for inclusive, accessible English Language Teaching (ELT). *Instructed Second Language Acquisition*, 7(2), 130-165. <https://doi.org/10.1558/isla.26019>
- Tian, W., Lê, H. D., & Thúy Nguyễn, N. (2023). Vietnamese university EFL teachers' practices and literacy in classroom assessment: A sociocultural perspective. *System*, 116, Article 103066. <https://doi.org/10.1016/j.system.2023.103066>
- Tian, W., Louw, S., & Khan, M. K. (2021). Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes. *System*, 103, Article 102682. <https://doi.org/10.1016/j.system.2021.102682>
- TONELLI, J. A. R., & QUEVEDO-CAMARGO, G. (2018) Teaching EFL to children: reflections on (future) teachers' language assessment literacy. *Revista Horizontes de Lingüística Aplicada*, v. 17, p. 63-91. DOI: <https://doi.org/10.26512/rhla.v17i1.9209>
- Tsagari, D. (2021). Gauging the assessment literacy levels of English teachers in Norway. *The European Journal of Applied Linguistics and TEFL*, 10(1), 161–191.
- Tsagari, D., & Vogt, K. (2022). Introduction to the special issue: Contextualising language assessment literacy. *Studies in Language Assessment*, 11(1), ii-vii. https://www.altanz.org/uploads/5/9/0/8/5908292/0._sila_11_1__front_matter_introduction__tsagari__vogt.pdf
- Tsagari, D., & Vogt, K. (eds.) (2022) Contextualizing Language Assessment Literacy (Special Issue). *Studies in Language Assessment (SiLA)*, formerly *Papers in Language Testing and Assessment (PLTA)*, 11(1). <https://www.altanz.org/current-issue.html>
- Van Dyk, T., & Weideman, A. (2022). Taaltoetsing en die assessering van taalvermoë [Language testing and the assessment of language ability]. In W. Carstens & T. van Dyk (Eds.), *Toegepaste Taalkunde in Afrikaans [Applied linguistics in Afrikaans]* (pp. 125-1420). Van Schaik. <https://toegepaste-taalkunde.co.za/hoofstuk-7/>
- Villa Larenas, S., & Brunfaut, T. (2023). But who trains the language teacher educator who trains the language teacher? An empirical investigation of Chilean EFL teacher educators' language assessment literacy. *Language Testing*, 40(3), 463-492. <https://doi.org/10.1177/02655322221134218>
- Villa Larenas, S., & Brunfaut, T. (2024). How do teacher educators learn about and teach language assessment? In B. Baker, & L. Taylor (Eds.), *Language assessment literacy and competence: Volume 1: Research and reflections from the field* (pp. 66-87). (Studies in Language Testing; Vol. 55). Cambridge University Press. <https://www.cambridgeenglish.org/images/717470-silt-volume-55.pdf>

- Vogt, K. & Tsagari, D. (2022). Language assessment literacy and test validation: Highlighting the role of language teachers. *Studies in Language Assessment*, 11(1), 1–15.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0006/4282557/SiLA-11.1-Vogt_and_Tsagari.pdf
- Vogt, K., Bøhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching*, 57(3), 325–340. doi:10.1017/S0261444824000090
- Weideman, A. (2024). Yardsticks for the future of language assessment: Disclosing the meaning of measurement. In M. R. Salaberry, A. Weideman, A., & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 220-234). Routledge. <https://doi.org/10.4324/9781003384922-12>
- Wicaksono, B. H., Ismail, S. M., Sultanova, S. A., & Abeda, D. (2023). I like language assessment: EFL learners' voices about self-assessment, self-efficacy, grit tendencies, academic resilience, and academic demotivation in online instruction. *Language Testing in Asia*, 13, Article 37 (2023).
<https://doi.org/10.1186/s40468-023-00252-2>
- Wu, P., & Lei, C. (2023). Developing feedback literacy through dialogue-supported performances of multi-draft writing in a postgraduate class. *Assessing Writing*, 57, Article 100759. <https://doi.org/10.1016/j.asw.2023.100759>
- Wu, P., Yu, S., & Luo, Y. (2023). The development of teacher feedback literacy in situ: EFL writing teachers' endeavor to human-computer-AWE integral feedback innovation. *Assessing Writing*, 57, Article 100739. <https://doi.org/10.1016/j.asw.2023.100739>
- Wudthayagorn, J. (2022). An exploration of the English Exit Examination policy in Thai public universities. *Language Assessment Quarterly*, 19(2), 107-123.
<https://doi.org/10.1080/15434303.2021.1937174>
- Xu, J., & Zheng, Y. (2024). Understanding the interconnectedness of L2 learners' writing assessment literacy and engagement: A longitudinal survey. *System*, 123, 103318.
<https://doi.org/10.1016/j.system.2024.103318>
- Xu, J., Zheng, Y., & Braund, H. (2023). Voices from L2 learners across different languages: Development and validation of a student writing assessment literacy scale. *Journal of Second Language Writing*, 60, 100993. <https://doi.org/10.1016/j.jslw.2023.100993>
- Xu, Y., Chen, J., & He, L. (2023). How Personal, Experiential, and Contextual Factors Mediate EFL Teachers' Conceptions of Assessment: A Narrative Study. *Chinese Journal of Applied Linguistics*, 46(2), 237-252.
<https://doi.org/10.1515/CJAL-2023-0206>
- Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
- Yang, R., & May, L. (2023). Understanding and Implementing Learning-Oriented Assessment: Evidence From Three Chinese Junior High School EFL Teachers. *Chinese Journal of Applied Linguistics*, 46(2), 308-325.
<https://doi.org/10.1515/CJAL-2023-0210>
- Yeo, M. (2023). Peer Review Circles: The Value of Working in Triadic Groups and Oral Observational Peer Feedback in the Development of Student Feedback Literacy. *Chinese Journal of Applied Linguistics*, 46(2), 270-288.
<https://doi.org/10.1515/CJAL-2023-0208>
- Yongqi GU, P. & LAM, R. (2023). Developing Assessment Literacy for Classroom-Based Formative Assessment. *Chinese Journal of Applied Linguistics*, 46(2), 155-161.

- <https://doi.org/10.1515/CJAL-2023-0201>
- Yu–Ting, K. (2023). Empowering preservice English teachers with language assessment literacy concepts and practices: Application of Vygotskian concept-based language instruction. *Modern Language Journal*, 107, 68–94.
<https://doi.org/10.1111/modl.12818>
- Yu, C., & Xu, W. (2024). Writing assessment literacy and its impact on the learning of writing: A netnography focusing on Duolingo English Test examinees. *Language Testing in Asia*, 14, Article 24. <https://doi.org/10.1186/s40468-024-00297-x>
- Yu, G. & Xu, J. (Eds.). (2024). Language test validation in a digital age. Studies in Language Testing Volume 52. Cambridge University Press & Assessment.
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48, Article 100525.
<https://doi.org/10.1016/j.asw.2021.100525>
- Yu, S., Zhang, E. D., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing*, 53, Article 100643.
<https://doi.org/10.1016/j.asw.2022.100643>
- Zabala Delgado, J., & Rodriguez, C. (2022). Language assessment literacy of language teachers in the context of adult education in Spain. *Studies in Language Assessment*, 11(1), 64–91.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0009/4282560/SiLA-11.1-Delgado_and_Rodriguez.pdf
- Zhang-Wu, Q. (2023). Exploring multilingual students' feedback literacy in an asynchronous online writing course. *Assessing Writing*, 56, Article 100718.
<https://doi.org/10.1016/j.asw.2023.100718>
- Zhang, F., Min, H., He, P., Chen, S., & Ren, S. (2023). Understanding EFL students' feedback literacy development in academic writing: A longitudinal case study. *Assessing writing*, 58, Article 100770. <https://doi.org/10.1016/j.asw.2023.100770>
- Zhang, T., & Mao, Z. (2023). Exploring the development of student feedback literacy in the second language writing classroom. *Assessing Writing*, 55, Article 100697.
<https://doi.org/10.1016/j.asw.2023.100697>
- Zhou, J., Deneen, C., Tai, J., & Dawson, P. (2023). Feedback seeking by first-year Chinese international students: Understanding practices and challenges. *Assessing Writing*, 57, Article 100757. <https://doi.org/10.1016/j.asw.2023.100757>
- Zimmermann, S., Schwarz, L., Peters, A. & Depner, G. (2024). Enhancing teachers and test takers' assessment literacy? Insights from test preparation for the digital TestDaF. In: B. Baker & L. Taylor (Eds.), *Language Assessment Literacy and Competence Volume 1: Research and Reflections from the Field* (pp. 110–128). Cambridge: Cambridge University Press.

Approaches to Assessment

- Abbott, M., Lee, K. & Ricioppo, S. (2021). Does portfolio-based language assessment align with learning-oriented assessment? Evidence from literacy learners and their Instructors. *Canadian Journal of Applied Linguistics / Revue Canadienne de Linguistique Appliquée*, 24(2), 260–285. <https://doi.org/10.37213/cjal.2021.31338>

- Abdolrezapour, P., & Ghanbari, N. (2021). Enhancing learning potential score in EFL listening comprehension and self-regulation through self-regulated dynamic assessment procedures. *Language Testing in Asia*, 11, Article 10. <https://doi.org/10.1186/s40468-021-00126-5>
- Abdullaeva, B. S., Çakmak, F., & Abdullaev, D. (2024). Paper assessment or online assessment: Exploring the impact of assessment modes on EFL students' language learning outcomes and personal development. *Language Testing in Asia*, 14(1), Article 35. <https://doi.org/10.1186/s40468-024-00309-w>
- Abrar-ul-Hassan, S., Douglas, D., & Turner, J. (2021). Revisiting second language portfolio assessment in a new age. *System*, 103, Article 102652. <https://doi.org/10.1016/j.system.2021.102652>
- Acar, A. S. (2023). Genre pedagogy: A writing pedagogy to help L2 writing instructors enact their classroom writing assessment literacy and feedback literacy. *Assessing Writing*, 56, Article 100717. <https://doi.org/10.1016/j.asw.2023.100717>
- Ahmed Abdel-Al Ibrahim, K., Cuba Carbajal, N., Zuta, M. E. C., & Bayat, S. (2023). Collaborative learning, scaffolding-based instruction, and self-assessment: Impacts on intermediate EFL learners' reading comprehension, motivation, and anxiety. *Language Testing in Asia*, 13, Article 16. <https://doi.org/10.1186/s40468-023-00229-1>
- Al-Rashidi, A. H., Asif, M., Vanani, M. G., & Aberash, A. (2022). Learner-oriented assessment (LOA) practice: The comparative study of self-assessment, peer assessment, and teacher assessment on EFL learners' writing complicity, accuracy, and fluency (CAF), speaking CAF, and attitude. *Language Testing in Asia*, 12, Article 59. <https://doi.org/10.1186/s40468-022-00209-x>
- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Aldosari, M., Heydarnejad, T., Hashemifardnia, A., & Abdalgane, M. (2023). The interplay among self-assessment, using reflection for assessment, classroom enjoyment, and immunity: Into prospects of effective language learning. *Language Testing in Asia*, 13, Article 1. <https://doi.org/10.1186/s40468-022-00213-1>
- Ali, M. H., & Abdulrazzaq, D. M. (2024). Dynamic versus static assessment: Effects on Iraqi EFL learners' resilience, burnout, positive orientation, and academic success. *Language Testing in Asia*, 14, Article 37. <https://doi.org/10.1186/s40468-024-00308-x>
- Allal, L. (2021). Involving primary school students in the co-construction of formative assessment in support of writing. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 584–601. <https://doi.org/10.1080/0969594X.2021.1951164>
- Archila, P. A., Molina, J., Danies, G., Truscott de Mejía, A.-M., & Restrepo, S. (2022). Using formal formative assessment (FFA) to promote undergraduates' bilingual written scientific argumentation (BWSA). *International Journal of Bilingual Education and Bilingualism*, 25(9), 3137–3150. <https://doi.org/10.1080/13670050.2021.2021386>
- Arques, A. C., & Ferrero, C. L. (2023). Peer-feedback of an occluded genre in the Spanish language classroom: A case study. *Assessing Writing*, 57, Article 100756. <https://doi.org/10.1016/j.asw.2023.100756>

- Baker, B., Polikar, S., & Homayounzadeh, M. (2021). "I was not thoughtful enough before:" Exploiting the learning-oriented-assessment potential of a test taker-oriented rubric for summative assessment. In A. Gebriel. (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 162–181). Routledge.
<https://doi.org/10.4324/9781003014102-12>
- Ball, J. (2023). Encouraging EAP students to engage with the literature of their disciplines. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 61–72). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Bao, G. (2024). Differential effects of EFL proficiency and mode on noun phrase use. *System*, 121, 103244. <https://doi.org/10.1016/j.system.2024.103244>
- Barker, B. (2023). EAL/D or an additional need? A cycle of learning, teaching and assessment to discern if an EAL/D student has additional needs to learning English. *Studies in Language Assessment*, 12(2), 272-285.
<https://doi.org/10.58379/XSLO8306>
- Batista, E. G. & Moraes, I. T. (2020). Avaliação orientada para a aprendizagem no ensino de línguas para crianças [Learning-oriented language assessment in the context of young learners]. *REVELLI*, 12, 1-19. <https://doi.org/10.51913/revelli.v12i0.10234>
- Berberovic, V. (2022). *Implementing peer feedback for writing tasks* (Research Notes Issue 83). Cambridge University Press and Assessment.
<https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing*, 40(1), 133-152. <https://doi.org/10.1177/02655322211064629>
- Bowles, M. A. (2022). Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, 39(3), 355–376. <https://doi.org/10.1177/02655322221076033>
- Braund, H., DeLuca, C., Panadero, E., & Cheng, L. (2021). Exploring formative assessment and co-regulation in kindergarten through interviews and direct observation. *Frontiers in Education: Educational Psychology*, 6, Article 732373.
<https://doi.org/10.3389/feduc.2021.732373>
- Braund, H., DeLuca, C., Panadero, E., & Cheng, L. (2021). Exploring formative assessment and co-regulation in kindergarten through interviews and direct observation. *Frontiers in Education: Educational Psychology*, 6, Article 732373.
<https://doi.org/10.3389/feduc.2021.732373>
- Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, 40(1), 15-23. <https://doi.org/10.1177/02655322221127896>
- Brunfaut, T., & Harding, L. (2024). Diagnostic approaches in teaching and assessing listening. In E. Wagner, A. O. Batty, & E. Galaczi (Eds.), *The Routledge Handbook of Second Language Acquisition and Listening* (pp. 255-267). (Routledge Handbooks in Second Language Acquisition). Routledge.
<https://www.routledge.com/The-Routledge-Handbook-of-Second-Language-Acquisition-and-Listening/Wagner-Batty-Galaczi/p/book/9781032113647>
- Burns, A. (2023). Action research for sustained and powerful professional learning. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 5–14). Cambridge University Press & Assessment.

- <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Burns, A., O'Keefe, S., & Savage, J. (2024). *Research notes 87: Findings of action research in ELICOS program 2023*. Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/714674-research-notes-87.pdf>
- Burton, J. D. (2023). Reflections on the past and future of language testing and assessment: An emerging scholar's perspective. *Language Testing*, 40(1), 24-30.
<https://doi.org/10.1177/02655322221126607>
- Çakmak, F., Ismail, S. M., & Karami, S. (2023). Advancing learning-oriented assessment (LOA): Mapping the role of self-assessment, academic resilience, academic motivation in students' test-taking skills, and test anxiety management in Telegram-assisted-language learning. *Language Testing in Asia*, 13, Article 20.
<https://doi.org/10.1186/s40468-023-00230-8>
- Campbell, M. & Rosenhan, C. (2024). Assessment and creativity through a translingual lens: transdisciplinary insights. *Applied Linguistics Review*, 15(5), 1839-1859.
<https://doi.org/10.1515/applirev-2023-0090>
- Carr, N. (2023). Feedback on writing through the lens of activity theory: An exploration of changes to peer-to-peer interactions. *Assessing Writing*, 56, Article 100720.
<https://doi.org/10.1016/j.asw.2023.100720>
- Chan, J. Y. H. (2022). The evolution of assessment in English pronunciation: The case of Hong Kong (1978-2018). *Language Assessment Quarterly*, 19(1), 1-26.
<https://doi.org/10.1080/15434303.2021.1935960>
- Chappuis, S., Brookhart, S., & Chappuis, J. (2021). *Ten assessment literacy goals for school leaders*. SAGE.
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723.
<https://doi.org/10.1016/j.asw.2023.100723>
- Chong, S. W., & Isaacs, T. (2023). An ecological perspective on classroom-based assessment. *TESOL Quarterly*, 57(4), 1558-1570. <http://doi.org/10.1002/tesq.3201>
- Chong, S. W., & Reinders, H. (Eds.). (2023). *Innovation in learning-oriented language assessment*. Springer.
- Chong, S. W., Isaacs, T., & McKinley, J. (2023). Research timeline: Ecological systems theory and L2 research. *Language Teaching*, 56(3), 333-348.
<http://doi.org/10.1017/S0261444822000283>
- Clews, P. (2022). *Towards a peer feedback scaffold* (Research Notes Issue 83). Cambridge University Press and Assessment.
<https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- deBoer, M., Leontjev, D., & Friederich, L. (2023). From language to function: developing self-and peer-assessment tools. *ELT Journal*, 77(1), 94-104.
<https://doi.org/10.1093/elt/ccac014>
- Derakhshan, A., & Fathi, J. (2024). Growth mindset, self-efficacy, and self-regulation: A symphony of success in L2 speaking. *System*, 123, 103320.
<https://doi.org/10.1016/j.system.2024.103320>
- Derakhshan, A., & Ghiasvand, F. (2022). Demystifying Iranian EFL teachers' perceptions and practices of learning-oriented assessment (LOA): challenges and prospects in focus. *Language Testing in Asia*, 12, Article 55.
<https://doi.org/10.1186/s40468-022-00204-2>

- Di Zhang, E., & Yu, S. (2023). Investigating the relationship between linguistic changes in L2 writers' paraphrasing, paraphrasing performance and L2 proficiency. *Applied Linguistics Review*, 14(5), 1451-1473. <https://doi.org/10.1515/applirev-2021-0014>
- Duke, G. (2023). Training L2 learners to be autonomous through instructional scaffolding. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 83–104). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Ebrahimi, B., & Ebadi, S. (2024). Exploring the impact of screencast-based ipsative assessment on EFL students' writing development. *Language Testing in Asia*, 14, Article 41. <https://doi.org/10.1186/s40468-024-00315-y>
- Estaji, M., & Banitalebi, Z. (2023). A study of Test-Taking strategies of Iranian IELTS repeaters: Any change in the strategy use? *International Journal of Testing*, 23(3), 205-230. <https://doi.org/10.1080/15305058.2023.2195662>
- Estaji, M., & Kogani, M. (2024). Assessment polarity or praxis parity? Uncovering the learning-oriented assessment (LOA) landscape of novice and experienced Iranian EFL teachers. *Language Testing in Asia*, 14, Article 34. <https://doi.org/10.1186/s40468-024-00310-3>
- Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices. *Language Testing in Asia*, 11, Article 2. <https://doi.org/10.1186/s40468-020-00117-y>
- Fazel, I., & Ali, A. M. (2022). EAP teachers' knowledge and use of learning-oriented assessment: A cross-contextual study. *System*, 104, Article 102685. <https://doi.org/10.1016/j.system.2021.102685>
- Fernández Álvarez, M., & Gordenstein Montes, A. L. (Eds.). (2023). *Global perspectives on effective assessment in English language teaching*. IGI Global. <https://doi.org/10.4018/978-1-6684-8213-1>
- Fox, J., Abdulhamid, N. & Turner, C.E. (2022). Classroom based assessment. In G. Fulcher & L. Harding (Eds.), *Routledge handbook of language testing*, 2nd ed. (pp. 119-135). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781003220756-6>
- Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. *Studies in Second Language Acquisition*, 44(1), 2-34. <https://doi.org/10.1017/S0272263120000388>
- Fukuda, S. T., Lander, B. W., & Pope, C. J. (2022). Formative assessment for learning how to learn: Exploring university student learning experiences. *RELC Journal*, 53(1), 118-133. <https://doi.org/10.1177/0033688220925927>
- Gao, S., Gales, M., & Xu, J. (2024). Detecting aberrant responses in automated L2 spoken English assessment. In C. Chapelle, G. H. Beckett & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 96-117). Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.07>
- Gebril, A. (2021). Learning-oriented assessment: Main issues and an overview. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 1–10). Routledge. <https://doi.org/10.4324/9781003014102-1>
- Gebril, A. (2021). *Learning-oriented language assessment: Putting theory into practice*. Routledge. <https://doi.org/10.4324/9781003014102>
- Gebril, A. (2021). Nation-wide assessment reform in Egypt: Challenges and potentials of learning-oriented assessment in an examinations-based instructional setting. In A.

- Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 109–122). Routledge. <https://doi.org/10.4324/9781003014102-9>
- Ghaneiarani, S., Alavi, S. M., & Kaivanpanah, S. (2024). Enhancing writing ability of Iranian EFL learners through learning-oriented assessment: Peer and teacher feedback implementation. *Language Testing in Asia*, 14, Article 39. <https://doi.org/10.1186/s40468-024-00298-w>
- Ginther, A. (2023). Administration, labor, and love. *Language Testing*, 40(1), 31-39. <https://doi.org/10.1177/02655322221127365>
- Goh, R., & Kiat Tan, K. (2023). Teachers' qualitatively different ways of experiencing assessment feedback: Implications for teacher assessment literacy. *Chinese Journal of Applied Linguistics*, 46(2), 253-269. <https://doi.org/10.1515/CJAL-2023-0207>
- Gomez Laich, M., & Taguchi, N. (2023). Effects of task complexity on the learning of genre specific rhetorical moves and linguistic forms: the case of contrast and argumentative essays. *International Review of Applied Linguistics in Language Teaching*, 61(2), 571-599. <https://doi.org/10.1515/iral-2021-0111>
- Gottlieb, M. (2021). *Assessment in multiple languages: A handbook for school and district leaders*. SAGE.
- Gottlieb, M. (2021). *Classroom assessment in multiple languages: A handbook for teachers*. Corwin.
- Gràcia, M., Adam-Alcocer, A., & Castillo Mardones, P. (2023). Exploring the impact of a teacher development programme using a digital application on linguistic interactions in the classroom: a multiple case study. *Applied Linguistics Review*, 14(5), 1305–1343. <https://doi.org/10.1515/applirev-2020-0132>
- Graham, S., Harbaugh-Schattenkirk, A. G., Aitken, A., Harris, K. R., Ng, C., Ray, A., Wilson, J. M., & Wdowin, J. (2022). Writing motivation questionnaire: Validation and application as a formative assessment. *Assessment in Education: Principles, Policy & Practice*, 29(2), 238–261. <https://doi.org/10.1080/0969594X.2022.2080178>
- Greenblatt, P., & McDonald, P. (2024). Using Multimodal Public Service Announcement Video Projects as Communicative Language Assessments. *RELC Journal*, 55(2), 559–566. <https://doi.org/10.1177/00336882221102234>
- Gujord, A.-K. H. (2023). Who succeeds and who fails? Exploring the role of background variables in explaining the outcomes of L2 language tests. *Language Testing*, 40(2), 227–248. <https://doi.org/10.1177/02655322221100115>
- Haddadian, G., Radmanesh, S., & Haddadian, N. (2024). Construction and validation of a Computerized Formative Assessment Literacy (CFAL) questionnaire for language teachers: An exploratory sequential mixed-methods investigation. *Language Testing in Asia*, 14, Article 33. <https://doi.org/10.1186/s40468-024-00303-2>
- Haghi, I., Clark, T. (2024). Assessment in the EAP Classroom. In E. Zehir Topkaya & H. Çelik (Eds.), *Teaching English for Academic Purposes*. Palgrave Macmillan/Cham. https://doi.org/10.1007/978-3-031-72545-6_13
- Hammad Al-Rashidi, A., Vadivel, B., Ramadan Khalil, N., & Basim, N. (2023). The comparative impacts of portfolio-based assessment, self-assessment, and scaffolded peer assessment on reading comprehension, vocabulary learning, and grammatical accuracy: insights from working memory capacity. *Language Testing in Asia*, 13, Article 24 (2023). <https://doi.org/10.1186/s40468-023-00237-1>
- Han, C., & Zhao, X. (2021). Accuracy of peer ratings on the quality of spoken-language interpreting. *Assessment and Evaluation in Higher Education*, 46(8), 1299–1313.

- <https://doi.org/10.1080/02602938.2020.1855624>
- Harrasi, K. T. S. AL. (2023). Reexamining feedback in the context of different rhetorical patterns of writing. *Language Testing in Asia*, 13, Article 4 (2023).
<https://doi.org/10.1186/s40468-023-00216-6>
- Heidari Vinchek, M., Mirzaei, A., & Roohani, A. (2024). A cognitive diagnostic approach to IELTS speaking test: Unveiling the subskills and test-takers' perceptions. *Language Testing in Asia*, 14, Article 42. <https://doi.org/10.1186/s40468-024-00311-2>
- Heritage, M. (2021). *Formative assessment: Making it happen in the classroom* (2nd ed.). SAGE.
- Heydarnejad, T., Ismail, S. M., Shakibaei, G., & Saeedian, A. (2022). Modeling the impact of L2 grit on EFL learners' core of self-assessment and foreign language anxiety. *Language Testing in Asia*, 12, Article 50. <https://doi.org/10.1186/s40468-022-00200-6>
- Heydarnejad, T., Saberi Dehkordi, E., & Parmadi, B. (2022). Learning-oriented assessment in the classroom: the contribution of self-assessment and critical thinking to EFL learners' academic engagement and self-esteem. *Language Testing in Asia*, 12, Article 60. <https://doi.org/10.1186/s40468-022-00210-4>
- Hirsu, L., Nitecka Barche, M. & Futro, D. (2024). Assessment and creativity through a translingual lens: introduction. *Applied Linguistics Review*, 15(5), 1831-1838.
<https://doi.org/10.1515/applirev-2023-0115>
- Homayouni, M. (2022). Peer assessment in group-oriented classroom contexts: on the effectiveness of peer assessment coupled with scaffolding and group work on speaking skills and vocabulary learning. *Language Testing in Asia*, 12, Article 61. <https://doi.org/10.1186/s40468-022-00211-3>
- Hong, J. (2023). Content teachers' and lecturers' corrective feedback in EMI classes in high school and university settings. *Studies in Second Language Learning and Teaching*, 13(2), 451–469. <https://doi.org/10.14746/ssllt.38282>
- Hosseinpur, R. M., & Kazemi, Z. (2022). Composing strategies employed by high-and low-performing Iranian EFL students in essay writing classes. *Assessing Writing*, 51, Article 100601. <https://doi.org/10.1016/j.asw.2021.100601>
- Hougham, D., Clenton, J., & Uchihara, T. (2024). Disentangling the contributions of shorter vs. Longer lexical bundles to L2 oral fluency. *System*, 121, 103243. <https://doi.org/10.1016/j.system.2024.103243>
- Huhta, A., Harsch, C., Leontjev, D., & Nieminen, L. (2024). The diagnosis of writing in a second or foreign language. Routledge.
- Inbar-Lourie, O. (2022). EMI programs and formative assessment: Implications for the assessment literacy of content lecturers. *Journal of English Medium Instruction* 1(2), 204-231. <https://doi.org/10.1075/jemi.21014.inb>
- Isbell, D. R. (2021). Can the test support student learning? Validating the use of a second language pronunciation diagnostic. *Language Assessment Quarterly*, 18(4), 331–356. <https://doi.org/10.1080/15434303.2021.1874382>
- Isbell, D., & Sakai, M. (2022). Pronunciation assessment in classroom contexts. In T. Derwing, J. Levis, & S. Sonsaat (Eds.), *Pronunciation in Second Language Learning and Teaching: Innovations and Developments in Research and Teaching* (pp. 194-214). Routledge.
- Isbell, D.R., & Lee, J. (2022). Self-Assessment of comprehensibility and accentedness in second language Korean. *Language Learning*, 72(3), 806-852.
<https://doi.org/10.1111/lang.12497>

- Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, 12, Article 40. <https://doi.org/10.1186/s40468-022-00191-4>
- Janssen, G. (2022). Local placement test retrofit and building language assessment literacy with teacher stakeholders: A case study from Colombia. *Language Testing*, 39(3), 377–400. <https://doi.org/10.1177/02655322221076153>
- Jia, L., Cai, J., & Wang, J. (2023). Promoting learning potential among students of L2 Chinese through dynamic assessment. *Language Assessment Quarterly*, 20(1), 66–87. <https://doi.org/10.1080/15434303.2022.2156870>
- Jung, D., & Randazzo, K. (2022). *The upward cycle: learner progress through critical reflection and strategic response* (Research Notes Issue 83). Cambridge University Press and Assessment. <https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Jung, J., & Fu, X. (2023). The impact of pragmalinguistic support on video-conferenced collaborative suggestion-giving task. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1083–1107. <https://doi.org/10.1515/iral-2021-0138>
- Kahn-Horwitz, J., & Goldstein, Z. (2024). English foreign language reading and spelling diagnostic assessments informing teaching and learning of young learners. *Language Testing*, 41(1), 60–88. <https://doi.org/10.1177/02655322231162838>
- Kang, M., & Lam, R. (2024). Understanding university English instructors' assessment literacy: A formative assessment perspective. *Language Testing in Asia*, 14, Article 52. <https://doi.org/10.1186/s40468-024-00323-y>
- Kao, Y-T. (2022). Effects of group dynamic assessment on L2 Chinese learners' literacy development: learners' responsiveness to interactive mediation. *Applied Linguistics Review*, 13(5), 843–871. <https://doi.org/10.1515/applirev-2019-0077>
- Kennedy, E., & Shiel, G. (2022). Writing assessment for communities of writers: rubric validation to support formative assessment of writing in Pre-K to grade 2. *Assessment in Education: Principles, Policy & Practice*, 29(2), 127–149. <https://doi.org/10.1080/0969594X.2022.2047608>
- Kim, H. S., Lee, J. H., & Lee, H. (2024). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, 28(1), 7-28. <https://doi.org/10.1177/1362168820981394>
- Kim, S., & Kessler, M. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, Article 100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Kim, Y., & Emeljanova, L. (2021). The effects of written corrective feedback on the accuracy of L2 writing: Comparing collaborative and individual revision behavior. *Language Teaching Research*, 25(2), 234–255. <https://doi.org/10.1177/1362168819831406>
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *English for Academic Purposes*, 62, Article 101247. <https://doi.org/10.1016/j.jeap.2023.101247>
- Kirkwood, M. (2023). Using ipsative feedback for personal bests in grammatical accuracy in L2 writing. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 105–128). Cambridge University Press & Assessment.

- <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142-161.
<https://doi.org/10.1080/15434303.2021.2023542>
- Koster, T., & Morgan, G. (2023). Investigating students' perceptions of effective engagement methods for online and multi-modal ELICOS classes. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 43–60). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Kushki, A., Rahimi, M., & Davin, K. J. (2022). Dynamic assessment of argumentative writing: mediating task response. *Assessing Writing*, 52, Article 100606.
<https://doi.org/10.1016/j.asw.2022.100606>
- Kyle, K., & Eguchi, M. (2024). Evaluating NLP models with written and spoken L2 samples. *Research Methods in Applied Linguistics*, 3(2), Article 100120.
<https://doi.org/10.1016/j.rmal.2024.100120>
- Lam, D. M. K. (2021). Feedback as a learning-oriented assessment practice: Principles, opportunities, and challenges. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 86–107). Routledge.
<https://doi.org/10.4324/9781003014102-7>
- Lam, R. (2024). E-portfolios for language assessment. In R. Lam (Ed.), *Integrating e-portfolios into L2 classrooms: Education for future* (pp. 44-56). Multilingual Matters. <https://doi.org/10.2307/jj.21995549.8>
- Latif, M. W., & Wasim, A. (2024). Investigating EFL instructors' approaches to classroom-based assessment culture: An explanatory sequential mixed-method approach. *Language Testing in Asia*, 14, Article 57.
<https://doi.org/10.1186/s40468-024-00332-x>
- Lau, W. M. E., Aubrey, S., & Ellis, R. (2024). Error logs in the second language classroom: Exploring the relationship between learner engagement with written corrective feedback and improvements in writing accuracy. *System*, 123, 103329.
<https://doi.org/10.1016/j.system.2024.103329>
- Le Bouthillier, J., & Kristmanson, P. (2023). Becoming a French second language teacher: Supporting confidence and competence. *Second Language Teacher Education*, 2(1), 21–42. <https://doi.org/10.1558/slte.24011>
- Lee, H-W., Mullooly, A., Devine, A., & Galaczi, E. (2024). Exploring Interaction in Video-call Paired Speaking Tests: A Look at Scores, Language, and Perceptions, *Applied Linguistics*, 45(3), 533–552. <https://doi.org/10.1093/applin/amad046>
- Lee, I. (2023). Problematising written corrective feedback: A global Englishes perspective. *Applied Linguistics*, 44(4), 791–796. <https://doi.org/10.1093/applin/amad038>
- Lee, K. (2023). Diagnosing Chinese college-level English as a Foreign Language (EFL) learners' integrated writing capability: A Log-linear Cognitive Diagnostic Modeling (LCDM) study. *Assessing Writing*, 57, Article 100730.
<https://doi.org/10.1016/j.asw.2023.100730>
- Leontjev, D., & Pollari, P. (2023). Guiding and assessing development of L2 writing process: The role of peer collaboration. *Innovation in Language Learning and Teaching*, 17(2), 440–452. <https://doi.org/10.1080/17501229.2022.2058514>
- Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2), 189–218.

- <https://doi.org/10.1177/0265532220932481>
- Li, M., Zhang, X., & Reynolds, B. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: an ETS corpus study. *Applied Linguistics Review*, 14(4), 847–873. <https://doi.org/10.1515/applirev-2020-0129>
- Ma, M. (2023). Exploring learning-oriented assessment in EAP writing classrooms: Teacher and student perspectives. *Language Testing in Asia*, 13, Article 33 (2023). <https://doi.org/10.1186/s40468-023-00249-x>
- Ma, M., & Bui, G. (2022). Implementing continuous assessment in an academic English writing course: An exploratory study. *Assessing Writing*, 53, Article 100629. <https://doi.org/10.1016/j.asw.2022.100629>
- Mabrouk, F. (2023). Tunisian EFL learners' perceptions of the use of oral presentations in assessing speaking skills: An exploratory practice study. *Studies in Applied Linguistics and TESOL*, 23(3), 18–40. <https://doi.org/10.52214/salt.v23i2.12083>
- Mäkipää, T., & Soltyska, A. (2024). English teachers' perceptions of feedback practices under emergency remote teaching conditions: Insights for future use. *Studies in Language Assessment (SILA)*, 13(2), 73–114. <https://doi.org/10.58379/mszy9375>
- Maplethorpe, L., Kim, H., Hunte, M. R., Vincett, M., & Jang, E. E. (2022). Student-generated questions in literacy education and assessment. *Journal of Literacy Research*, 54(1), 74–97. <https://doi.org/10.1177/1086296X221076436>
- Martín-Gascón, B. (2023). Developing L2 learners' metaphoric competence: a case study of figurative motion constructions. *International Review of Applied Linguistics in Language Teaching*, 61(1), 79–109. <https://doi.org/10.1515/iral-2022-0043>
- Matta, M., Mercer, S. H., & Keller-Margulis, M. A. (2022). Evaluating validity and bias for hand-calculated and automated written expression curriculum-based measurement scores. *Assessment in Education: Principles, Policy & Practice*, 29(2), 200–218. <https://doi.org/10.1080/0969594X.2022.2043240>
- Mawer, V. (2023). Am I talking enough? Evaluating the effectiveness of quantitative speaking feedback. In A. Burn & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 73–82). Cambridge University Press & Assessment. <https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2021). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 34(7), 868–897. <https://doi.org/10.1080/09588221.2019.1645181>
- Meletiadou, E., & Tzagari, D. (2022). Exploring EFL teachers' perceptions of the use of peer assessment in external exam-dominated writing classes. *Languages*, 7(1), 16. <https://doi.org/10.3390/languages7010016>
- Mesina, V. G. (2023). The “Problem” of L2 writers in college composition placement: Identity, outcomes, and the future of directed self placement. *Studies in Applied Linguistics and TESOL*, 22(2), 23–36. <https://doi.org/10.52214/salt.v22i2.10682>
- Min, S., & He, L. (2022). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, 39(1), 90–116. <https://doi.org/10.1177/0265532221995475>
- Min, S., Cai, H., & He, L. (2022). Application of bi-factor MIRT and higher-order CDM models to an in-house EFL listening test for diagnostic purposes. *Language Assessment Quarterly*, 19(2), 189–213. <https://doi.org/10.1080/15434303.2021.1980571>

- Morton, T., & Nashaat-Sobhy, N. (2024). Exploring Bases of achievement in content and Language Integrated assessment in a bilingual education program. *TESOL Quarterly*, 58(1), 5–31. <https://doi.org/10.1002/tesq.3207>
- Nakamura, S., Phung, L., & Reinders, H. (2021). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*, 43(2), 428–441. <https://doi.org/10.1017/S0272263121000437>
- Namaziandost, E., Kargar Behbahani, H., & Heydarnejad, T. (2024). Tapping the alphabets of learning-oriented assessment: Self-assessment, classroom climate, mindsets, trait emotional intelligence, and academic engagement are in focus. *Language Testing in Asia*, 14, Article 21. <https://doi.org/10.1186/s40468-024-00293-1>
- Nguyen, T. T., Richardson, T., Nguyen, A. N., Vu, T. N., & Dang, T. T. H. (2023). A systematic review of potential opportunities and challenges to the use of portfolios in Vietnam as an assessment tool. *Innovation in Language Learning and Teaching*, 17(5), 894–908. <https://doi.org/10.1080/17501229.2023.2194038>
- Nicholas, A., & Blake, J. (2024). Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing. *Research Methods in Applied Linguistics*, 3(3), Article 100164. <https://doi.org/10.1016/j.rmal.2024.100164>
- Norris, J.M., Sasayama, S. and Kim, M. (2023), Simulating real-world context in an email writing task: Implications for task-based language assessment. *ETS Research Report Series*, 2023, 1-22. <https://doi.org/10.1002/ets2.12366>
- Nurjain, A., Salazar-Espinoza, D. E., Saenko, N., & Bina, E. (2023). Learner-oriented assessment matters: testing the effects of academic buoyancy, reflective thinking, and learner enjoyment in self-assessment and test-taking anxiety management of the EFL learners. *Language Testing in Asia*, 13, Article 30 (2023). <https://doi.org/10.1186/s40468-023-00247-z>
- O'Sullivan, B. (2023). Language Testing. In L. Wei, Z. Hua, & J. Simpson (Eds.), *Routledge Handbook of Applied Linguistics* (2nd ed., pp.136-149). Routledge. <https://doi.org/10.4324/9781003082644>
- Oliver, G. P., & Gregory, K. M. (2024). Integrated Performance Assessments: providing equitable instruction and assessment for ELLs/MLLs. *TESOL Quarterly*, 58(4), 1815–1830. <https://doi.org/10.1002/tesq.3324>
- Oliver, R., Chen, H., & Dovchin, S. (2024). Review of selected research in applied linguistics published in Australia (2015–2022). *Language Teaching*, 57(3), 341–376. [doi:10.1017/S0261444824000077](https://doi.org/10.1017/S0261444824000077)
- Orabah, S. S. B., Bijani, H., & Ismail, S. M. (2022). Assessing English language teachers' understanding and practices of student-centered learning in Oman. *Language Testing in Asia*, 12, Article 41. <https://doi.org/10.1186/s40468-022-00184-3>
- Osidak, V., Kvasova, O., Kavytska, T., & Drobotun, V. (2024). Book Creator as a tool of formative assessment in the TEFL classroom: Emergency remote teaching during wartime. *Studies in Language Assessment (SILA)*, 13(2), 1-37. <https://doi.org/10.58379/RBXT4093>
- Ożańska-Ponikwia, K., Piechurska-Kuciel, E., & Skłacka, K. (2023). Emotional intelligence as a mediator in the relationship between neuroticism and L2 achievement. *Applied Linguistics Review*, 14(1), 67-86. <https://doi.org/10.1515/applirev-2020-0008>
- Papi, M., Wolff, D., Nakatsukasa, K., & Bellwoar, E. (2021). Motivational factors underlying learner preferences for corrective feedback: Language mindsets and achievement goals. *Language Teaching Research*, 25(6), 858–877.

- <https://doi.org/10.1177/13621688211018808>
- Parker, S. (2023). Building connection: An EAP course that facilitates engagement with the university community. In A. Burn, & S. O'Keefe (Eds.), *Findings of the action research in ELICOS Program 2022* (pp. 15–42). Cambridge University Press & Assessment.
<https://www.cambridgeenglish.org/Images/690835-research-notes-85.pdf>
- Parkinson, J., Watterson, G., Sawczuc da Silva, A., & Kithulgoda, E. (2023). Rhetorical features of high-graded and low-graded engineering case study responses. *Journal of English for Academic Purposes*, 61, Article 101206.
<https://doi.org/10.1016/j.jeap.2022.101206>
- Patra, I., Alazemi, A., Al-Jamal, D., & Gheisari, A. (2022). The effectiveness of teachers' written and verbal corrective feedback (CF) during formative assessment (FA) on male language learners' academic anxiety (AA), academic performance (AP), and attitude toward learning (ATL). *Language Testing in Asia*, 12, Article 19.
<https://doi.org/10.1186/s40468-022-00169-2>
- Pérez Castillejo, S. (2023). Prior processing, foreign language classroom anxiety, and L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 61(2), 519-544. <https://doi.org/10.1515/iral-2021-0091>
- Perry, S. (2024). Podcasting as a form of assessment: Increasing student motivation in academic English-speaking assessment. *Language Testing in Asia*, 14, Article 15.
<https://doi.org/10.1186/s40468-024-00288-y>
- Phakiti, A., & Isaacs, T. (2021). Classroom assessment and validity: Psychometric and edumetric approaches. *The European Journal of Applied Linguistics and TEFL*, 10(1), 3–21.
- Piyumi Udeshinee, W. A., Knutsson, O., Barbutiu, S. M., & Jayathilake, C. (2024). Re-designing a regulatory scale for dynamic assessment in the synchronous text chat environment in collaboration with teachers. *Computer Assisted Language Learning*, 37(7), 1527-1553. <https://doi.org/10.1080/09588221.2022.2092153>
- Pomarolli, G., Artoni, D., Boschiero, M., & Piccinin, S. (2023). Inclusive and accessible teaching of Russian L2 to absolute beginners: A comparison between TBLT and PPP. *Instructed Second Language Acquisition*, 7(2), 166-193.
<https://doi.org/10.1558/isla.26047>
- Pourdana, N., & Asghari, S. (2021). Different dimensions of teacher and peer assessment of EFL learners' writing: Descriptive and narrative genres in focus. *Language Testing in Asia*, 11, Article 6. <https://doi.org/10.1186/s40468-021-00122-9>
- Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, 12, Article 7. <https://doi.org/10.1186/s40468-022-00156-7>
- Prasad, K., Aladini, A., Normurodova, N. Z., & Belton, B. (2023). Take language assessment easy: the mediator impacts of self-assessment, test-taking skills in predicting student evaluation apprehension, foreign language learning self-esteem, and language achievement in online classes. *Language Testing in Asia*, 13, Article 34 (2023). <https://doi.org/10.1186/s40468-023-00246-0>
- Purpura, J. E. (2024). Learning-oriented assessment. In A. J. Kunnan (Ed.), *Companion to Language Assessment* (pp. 100-124). Oxford, UK: John Wiley & Sons.
- Purpura, J.E (2021). A rationale for using a scenario-based assessment to measure competency-based, situated second and foreign language proficiency. In M.

- Masperi, C Cervini, Y Bardière (Eds.), *Évaluation des acquisitions langagières: du formatif au certificatif* (pp. 54-96). MediAzioni. <http://www.mediazioni.sitlec.unibo.it>.
- Qin, L., & Nian, Y. (2024). Unpacking Chinese lower-proficiency EFL learners' conceptual development of English subjunctives with group dynamic assessment: Efficacy, sustainability and reciprocity. *System*, 121, 103248. <https://doi.org/10.1016/j.system.2024.103248>
- Rahman, K. A., Hasan, M. K., Namaziandost, E., & Ibna Seraj, P. M. (2021). Implementing a formative assessment model at the secondary schools: Attitudes and challenges. *Language Testing in Asia*, 11, Article 18. <https://doi.org/10.1186/s40468-021-00136-3>
- Rastgou, A. (2022). How feedback conditions broaden or constrain knowledge and perceptions about improvement in L2 writing: A 12-week exploratory study. *Assessing Writing*, 53, Article 100633. <https://doi.org/10.1016/j.asw.2022.100633>
- Rezaei, A., & Barkaoui, K. (2021). Peer and teacher assessment of L2 writing in high- and low-stakes conditions. *International Journal of Applied Linguistics*, 172(2), 199–228. <https://doi.org/10.1075/itl.20006.rez>
- Rezaghah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Language Testing Journal*, 12(2). 76-94, <https://doi.org/10.22034/ijlt.2022.157126>
- Riswanto, Teferi, H., & Ahmed Abdel-Al Ibrahim, K. (2023). Cultivating EFL learners' productive skills by employing dynamic and non-dynamic assessments: Attitude in focus. *Language Testing in Asia*, 13, Article 18 (2023). <https://doi.org/10.1186/s40468-023-00228-2>
- Ryan, M., Khosronejad, M., Barton, G., Myhill, D., & Kervin, L. (2022). Reflexive writing dialogues: Elementary students' perceptions and performances as writers during classroom experiences. *Assessing Writing*, 51, Article 100592. <https://doi.org/10.1016/j.asw.2021.100592>
- Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: Pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *Journal of English for Academic Purposes*, 61, Article 101208. <https://doi.org/10.1016/j.jeap.2022.101208>
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50–61. <https://doi.org/10.1016/j.asw.2017.09.003>
- Sagre, A., Herazo, J. D., & Davin, K. J. (2022). Contradictions in teachers' classroom dynamic assessment implementation: An activity system analysis. *TESOL Quarterly*, 56(1), 154-177. <https://doi.org/10.1002/tesq.3046>
- Salamoura, A., & Morgan, S. (2021). Learning-oriented assessment from a teacher's perspective: Insights from teachers' action research. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp.182–206). Routledge. <https://doi.org/10.4324/9781003014102-13>
- Saville, N. (2021). Learning-oriented assessment: Basic concepts and frameworks in using assessment to support language learning. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 13–33). Routledge. <https://doi.org/10.4324/9781003014102-3>
- Sawaki, Y., Ishii, Y., & Oi, Y. (2024). The efficacy of automated feedback on summary content in an online formative assessment module for undergraduate L2 writing

- instruction in Japan. *Studies in Language Assessment (SILA)*, 13(2), 38–72.
<https://doi.org/10.58379/mjfp3942>
- Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, 14, Article 50. <https://doi.org/10.1186/s40468-024-00325-w>
- Schissel, J., Korne, H., & López-Gopar, M. (2021). Grappling with translanguaging for teaching and assessment in culturally and linguistically diverse contexts: Teacher perspectives from Oaxaca, Mexico. *International Journal of Bilingual Education and Bilingualism*, 24(3), 340–356. <https://doi.org/10.1080/13670050.2018.1463965>
- Seymour, J., & Angelo, D. (2023). Seeing the positives in assessment. Contributing to a “literature of doing” school-based Aboriginal language revival programs. *Studies in Language Assessment*, 12(2), 168–204.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4820543/SiLA_12_2_Seymour_and_Angelo.pdf
- Shen, B., & Bai, B. (2024). Enhancing Chinese university students’ writing performance and self-regulated learning (Srl) writing strategy use through a strategy-based intervention. *System*, 122, 103249. <https://doi.org/10.1016/j.system.2024.103249>
- Shrestha, P. N. (2022). Examining evaluative language used in assessment feedback on business students’ academic writing. *Assessing Writing*, 54, Article 100664.
<https://doi.org/10.1016/j.asw.2022.100664>
- Sippel, L., & Martin, I. A. (2024). Is corrective feedback during telecollaboration beneficial? The effects of peer and teacher corrections on L2 writing proficiency. *Journal of Second Language Writing*, 64, Article 101098.
<https://doi.org/10.1016/j.jslw.2024.101098>
- Skar, G. B., & Huebner, A. (2022). Predicting first grade students’ writing proficiency. *Assessment in Education: Principles, Policy & Practice*, 29(2), 219–237.
<https://doi.org/10.1080/0969594X.2022.2057424>
- Song, Y., Sparks, J. R., & Guzman-Orth, D. (2021). Investigating student performance on an assessment tool to support the development of English learners’ argumentation skills. *Computer Assisted Language Learning*, 35(9), 2291–2311.
<https://doi.org/10.1080/09588221.2021.1879160>
- Sun, Y., Lan, G., & Zhang, L. (2023). Pedagogical values of translingual practices in improving student feedback literacy in academic writing. *Assessing Writing*, 56, Article 100715. <https://doi.org/10.1016/j.asw.2023.100715>
- Sundqvist, P., & Uztosun, M. S. (2024). Extramural English in Scandinavia and Asia: scale development, learner engagement, and perceived speaking ability. *TESOL Quarterly*, 58(4), 1638–1665. <https://doi.org/10.1002/tesq.3296>
- Suzuki, Y., Eguchi, M., & de Jong, N. (2022). Does the reuse of constructions promote fluency development in task repetition? A usage-based perspective. *TESOL Quarterly*, 56(4), 1290–1319. <https://doi.org/10.1002/tesq.3103>
- Tavassoli, K., & Rahmatollahi, M. (2024). Differential impact of asynchronous and synchronous computer-assisted dynamic assessment on higher-order and lower-order writing skills. *Journal of Language Horizons*, 8(1), 33–61.
<https://doi.org/10.22051/lghor.2023.41130.1720>
- Tavassoli, K., Oskouiefar, M., & Ghamoushi, M. (2023). The effect of mobile-assisted learning-oriented assessment on EFL learners’ writing ability. *Iranian Journal of*

- Applied Linguistics, 26(2), 6-6. <http://ijal.khu.ac.ir/article-1-3233-en.html>.
- Tenca, E., Hartle, S., & Piccinin, S. (2023). Teaching oral skills in Higher Education (HE) language learning: Piloting a model for inclusive, accessible English Language Teaching (ELT). *Instructed Second Language Acquisition*, 7(2), 130-165. <https://doi.org/10.1558/isla.26019>
- Teng, L. S. (2022). Explicit strategy-based instruction in L2 writing contexts: A perspective of self-regulated learning and formative assessment. *Assessing Writing*, 53, Article 100645. <https://doi.org/10.1016/j.asw.2022.100645>
- Teng, M. F., & Zhang, L. J. (2024). Validating a new method for assessing young foreign language learners' metacognitive knowledge. *Research Methods in Applied Linguistics*, 3(2), Article 100121. <https://doi.org/10.1016/j.rmal.2024.100121>
- Teng, M. F., Wang, C., & Zhang, L. J. (2022). Assessing self-regulatory writing strategies and their predictive effects on young EFL learners' writing performance. *Assessing Writing*, 51, Article 100573. <https://doi.org/10.1016/j.asw.2021.100573>
- Toprak, T. E., & Cakir, A. (2021). Examining the L2 reading comprehension ability of adult ELLs: Developing a diagnostic test within the cognitive diagnostic assessment framework. *Language Testing*, 38(1), 106–131. <https://doi.org/10.1177/0265532220941470>
- Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238. <https://doi.org/10.1177/02655322211050600>
- Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2022). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Language Annals*, 55(1), 135–154. <https://doi.org/10.1111/flan.12571>
- Vakili, S., & Ebadi, S. (2022). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning*, 35(3), 345-380. <https://doi.org/10.1080/09588221.2019.1698616>
- Vasu, K. A., Mei Fung, Y., Nimehchisalem, V., & Md Rashid, S. (2022). Self-regulated learning development in undergraduate ESL writing classrooms: Teacher feedback versus self-assessment. *RELC Journal*, 53(3), 612–626. <https://doi.org/10.1177/0033688220957782>
- Vogt, K., & Antia, B. E. (2024). *Multilingual assessment: Finding the nexus?* Peter Lang. <https://doi.org/10.3726/b21992>
- Vogt, K., & Tzagari, D. (2024). Insights from emergency remote language assessment for a post-pandemic world. *Studies in Language Assessment (SILA)*, 13(2), 115–146. <https://doi.org/10.58379/jjzp4250>
- Wang, D. & East, M. (2024). Integrating translanguaging into assessment: students' responses and perceptions. *Applied Linguistics Review*, 15(5), 1911–1937. <https://doi.org/10.1515/applirev-2023-0087>
- Wang, D., Zhang, L., & Huang, Y. (2024). Tracking the development of logical metaphor usage in argumentative writing: A longitudinal study with EFL learners. *System*, 122, 103292. <https://doi.org/10.1016/j.system.2024.103292>
- Wang, I. K., & Cohen, A. D. (2022). Self-access strategy instruction for academic writing vocabulary: What learners actually do. *Applied Linguistics*, 44(6), 976–1009.

- <https://doi.org/10.1093/applin/amac056>
- Wang, T-H., Sun, Y., & Huang, N-W. (2023). Implementation of web-based dynamic assessment in improving low English achiever's learning effectiveness. *Computer Assisted Language Learning*, 36(8), 1450–1476.
<https://doi.org/10.1080/09588221.2021.1998129>
- Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, 53, Article 100641.
<https://doi.org/10.1016/j.asw.2022.100641>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, Article 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Watson, S. (2022). *Formative speaking assessment: Beginning with story* (Research Notes Issue 83). Cambridge University Press and Assessment.
<https://www.cambridgeenglish.org/Images/658686-research-notes-83.pdf>
- Wicaksono, B. H., Ismail, S. M., Sultanova, S. A., & Abeda, D. (2023). I like language assessment: EFL learners' voices about self-assessment, self-efficacy, grit tendencies, academic resilience, and academic demotivation in online instruction. *Language Testing in Asia*, 13, Article 37 (2023).
<https://doi.org/10.1186/s40468-023-00252-2>
- Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, 29(2), 183–199.
<https://doi.org/10.1080/0969594X.2022.2025762>
- Winke, P. M. (2023). Forty years of Language Testing, and the changing paths of publishing. *Language Testing*, 40(1), 3-7.
<https://doi.org/10.1177/02655322221136802>
- Wolf, M. K., & Lopez, A. (2022). Developing a technology-based classroom assessment of academic reading skills for English language learners and teachers: validity evidence for formative use. *Languages*, 7(2), Article 71.
<https://doi.org/10.3390/languages7020071>
- Wu, X. M., Zhang, L. J., & Dixon, H. R. (2021). Implementing Assessment for Learning (AfL) in Chinese university EFL classes: Teachers' values and practices. *System*, 101, Article 102589. <https://doi.org/10.1016/j.system.2021.102589>
- Xia, D., Chen, Y., & Pae, H. (2023). Lexical and grammatical collocations in beginning and intermediate L2 argumentative essays: a bigram study. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1421-1453.
<https://doi.org/10.1515/iral-2021-0188>
- Xie, Q., & Lei, Y. (2022). Diagnostic assessment of L2 academic writing product, process and self-regulatory strategy use with a comparative dimension. *Language Assessment Quarterly*, 19(3), 231-263.
<https://doi.org/10.1080/15434303.2021.1903470>
- Xu, J. (2022). Incremental intelligence matters: How L2 writing mindsets impact feedback orientation and self-regulated learning writing strategies. *Assessing Writing*, 51, Article 100593. <https://doi.org/10.1016/j.asw.2021.100593>
- Xu, J., & Qiu, X. (2024). Engaging L2 Learners in Information-gap Tasks: How Task Type and Topic Familiarity Affect Learner Engagement. *RELC Journal*, 55(1), 29–45.
<https://doi.org/10.1177/00336882211061628>

- Xu, Z., Zhang, L., & Parr, J. (2023). Incorporating peer feedback in writing instruction: examining its effects on Chinese English-as-a-foreign-language (EFL) learners' writing performance. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1337-1364. <https://doi.org/10.1515/iral-2021-0078>
- Yan, Q. (2024). Exploring Chinese university EFL students' perceptions of formative assessment: A qualitative study. *System*, 125, 103391. <https://doi.org/10.1016/j.system.2024.103391>
- Yan, Z., King, R. B., & Haw, J. Y. (2021). Formative assessment, growth mindset, and achievement: Examining their relations in the East and the West. *Assessment in Education: Principles, Policy & Practice*, 28(5-6), 676-702. <https://doi.org/10.1080/0969594X.2021.1988510>
- Yang, S. (2021). Diagnosing reading problems for low-level Chinese as second language learners. *System*, 97, Article 102433. <https://doi.org/10.1016/j.system.2020.102433>
- Yao, Y., Guo, N. S., Li, C., & McCampbell, D. (2021). How university EFL writers' beliefs in writing ability impact their perceptions of peer assessment: Perspectives from implicit theories of intelligence. *Assessment and Evaluation in Higher Education*, 46(1), 151-167. <https://doi.org/10.1080/02602938.2020.1750559>
- Yao, Y., Zhu, X., Zhu, S., & Jiang, Y. (2023). The impacts of self-efficacy on undergraduate students' perceived task value and task performance of L1 Chinese integrated writing: A mixed-method research. *Assessing Writing*, 55, Article 100687. <https://doi.org/10.1016/j.asw.2022.100687>
- Yi, Y.-S. (2021). On the usefulness of CDA-based score reporting: Implications for self-regulated learning. *Language Testing in Asia*, 11, Article 13. <https://doi.org/10.1186/s40468-021-00127-4>
- Yilmaz, Y., Arroyo, D., Carver, C., Choi, J., & Dibartolomeo, M. (2023). The effects of distributed versus massed corrective feedback on the acquisition of Spanish differential object marking. *Modern Language Journal*, 107, 756-781. <https://doi.org/10.1111/modl.12862>
- Yu, G. & Xu, J. (Eds.). (2024). Language test validation in a digital age. Studies in Language Testing Volume 52. Cambridge University Press & Assessment.
- Yu, G., & Green, A. (2021). Preparing for admissions tests in English. *Assessment in Education: Principles, Policy & Practice*, 28(1), 1-12. <https://doi.org/10.1080/0969594X.2021.1880120>
- Yu, S. (2021). Giving genre-based peer feedback in academic writing: sources of knowledge and skills, difficulties and challenges. *Assessment and Evaluation in Higher Education*, 46(1), 36-53. <https://doi.org/10.1080/02602938.2020.1742872>
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48, Article 100525. <https://doi.org/10.1016/j.asw.2021.100525>
- Yung, K. W.-H. (2023). Engaging exam-oriented students in communicative language teaching by 'packaging' learning English through songs as exam practice. *RELC Journal*, 54(1), 280-290. <https://doi.org/10.1177/0033688220978542>
- Zhan, Y., Lee, J. C. K., & Wan, Z. H. (2022). Who perceives more value of teacher feedback? Exploring the roles of college students' possible second language selves and language learning strategies. *Language Testing in Asia*, 12, Article 62. <https://doi.org/10.1186/s40468-022-00212-2>

- Zhang, Y., & MacWhinney, B. (2023). Using diagnostic feedback to enhance the development of phonetic knowledge of an L2: A CALL design based on the unified competition model and the implementation with the Pinyin Tutor. *Language Testing in Asia*, 13, Article 35 (2023). <https://doi.org/10.1186/s40468-023-00232-6>
- Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, Article 100586. <https://doi.org/10.1016/j.asw.2021.100586>
- Zhao, C. G., & Qi, Q. (2023). Implementing learning-oriented assessment (LOA) among limited-proficiency EFL students: Challenges, strategies, and students' reactions. *TESOL Quarterly*, 57(2), 566–594. <https://doi.org/10.1002/tesq.3167>
- Zhao, H., & Zhao, B. (2023). Co-constructing the assessment criteria for EFL writing by instructors and students: A participative approach to constructively aligning the CEFR, curricula, teaching and learning. *Language Teaching Research*, 27(3), 765–793. <https://doi.org/10.1177/1362168820948458>
- Zhou, Q., Chen, L., & Hou, X. (2022). Exploring Chinese EFL undergraduates' writing from sources: Self-efficacy and performance. *Assessing Writing*, 54, Article 100663. <https://doi.org/10.1016/j.asw.2022.100663>
- Zimmermann, S., Schwarz, L., Peters, A. & Depner, G. (2024). Enhancing teachers and test takers' assessment literacy? Insights from test preparation for the digital TestDaF. In: B. Baker & L. Taylor (Eds.), *Language Assessment Literacy and Competence Volume 1: Research and Reflections from the Field* (pp. 110–128). Cambridge: Cambridge University Press.

Fairness and Ethics

- Antia, B. (2021). Multilingual examinations: Towards a schema of politicization of language in end of high school examinations in sub-Saharan Africa. *International Journal of Bilingual Education and Bilingualism*, 24(1), 138–153. <https://doi.org/10.1080/13670050.2018.1450354>
- Arce, A. J., & Young, M. J. (2022). Test efficacy: Refocusing validation from college exams to candidates. *International Journal of Testing*, 22(1), 100–119. <https://doi.org/10.1080/15305058.2021.2019752>
- Azizi, Z. (2022). Fairness in assessment practices in online education: Iranian University English teachers' perceptions. *Language Testing in Asia*, 12, Article 14. <https://doi.org/10.1186/s40468-022-00164-7>
- Azizi, Z., & Ismail, S. M. (2023). What ethical requirements should be considered in language classroom assessment? Insights from high school students. *Language Testing in Asia*, 13, Article 23. <https://doi.org/10.1186/s40468-023-00235-3>
- Bazvand, A. D. (2023). Exploring student teachers' perceptions of assessment ethics across university-based teacher education programs in Iran. *Language Testing in Asia*, 13, Article 9. <https://doi.org/10.1186/s40468-022-00205-1>
- Carlsen, C. H., & Rocca, L. (2021). Language test misuse. *Language Assessment Quarterly*, 18(5), 477–491. <https://doi.org/10.1080/15434303.2021.1947288>
- Chapman, M., Kemp, J. A., Kim, A. A., MacGregor, D., & MacMillan, F. Testing young multilingual learners remotely: Prioritizing equity, fairness, and social justice. In K. Sadeghi & D. Douglas (Eds.), *Fundamental considerations in technology mediated*

- language assessment (pp. 153-170). Routledge.
<https://doi.org/10.4324/9781003292395-13>
- Chen, X., Aryadoust, V., & Zhang, W. (2024). A systematic review of differential item functioning in second language assessment. *Language Testing*, 0(0).
<https://doi.org/10.1177/02655322241290188>
- Cheng, L. & Sultana, N. (2022). Washback: Looking backward and forward. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp 136–152).
<https://doi.org/10.4324/9781003220756-12>
- Cheng, L., & Zheng, Y. (2021). Measuring anxiety. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 177–186). <https://doi.org/10.4324/9781351034784-20>
- Cheng, L., Im, G-H., & Jabeen, R. (2021). Whose English(es). are we assessing and by whom? In B. Lantaigne, C. Coombe., & J.D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp. 169–179). Springer.
https://doi.org/10.1007/978-981-33-4232-3_12
- Christensen, L. L., Shyyan, V. V., & MacMillan, F. (2023). Toward a systematic accessibility review process for English language proficiency tests for young learners. *Language Testing*, 40(4), 856-876. <https://doi.org/10.1177/02655322231168386>
- Crosby, S. (2023). A fair go: Translanguaging and assessment practices in a New Zealand junior college. *Studies in Language Assessment*, 12(2), 117-150.
- Dalal, D. K., Randall, J. G., Cheung, H. K., Gorman, B. C., Roch, S. G., & Williams, K. J. (2022). Is there bias in alternatives to standardized tests? An investigation into letters of recommendation. *International Journal of Testing*, 22(1), 21-42.
<https://doi.org/10.1080/15305058.2021.2019751>
- East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–273.
<https://doi.org/10.1017/s0261444823000034>
- Erguvan, I. D. (2021). The rise of contract cheating during the COVID-19 pandemic: A qualitative study through the eyes of academics in Kuwait. *Language Testing in Asia*, 11, Article 34. <https://doi.org/10.1186/s40468-021-00149-y>
- Erlam, R., & Rau, W. T. (2023). Fairer for all: Teachers identifying and solving problems with the assessment of English language learners in the mainstream classroom. *Studies in Language Assessment*, 12(2), 94-116.
- Fan, J. (2022). Ethics and fairness. In K. Luo (Ed.), *Major Issues in Language Assessment* (pp. 207-220). Foreign Language Testing and Research Press.
- Fox, J., & Artemeva, N. (2022). *Reconsidering context in language assessment: Transdisciplinary perspectives, social theories, and validity*. Routledge.
<https://doi.org/10.4324/9781351184571>
- Franken, M. (2023). The new national literacy tests for post-primary students in Aotearoa New Zealand: How process and design issues undermine principles of a strong and fair qualification. *Studies in Language Assessment*, 12(2), 59-93.
- Frost, K. (2021). Negotiating the boundaries of responsibility: Rethinking test takers and the ethics of testing. *Papers in Language Testing and Assessment*, 10(1), 70–83.
http://www.altanz.org/uploads/5/9/0/8/5908292/5._frost.pdf
- Furtoso, V. B., Egido, A. A., & Moraes, I. (2022). Social justice in language assessment: Intercultural communicative competence in the Celpe-Bras exam. In E. Meletiadiou (Ed.), *Handbook of research on fostering social justice through intercultural and*

- multilingual communication (pp. 162-186). IGI Global.
<https://doi.org/10.4018/978-1-6684-5083-3.ch009>
- Grapin, S. E. (2023). Assessment of English learners and their peers in the content areas: Expanding what “Counts” as evidence of content learning. *Language assessment quarterly*, 20(2), 215–234. <https://doi.org/10.1080/15434303.2022.2147072>
- Guapacha Chamorro, M. E. (2022). Cognitive validity evidence of computer- and paper-based writing tests and differences in the impact on EFL test-takers in classroom assessment. *Assessing Writing*, 51, Article 100594.
<https://doi.org/10.1016/j.asw.2021.100594>
- Hosseinnia, M., & Kafi, Z. (2024). Constructing and validating a Code of Ethics in Testing Inventory: Investigating EFL instructors’ perspectives. *Language Testing in Asia*, 14, Article 56. <https://doi.org/10.1186/s40468-024-00331-y>
- Hudson, C., Angelo, D., & Creagh, S. (2023). Instantiating justice, fairness and inclusiveness in English as an Additional Language/Dialect assessment frameworks: Unpacking the evidence base for the Bandscales State Schools (Queensland). *Studies in Language Assessment*, 12(2), 235–271.
- Inbar-Lourie, O., & Shohamy, E. (2021). The Conflict and Consequences of Two Assessment Measures in Israel: Global PISA vs. the National MEITZAV. In B. Lantaigne (Ed.). *Challenges in Language Testing around the World* (pp. 191-200). Springer International Publishing
- Isaacs, T., & Winke, P. M. (2024). Purposeful turns for more equitable and transparent publishing in language testing and assessment. *Language Testing*, 41(1), 3-8.
doi:10.1177/02655322231203234
- Isbell, D. R., & Kim, J. (2023). Developer involvement and COI disclosure in high-stakes English proficiency test validation research: A systematic review. *Research Methods in Applied Linguistics*, 2(3), Article 100060.
<https://doi.org/10.1016/j.rmal.2023.100060>
- Isbell, D. R., Kremmel, B., & Kim, J. (2023). Remote Proctoring in Language Testing: Implications for Fairness and Justice. *Language Assessment Quarterly*, 20(4-5), 469–487. <https://doi.org/10.1080/15434303.2023.2288251>
- Jalilzadeh, K., Rashtchi, M., & Mirzapour, F. (2024). Cheating in online assessment: A qualitative study on reasons and coping strategies focusing on EFL teachers’ perceptions. *Language Testing in Asia*, 14, Article 29.
<https://doi.org/10.1186/s40468-024-00304-1>
- Khan, K. (2022). The securitisation of language borders and the (re) production of inequalities. *TESOL Quarterly*, 56(4), 1458-1470. <https://doi.org/10.1002/tesq.3186>
- Klieger, D., Bochenek, J., Ezzo, C., Holtzman, S., Cline, F., & Olivera-Aguilar, M. (2022). Using third-party evaluations to assess socioemotional skills in graduate and professional school admissions. *International Journal of Testing*, 22(1), 72-99.
<https://doi.org/10.1080/15305058.2021.2019748>
- Lane, P. (2023). From silence to silencing? Contradictions and tensions in language R=revitalization. *Applied Linguistics*, 44(5), 833–847.
<https://doi.org/10.1093/applin/amac075>
- López-Gopar, M. E. (2021). Citizenship in language testing: A call for respectful collaborations. *Language Testing*, 38(4), 649–653.
<https://doi.org/10.1177/02655322211022061>

- Mäkipää, T., Hahl, K., & Luodonpää-Manni, M. (2021). Teachers' perceptions of assessment and feedback practices in Finland's foreign language classes during the Covid-19 pandemic. *CEPS Journal*, 11, 219–240. <https://doi.org/10.26529/cepsj.1108>
- McElwee, S., Y.F. Cheung, K., R.T. Cromie, S., Shannon, M., & Gallacher, T. (2021). Revising the BioMedical Admissions Test (BMAT) to improve impact and washback for candidates and support fair access to test preparation. *Assessment in Education: Principles, Policy & Practice*, 28(1), 77–95. <https://doi.org/10.1080/0969594X.2019.1639133>
- Moradi, E., Ghabanchi, Z., & Pishghadam, R. (2022). Reading comprehension test fairness across gender and mode of learning: Insights from IRT-based differential item functioning analysis. *Language Testing in Asia*, 12, Article 39. <https://doi.org/10.1186/s40468-022-00192-3>
- Newman, D. A., Tang, C., Song, Q. C., & Wee, S. (2022). Dropping the GRE, keeping the GRE, or GRE-optional admissions? Considering tradeoffs and fairness. *International Journal of Testing*, 22(1), 43-71. <https://doi.org/10.1080/15305058.2021.2019750>
- Niessen, A. S. M., & Neumann, M. (2022). Using personal statements in college admissions: An investigation of gender bias and the effects of increased structure. *International Journal of Testing*, 22(1), 5-20. <https://doi.org/10.1080/15305058.2021.2019749>
- Nieto, D., & Nguyen, A. (2023). Strategizing assessment to support emerging bilingual students: a focus on the home language. *Multicultural Learning and Teaching*, 18(2), 239-250. <https://doi.org/10.1515/mlt-2021-0005>
- Noori, M. (2022). Road not taken" in language testing: Sociocultural implications of test and teaching contents. *TESOL Quarterly*, 56(4), 1486-1503. <https://doi.org/10.1002/tesq.3189>
- Oliver, R., Chen, H., & Dovchin, S. (2024). Review of selected research in applied linguistics published in Australia (2015–2022). *Language Teaching*, 57(3), 341–376. doi:10.1017/S0261444824000077
- Pelleriti, M. (2023). Item writing and dyslexic students: Promoting best practices. *Iperstoria*, 21, 217–231. <https://iperstoria.it/article/view/1285/1315>
- Prabowo, M. Y. (2022). Assessing equivalency between paper-and-pencil and computer-based English competency test. In R. E. Capel & Q. S. Hoon (Eds.), Conference proceedings iELT-CON2021 (pp. 42–51). Penang English Language Learning and Teaching Association. https://pellta.org/uploads/8/1/8/9/81899300/iel-con_2021_proceedings.pdf
- QUEVEDO-CAMARGO, G., & PINHEIRO, L. A. (2021) Ética na avaliação de línguas adicionais: da postura docente ao instrumento avaliativo [Ethics in the assessment of additional languages: from the teaching stance to the assessment instrument]. *Estudos em Avaliação Educacional* (Online), v. 32. DOI: <https://doi.org/10.18222/eae.v32.8641>
- Rahmati, T., & Nushi, M. (2023). EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography. *Iranian Journal of Language Teaching Research*, 11(3), 31-52. <https://doi.org/10.30466/ijltr.2023.121405>
- Randez, R. A., & Cornell, C. (2023). Advancing equity in language assessment for learners with disabilities. *Language Testing*, 40(4), 984-999. <https://doi.org/10.1177/02655322231169442>

- Rasooli, A., DeLuca, C. & Cheng, L. (2022). Beginning teacher candidates' approaches to grading and assessment conceptions—Implications for teacher education in assessment. *Educational Research for Policy and Practice*, 22, 63–90. <https://doi.org/10.1007/s10671-022-09320-5>
- Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2024). Classroom Assessment Fairness Inventory: A New Instrument to Support Fairness in Classroom Assessment. *Assessment in Education*, 30-5-6. <https://doi.org/10.1080/0969594X.2023.2255936>
- Rezai, A. (2022). Fairness in classroom assessment: development and validation of a questionnaire. *Language Testing in Asia*, 12, Article 17. <https://doi.org/10.1186/s40468-022-00162-9>
- Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: insights from Iranian university teachers. *Language Testing in Asia*, 12, Article 8. <https://doi.org/10.1186/s40468-022-00157-6>
- Rolstad, K., & MacSwan, J. (2024). Bilingual language assessment: A persistent case of bias. In A. K. Kibler, A. Walqui, G. C. Bunch, & C. J. Faltis (Eds.), *Equity in multilingual schools and communities: Celebrating the contributions of Guadalupe Valdés* (pp. 54-66). Multilingual Matters. <https://doi.org/10.2307/lj.22452326.9>
- Sadeghi, K. & Douglas, D. (Eds.). (2023). *Fundamental considerations in technology mediated language assessment*. Routledge. <https://doi.org/10.4324/9781003292395>
- Saito, H., Sawaki, Y., & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and use for Japanese university admission. *Language Assessment Quarterly*, 19(4), 422-448. <https://doi.org/10.1080/15434303.2022.2083965>
- Salaberry, M. R. & Weideman, A. (2024). Context, construct and ethics. In M. R., Salaberry, A. Weideman, A. & W-L. Hsu. (Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 3-27). Routledge. https://doi.org/10.4324/9781003384922_-2
- Schissel, J. L., & Khan, K. (2021). Responsibilities and opportunities in language testing with respect to historicized forms of socio-political discrimination: A matter of academic citizenship. *Language Testing*, 38(4), 640–648. <https://doi.org/10.1177/02655322211028590>
- Schmidgall, J., Huo, Y., Cid, J., & Wei, Y. (2024). Investigating fairness claims for a general-purposes assessment of English proficiency for the international workplace: Do full-time employees have an unfair advantage over full-time students? ETS Research Report Series, 2024(1), 1–20. <https://doi.org/10.1002/ets2.12380>
- Shahmirzadi, N. (2024). Simulation study: Seeking fairness in a high-stakes placement test. *Language Testing in Focus: An International Journal*. 10, 18–28, (e-ISSN: 2717-9087). <https://doi.org/10.32038/ltf.2024.10.02>
- Shahmirzadi, N., Siyyari, M., Marashi, H., & Gramipour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA. *Journal of Foreign Languages Research*, 10(1), 152–165, (ISC), (ISSN: 2588–4123, Online ISSN: 2588–7521). DOI: 10.22059/JFLR.2019.282328.634
- Shamsaldeen, F., Wang, J., & Ahn, S. (2024). Evaluating the fairness of a high-stakes college entrance exam in Kuwait. *Language Testing in Asia*, 14, Article 27. <https://doi.org/10.1186/s40468-024-00301-4>
- Shohamy, E. (2022). Critical language testing, multilingualism and social justice. *TESOL Quarterly*, 56(4), 1445-1457. <https://doi.org/10.1002/tesq.3185>

- Suto, I., Williamson, J., Ireland, J., & Macinska, S. (2023). On reducing errors in assessment instruments. *Research Papers in Education*, 38(3), 357-377. <https://doi.org/10.1080/02671522.2021.1968940>
- Taylor, L. (2023). Reframing the discourse and rhetoric of language testing and assessment for the public square. *Language Testing*, 40(1), 47-53. <https://doi.org/10.1177/02655322221127421>
- Taylor, L., & Banerjee, J. (2023). Accommodations in language testing and assessment: Safeguarding equity, access, and inclusion. *Language Testing*, 40(4), 847-855. <https://doi.org/10.1177/02655322231186221>
- Taylor, L., & Banerjee, J. (2023). Language assessment accommodations: Issues and challenges for the future. *Language Testing*, 40(4), 1000-1006. <https://doi.org/10.1177/02655322231186222>
- Uysal, H. (2022). Standardized tests and within-group segregation: The not-so-optimal ESL classroom for long-term English learners. *TESOL Quarterly*, 56(4), 1471-1485. <https://doi.org/10.1002/tesq.3187>
- Vidal Rodeiro, C., & Macinska, S. (2022). Equal opportunity or unfair advantage? The impact of test accommodations on performance in high-stakes assessments. *Assessment in Education : Principles, Policy & Practice*, 29(4), 462-481. <https://doi.org/10.1080/0969594X.2022.2121680>
- Vogt, K., & Antia, B. E. (2024). *Multilingual assessment: Finding the nexus?* Peter Lang. <https://doi.org/10.3726/b21992>
- Weideman, A., & Deygers, B. (2024). Validity and validation: An alternative perspective. In M. R. Salaberry, A. Weideman, & W-L. Hsu.(Eds.). *Ethics and context in second language testing: Rethinking validity in theory and practice* (pp. 28-50). Routledge. <https://doi.org/10.4324/9781003384922-3>
- Wolf, M. K., Yoo, H., Guzman-Orth, D., & Abedi, J. (2022). Investigating the effects of test accommodations with process data for English learners in a mathematics assessment. *Educational Assessment*, 27(1), 27-45. <https://doi.org/10.1080/10627197.2021.1982693>
- Yang, Z., & Yu, C.* (2024). Is the use of multiple-choice items and a holistically-scored paragraph translation task fair? Examining a large-scale translation subtest. *Asia Pacific Education Review*. Online First. <https://doi.org/10.1007/s12564-024-09993-y>

Language Testing Policy

- Arias, A., Vessey, R., & Sheyholislami, J. (2022). Jumping through hoops in the Canadian immigration system: A critical view of the immigrant's journey to citizenship. *TESOL Quarterly*, 56(4), 1430-1444. <https://doi.org/10.1002/tesq.3184>
- Barkaoui, K. (2024). The academic achievement of undergraduate students with different score profiles on the TOEFL iBT: A replication study. *Language Assessment Quarterly*. <https://doi.org/10.1080/15434303.2024.2438142>
- Bruce, E., Dunn, K., & Clark, T. (2023). Review of recommended two-year limit on IELTS validity: A summary of preliminary internal research. *Research Reports*. <https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/review-of-recommended-2-year-limit-on-ielts-validity-bruce-et-al-2023.pdf>
- Carlsen, C. H., & Rocca, L. (2022). Language test activism. *Language Policy*, 21(4), 597-616. <https://doi.org/10.1007/s10993-022-09614-7>

- Cook, B., Luke, J., Valoe, A., & Barkaoui, K. (2021). Institutional language policy and ESL teachers' conceptions of L2 writing and classroom assessment practices. *Canadian Modern Language Review*, 77(2), 93–109. <https://doi.org/10.3138/cmlr-2020-0013>
- Coombe, T., Clark, T., & Mohebbi, H. (2023). Volume Introduction: Special Issue in Honour of James Dean Brown's Five-Decade Contribution to Language Testing and Assessment. *Language Teaching Research Quarterly*, 37, 1-3.
- Cushing, S. T., Ren, H., & Tan, Y. (2024). The use of TOEFL iBT® in admissions decisions: Stakeholder perceptions of policies and practices. ETS Research Report Series, 2024(1), 1–16. <https://doi.org/10.1002/ets2.12375>
- De Jong, J. H.A.L. (2022) The Action-oriented approach and language testing: A critical view. In D. Little & N. Figueras (Eds.), *Reflecting on the Common European Framework of reference for languages and its companion volume* (pp. 53–63). *Multilingual Matters*. <https://doi.org/10.21832/9781800410206-009>
- Deygers, B. (2021). The CEFR companion volume: Between research-based policy and policy-based research. *Applied Linguistics*, 42(1), 186–191. <https://doi.org/10.1093/applin/amz024>
- Deygers, B., Bigelow, M., Lo Bianco, J., Nadarajan, D., & Tani, M. (2021). Low print literacy and its representation in research and policy. *Language Assessment Quarterly*, 18(5), 463–476. <https://doi.org/10.1080/15434303.2021.1903471>
- Dissake, E. M. K. (2022). Assessing litigant's language proficiency: the case of the Bafoussam Court of First Instance. *Language Policy*, 21(2), 217–234. <https://doi.org/10.1007/s10993-021-09603->
- Du Plessis, T. (2021). Institutional language policy and academic literacy in South African higher education. In A. Weideman, J. Read & T. Du Plessis (Eds.), *Assessing academic literacy in a multilingual society: Transition and transformation* (pp. 3-21). *Multilingual Matters*.
- Elder, C. (2021). Towards 'policy responsible' language assessment: Framing the Language Testing Research Centre's contribution in languages other than English. *Papers in Language Testing and Assessment*, 10(1), 4–29. http://www.altanz.org/uploads/5/9/0/8/5908292/2._elder.pdf
- Erguvan, I. D. (2022). University students' understanding of contract cheating: A qualitative case study in Kuwait. *Language Testing in Asia*, 12, Article 56. <https://doi.org/10.1186/s40468-022-00208-y>
- Erickson, G., & Tholin, J. (2022). Overall, a good test, but...—Swedish lower secondary teachers' perceptions and use of national test results of English. *Languages*, 7(1), Article 64. <https://doi.org/10.3390/languages7010064>
- Erickson, G., Borger, L., & Olsson, E. (2022). National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language Testing*, 39(3), 474–493. <https://doi.org/10.1177/02655322221075067>
- Fan, J., Frost, K., & Jin, Y. (2022). Local English testing in China's tertiary education: Contexts, policies, and practices. *Language Testing*, 39(3), 453–473. <https://doi.org/10.1177/02655322211070839>
- Fjørtoft, H., Lai, M. K., & Li, M. (2024). How curriculum and assessment policies affect the role of reading in an assessment culture: a New Zealand case study. *Assessment in Education: Principles, Policy & Practice*, 31(2), 1-31. <https://doi.org/10.1080/0969594X.2024.2365645>

- Frost, K. (2021). Negotiating the boundaries of responsibility: Rethinking test takers and the ethics of testing. *Papers in Language Testing and Assessment*, 10(1), 70–83.
http://www.altanz.org/uploads/5/9/0/8/5908292/5._frost.pdf
- Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56. <https://doi.org/10.32038/ltrq.2022.29.03>
- Gebril, A. (2021). Nation-wide assessment reform in Egypt: Challenges and potentials of learning-oriented assessment in an examinations-based instructional setting. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theory into practice* (pp. 109–122). Routledge. <https://doi.org/10.4324/9781003014102-9>
- Green, A. and Chan, S. (2022). Report to the Nursing and Midwifery Council on language test policy. *Nursing and Midwifery Council*, 1–27.
<https://www.nmc.org.uk/globalassets/sitedocuments/english-language-consultation/crella-report-to-the-nmc-june-22.pdf>
- Inbar-Lourie, O., & Shohamy, E. (2021). The Conflict and Consequences of Two Assessment Measures in Israel: Global PISA vs. the National MEITZAV. In B. Lantaigne (Ed.), *Challenges in Language Testing around the World* (pp. 191-200). Springer International Publishing
- Islam, M. S., Hasan, M. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: Principles, practices, and problems. *Language Testing in Asia*, 11, Article 1. <https://doi.org/10.1186/s40468-020-00116-z>
- Jin, Y. (2023). Decision making in large-scale language testing: Intersections of policy, practice and research. *Studies in Language Assessment*, 12(1), 64–92.
<https://doi.org/10.58379/HKKX5020>
- Kazemian, M. & Khonamri, F. (2022). Past, present, and future of language assessment: An interview with Dr. Hossein Farhady, *MEXTESOL Journal*, 46(4), 1–8.
https://www.mextesol.net/journal/index.php?page=journal&id_article=46395
- Khan, K. (2022). The securitisation of language borders and the (re) production of inequalities. *TESOL Quarterly*, 56(4), 1458-1470. <https://doi.org/10.1002/tesq.3186>
- Kılıçkaya, F. (2021). Reconsidering language assessment courses in teacher training programs during and after the COVID-19 pandemic. *Lublin Studies in Modern Languages and Literature*, 45(3), 45–55.
<https://doi.org/10.17951/lsmll.2021.45.3.45-55>
- Knoch, Ute. (2021). The challenges of providing expert advice in policy contexts. *Papers in Language Testing and Assessment*, 10(1), 30–48.
http://www.altanz.org/uploads/5/9/0/8/5908292/3._knoch.pdf
- Lam, D. M. K., Green, A., Murray, N., & Gayton, A. (2021). How are IELTS scores set and used for university admissions selection: A cross-institutional case study. *IELTS Research Reports Online Series*, 3. British Council, Cambridge Assessment English and IDP: IELTS Australia.
<https://www.ielts.org/for-researchers/research-reports/online-series-2021-3>
- Leung, C., Evans, M., & Liu, Y. (2021). English as an additional language assessment framework: Filling a void in policy and provision in school education in England. *Language Assessment Quarterly*, 18(3), 296–315.
<https://doi.org/10.1080/15434303.2020.1869745>
- Macqueen, S, Pill, J., & Knoch, U. (2021). Trust the test: Score-user perspectives on the roles of language tests in professional registration and skilled migration. *Papers in*

- Language Testing and Assessment*, 10(1), 49–69.
http://www.altanz.org/uploads/5/9/0/8/5908292/4._macqueen_pill___knoch.pdf
- Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103, Article 102674.
<https://doi.org/10.1016/j.system.2021.102674>
- McLeod, M., & Cheng, L. (2023). The Canadian English Language Proficiency Index Program (CELPPI) Test. *Language Assessment Quarterly*, 20(3), 339–352.
<https://doi.org/10.1080/15434303.2023.2237487>
- Nazari, M., & Molana, K. (2023). “Predators of emotions”: The role of school assessment policies in English language teachers' emotion labor. *TESOL Quarterly*, 57(4), 1226–1255. <https://doi.org/10.1002/tesq.3188>
- O'Sullivan, B., & Cheng, L. (2022). Lessons from the Chinese imperial examination system. *Language Testing in Asia*, 12, Article 52.
<https://doi.org/10.1186/s40468-022-00201-5>
- Oliver, R., Chen, H., & Dovchin, S. (2024). Review of selected research in applied linguistics published in Australia (2015–2022). *Language Teaching*, 57(3), 341–376.
[doi:10.1017/S0261444824000077](https://doi.org/10.1017/S0261444824000077)
- Park, H.I., Solon, M., Dehghan-Chaleshtori, M., & Ghanbar, H. (2022). Proficiency reporting practices in research on second language acquisition: Have we made any progress?. *Language Learning*, 72, 198–236. <https://doi.org/10.1111/lang.12475>
- Prabowo, M. Y., & Rahmadian, S. (2022). Computer-based English competency assessment for scholarship selection: Challenges, strategies, and implementation in the Ministry of Finance. *Jurnal Sositelknologi*, 21(1), 84–96.
<https://doi.org/10.5614/sostek.itbj.2022.21.1.9>
- Qin, C. Y., & Baker, B. (2025). Review of the Canadian English Language Proficiency Index Program (CELPPI). *Language Testing*, 42(1), 100–113.
<https://doi.org/10.1177/02655322241291764>
- Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education: Principles, Policy & Practice*, 29(3), 383–394.
<https://doi.org/10.1080/0969594X.2022.2103516>
- Rios, J. A., & Ihlenfeldt, S. D. (2021). State assessment score reporting practices for English learner parents. *Educational Measurement: Issues and Practice*, 40(3), 31–41. <https://doi.org/10.1111/emip.12424>
- Russell, B., Barron, C., Kim, H., & Jang, E. E. (2022). A mixed-method investigation into international university students' experience with academic language demands. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2022.934692>
- Saito, H., Sawaki, Y., & Kasahara, K. (2022). Evaluating fairness and justice of external English language test score interpretation and use for Japanese university admission. *Language Assessment Quarterly*, 19(4), 422–448.
<https://doi.org/10.1080/15434303.2022.2083965>
- Schildt, L., Deygers, B., & Weideman, A. (2024). Language testers and their place in the policy web. *Language Testing*, 41(2), 338–356.
<https://doi.org/10.1177/02655322231191133>
- Seymour, J., & Angelo, D. (2023). Seeing the positives in assessment. Contributing to a “literature of doing” school-based Aboriginal language revival programs. *Studies in Language Assessment*, 12(2), 168–204.

- https://arts.unimelb.edu.au/__data/assets/pdf_file/0008/4820543/SiLA_12_2_Seymour_and_Angelo.pdf
- Shin, D. (2024). Critical discursive approaches to evaluating policy-driven testing: Social impact as a target for validation. *Language Testing*, 41(1), 162–180.
<https://doi.org/10.1177/02655322231163863>
- Shin, D., Park, S., & Cho, E. (2023). A review study on discourse-analytical approaches to language testing policy in the South Korean context. *Language Testing in Asia*, 13, Article 44. <https://doi.org/10.1186/s40468-023-00257-x>
- Sims, M. L. (2023). Shifting perceptions of socially just writing assessment: Labor-based contract grading and multilingual writing instruction. *Assessing Writing*, 57, Article 100731. <https://doi.org/10.1016/j.asw.2023.100731>
- Spiby, R., & Clark, T. (2023). Responding to a transformed assessment landscape: The example of IELTS. In H. I. Işık, E. Kantarcıoğlu, A. Üner & C. Thomas (Eds.), *Re-Vision and Re-Form in English for academic purposes contexts after the pandemic* (pp. 83-93). Newcastle: Cambridge Scholars Publishing.
- Taylor, L., & Banerjee, J. (2023). Accommodations in language testing and assessment: Safeguarding equity, access, and inclusion. *Language Testing*, 40(4), 847-855.
<https://doi.org/10.1177/02655322231186221>
- Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing*, 39(2), 212–238.
<https://doi.org/10.1177/02655322211050600>
- Wudthayagorn, J. (2022). An exploration of the English Exit Examination policy in Thai public universities. *Language Assessment Quarterly*, 19(2), 107-123.
<https://doi.org/10.1080/15434303.2021.1937174>
- Yan, X., Chuang, P.-L., & Zhang, C. (2024). Economic return of English language proficiency: Do English proficiency scores predict employee income in China? *Studies in Language Assessment*, 13(1), 108-140.
<https://doi.org/10.58379/DLDI3148>
- Yang, R., & May, L. (2023). Understanding and Implementing Learning-Oriented Assessment: Evidence From Three Chinese Junior High School EFL Teachers. *Chinese Journal of Applied Linguistics*, 46(2), 308-325.
<https://doi.org/10.1515/CJAL-2023-0210>
- Yuksel, D., Soruç, A., Altay, M., & Curle, S. (2023). A longitudinal study at an English medium instruction university in Turkey: the interplay between English language improvement and academic success. *Applied Linguistics Review*, 14(3), 533-552.
<https://doi.org/10.1515/applirev-2020-0097>
- Zhao, R., Coss, M. D., Ruan, H., Li, B., & Ma, J. (2023). Examining secondary–postsecondary articulation of Chinese language programs: A survey of US college placement procedures. *Foreign Language Annals*, 56(3), 690–719.

Research in Language Assessment

Research Methodology

- Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2023). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition*, 46(2), 307–329. <https://doi.org/10.1017/S0272263123000487>
- Alpizar, D., Li, T., Norris, J. M., & Gu, L. (2023). Psychometric approaches to analyzing C-tests. *Language Testing*, 40(1), 107-132. <https://doi.org/10.1177/02655322211062138>
- Aryadoust, V., Foo, S., & Ng, L. Y. (2022). What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Language Testing*, 39(1), 56–89. <https://doi.org/10.1177/02655322211026876>
- Aryadoust, V., Min, S., & Chen, X. (2024). Investigating differential item functioning across interaction variables in listening comprehension assessment. *Studies in Educational Evaluation*, 80, 101322. <https://doi.org/10.1016/j.stueduc.2024.101322>
- Badham, L., & Furlong, A. (2023). Summative assessments in a multilingual context: What comparative judgment reveals about comparability across different languages in Literature. *International Journal of Testing*, 23:2, 111-134. <https://doi.org/10.1080/15305058.2022.2149536>
- Barkaoui, K. (2021). *Evaluating tests of second language development*. Peter Lang. <https://doi.org/10.3726/b18059>
- Barkaoui, K. (2024). Multi-faceted Rasch analysis for evaluating second language tests. In Kunnan, A. (Ed.), *The concise companion to language assessment* (pp. 602-618). Wiley-Blackwell.
- Barkaoui, K. (2024). The academic achievement of undergraduate students with different score profiles on the TOEFL iBT: A replication study. *Language Assessment Quarterly*. <https://doi.org/10.1080/15434303.2024.2438142>
- Barkaoui, K., & Hadidi, A. (2021). *Assessing changes in second Language writing performance*. Routledge. <https://doi.org/10.4324/9781003092346>
- Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 38(4), 511–535. <https://doi.org/10.1177/0265532220951504>
- Betts, J., Muntean, W., Kim, D., & Kao, S. (2022). Evaluating different scoring methods for multiple response items providing partial credit. *Educational and Psychological Measurement*, 82(1), 151–176. <https://doi.org/10.1177/0013164421994636>
- Bonvin, A., Brugger, L., & Berthele, R. (2023). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*, 61(2), 257-285. <https://doi.org/10.1515/iral-2020-0093>
- Buffington, J., Demos, A., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, 43(3), 635–662. <https://doi.org/10.1017/S0272263121000127>
- Bulut, H. C., Bulut, O., & Arikan, S. (2023). Evaluating group differences in online reading comprehension: The impact of item properties. *International Journal of Testing*, 23(1), 10-33. <https://doi.org/10.1080/15305058.2022.2044821>

- Chan, K. K. Y., Bond, T., & Yan, Z. (2023). Application of an Automated Essay Scoring engine to English writing assessment using Many-Facet Rasch Measurement. *Language Testing*, 40(1), 61-85. <https://doi.org/10.1177/02655322221076025>
- Clark, T., & Endres, H. (2021). Online diagnostic grammar testing for high school students—An international trial. *Assessment in Education: Principles, Policy & Practice*, 28(5–6), 602–632. <https://doi.org/10.1080/0969594X.2021.1970513>
- Curtis, A. & Cheng, L. (2024). The Complexities of Diversity: A Personal Journey Across Time and Space. *The Language Teacher*, 48(5), 14-17
<https://doi.org/10.37546/JALTTLT48.5-3>
- DeKeyser, R., & Li, S. (2021). Epilogue: Evidence for the validity of implicit aptitude and the need for construct validation. *Studies in Second Language Acquisition*, 43(3), 692–697. <https://doi.org/10.1017/S0272263121000140>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE.
- Dimitrov, D. M., & Atanasov, D. V. (2022). Testing for differential item functioning under the D-Scoring method. *Educational and Psychological Measurement*, 82(1), 107–121. <https://doi.org/10.1177/00131644211001524>
- Dobrić, N., Cesnik, H., Harsch, C. (2023). *Festschrift in honour of Günther Sigott: Advanced methods in language testing*. Peter Lang. <https://doi.org/10.3726/b20407>
- Drackert, A., Zimmermann, S., Marks, D. & Korflür, D. (2023). New test format – new research agenda: An overview of the technology-related research at g.a.s.t. In: *Conference Proceedings of the ALTE 8th International Conference: Language Assessment Fit for the Future* (pp. 34–38).
- Engelhard, G., & Wang, J. (2021). *Rasch models for solving measurement problems: Invariant measurement in the social sciences*. SAGE.
- Falk, C. F., & Feuerstahler, L. M. (2022). On the performance of semi- and nonparametric item response functions in computer adaptive tests. *Educational and Psychological Measurement*, 82(1), 57–75. <https://doi.org/10.1177/00131644211014261>
- Fan, J., & Zhang, X. (2024). Applying the Rasch model in language assessment. Foreign Language Teaching and Research Press.
- Fox, J. (2021). *A mathematical primer for social statistics* (2nd ed.). SAGE.
- Furr, R. M. (2021). *Psychometrics: An introduction* (4th ed.). SAGE.
- Graham, K. M. (2024). Examining the criterion validity of curriculum-based measurement of writing (CBM-W) with young EFL learners. *Research Methods in Applied Linguistics*, 3(2), Article 100108. <https://doi.org/10.1016/j.rmal.2024.100108>
- Gyllstad, H., McLean, S., & Stewart, J. (2021). Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. *Language Testing*, 38(4), 558–579. <https://doi.org/10.1177/0265532220979562>
- Hair Jr, J.F., Hult, G.T., Ringle, C.M., Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. (3rd ed.). SAGE.
- In'nami, Y., & Koizumi, R. (2021). *Factor structure and four-skill profiles of the Aptis test*. (ARAGs research reports, AR-G/2021/2). British Council.
<https://www.britishcouncil.org/factor-structure-and-four-skill-profiles-aptis-test>
- In'nami, Y., & Koizumi, R. (2022). The relationship between L2 listening and metacognitive awareness across listening tests and learner samples. *International Journal of*

- Listening*, 36(2), 100-117. <http://dx.doi.org/10.1080/10904018.2021.1955683>
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2023). Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach. *Language Learning*, 73(3), 759–798. <https://doi.org/10.1111/lang.12548>
- In'nami, Y., Mizumoto, A., Plonsky, L., & Koizumi, R. (2022). Promoting computationally reproducible research in applied linguistics: Recommended practices and considerations. *Research Methods in Applied Linguistics*, 1(3), Article 100030. <https://doi.org/10.1016/j.rmal.2022.100030>
- Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 305–315). Routledge. <https://doi.org/10.4324/9781351034784-33>
- Isbell, D. R., & Son, Y. A. (2022). Measurement properties of a standardized elicited imitation test: An integrative data analysis. *Studies in Second Language Acquisition*, 44(3), 859-885. <https://doi.org/10.1017/S0272263121000383>
- Isbell, D. R., & Son, Y.-A. (2023). Explanatory item response models for instrument validation: A tutorial based on an elicited imitation test. *Research Methods in Applied Linguistics*, 2(3), Article 100080. <https://doi.org/10.1016/j.rmal.2023.100080>
- Jeon, E. H., In'nami, Y., & Koizumi, R. (2022). Discussion, limitations, and future research. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 369-385). John Benjamins. <https://doi.org/10.1075/bpa.13>
- Jin, K.-Y., & Eckes, T. (2022). Detecting Differential Rater Functioning in Severity and Centrality: The dual DRF facets model. *Educational and Psychological Measurement*, 82(4), 757–781. <https://doi.org/10.1177/00131644211043207>
- Kahng, J., & Otonya, M. (2021). What elicited imitation can show us: Quantitative and qualitative analyses of longitudinal changes. *TESOL Quarterly*, 55(1), 284–295. <https://doi.org/10.1002/tesq.3020>
- Kang, H.-A., Han, S., Kim, D., & Kao, S.-C. (2022). Polytomous testlet response models for technology-enhanced innovative items: implications on model fit and trait inference. *Educational and Psychological Measurement*, 82(4), 811–838. <https://doi.org/10.1177/00131644211032261>
- Kollias, C. (2023). *Virtual standard setting: Setting cut scores*. Peter Lang. <https://doi.org/10.3726/b20407>
- Kroc, E., & Olvera Astivia, O. L. (2022). The importance of thinking multivariately when setting subscale cutoff scores. *Educational and Psychological Measurement*, 82(3), 517–538. <https://doi.org/10.1177/00131644211023569>
- Lamprianou, I., Tzagari, D., & Kyriakou, N. (2023). Experienced but detached from reality: Theorizing and operationalizing the relationship between experience and rater effects. *Assessing Writing*, 56, Article 100713. <https://doi.org/10.1016/j.asw.2023.100713>
- Latimer, N., & Chan, S. (2022). Eye-tracking L2 students taking online multiple-choice reading tests: Benefits and challenges. *International Journal of TESOL Studies*, 4(1), 83–104. <https://doi.org/10.46451/ijts.2022.01.07>
- Lee, C., & Qian, H. (2022). Hybrid threshold-based sequential procedures for detecting compromised items in a computerized adaptive testing licensure exam. *Educational*

- and *Psychological Measurement*, 82(4), 782–810.
<https://doi.org/10.1177/00131644211023868>
- Li, H., Hunter, C. V., & Bialo, J. A. (2022). A revisit of Zumbo's third generation DIF: how are we doing in language testing?. *Language Assessment Quarterly*, 19(1), 27-53.
<https://doi.org/10.1080/15434303.2021.1988953><https://doi.org/10.1080/15434303.2021.1963253>
- Li, S. (2024). Using the Rasch model to evaluate instrument and data quality in assessing requests in L2 Chinese. In S. Li (Ed.), *Pragmatics of Chinese as a second language* (pp. 161-184). Multilingual Matters. <https://doi.org/10.2307/jj.22679704.12>
- Li, S., Wen, T., Li, X., Feng, Y., & Lin, C. (2023). Comparing holistic and analytic marking methods in assessing speech act production in L2 Chinese. *Language Testing*, 40(2), 249–275. <https://doi.org/10.1177/02655322221113917>
- Liao, R. J. (2023). The use of generalizability theory in investigating the score dependability of classroom-based L2 reading assessment. *Language Testing*, 40(1), 86-106.
<https://doi.org/10.1177/02655322211070840>
- Liu, M., Al-Hoorie, A. H., & Hiver, P. (2024). Open Access in language testing and assessment: The case of two flagship journals. *Language Testing*, 41(4), 703–728.
<https://doi.org/10.1177/02655322241261417>
- Liu, X., & Jane Rogers, H. (2022). Treatments of differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, 82(2), 225–253.
<https://doi.org/10.1177/00131644211012050>
- Liu, X., & Read, J. (2021). Investigating the skills involved in reading test tasks through expert judgement and verbal protocol analysis: Convergence and divergence between the two methods. *Language Assessment Quarterly*, 18(4), 357–381.
<https://doi.org/10.1080/15434303.2021.1881964>
- Lundgren, E., & Eklöf, H. (2023). Questionnaire-taking motivation: Using response times to assess motivation to optimize on the PISA 2018 student questionnaire. *International Journal of Testing*, 23:4, 231-256. <https://doi.org/10.1080/15305058.2023.2214647>
- Luo, Y. (2021). A comparison of common IRT model-selection methods with mixed-format tests. *Measurement: Interdisciplinary Research and Perspectives*, 19(4), 199–212.
<https://doi.org/10.1080/15366367.2021.1878779>
- McKay, T. H., Teimouri, Y., Sağdıç, A., Salen, B., Reagan, D., & Malone, M. E. (2021). The cagey C-test construct: Some evidence from a meta-analysis of correlation coefficients. *System*, 99, Article 102526.
<https://doi.org/10.1016/j.system.2021.102526>
- Min, S., Cai, H., & He, L. (2022). Application of bi-factor MIRT and higher-order CDM models to an in-house EFL listening test for diagnostic purposes. *Language Assessment Quarterly*, 19(2), 189-213.
<https://doi.org/10.1080/15434303.2021.1980571>
- Monfils, L. F., & Manna, V. F. (2021). Time to achieving a designated criterion score level: A survival analysis study of test taker performance on the TOEFL iBT® test. *Language Testing*, 38(1), 154–176. <https://doi.org/10.1177/0265532220940709>
- Nicklin, C., & Vitta, J. P. (2022). Assessing Rasch measurement estimation methods across R packages with yes/no vocabulary test data. *Language Testing*, 39(4), 513–540.
<https://doi.org/10.1177/02655322211066822>
- Paek, I., Liang, X., & Lin, Z. (2021). Regarding item parameter invariance for the Rasch and the 2-parameter logistic models: An investigation under finite non-representative

- sample calibrations. *Measurement: Interdisciplinary Research and Perspectives*, 19(1), 39–54. <https://doi.org/10.1080/15366367.2020.1754703>
- Paquot, M., Rubin, R., & Vandeweerd, N. (2022). Crowdsourced adaptive comparative judgment: A community-based solution for proficiency rating. *Language Learning*, 72, 853–885. <https://doi.org/10.1111/lang.12498>
- Rahmati, T., & Nushi, M. (2023). EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography. *Iranian Journal of Language Teaching Research*, 11(3), 31–52. <https://doi.org/10.30466/ijltr.2023.121405>
- Rasooli, A., DeLuca, C., Cheng, L., & Mousavi, A. (2024). Classroom Assessment Fairness Inventory: A New Instrument to Support Fairness in Classroom Assessment. *Assessment in Education*, 30–5–6. <https://doi.org/10.1080/0969594X.2023.2255936>
- Raykov, T., & Marcoulides, G.A. (2023) Evaluating the discrepancy between scale reliability and Cronbach's coefficient alpha using latent variable modeling, *Measurement: Interdisciplinary Research and Perspectives*, 21(1), 29–37, DOI: 10.1080/15366367.2022.2031485
- Sarac, M., & Loken, E. (2023). Examining patterns of omitted responses in a large-scale English language proficiency test. *International Journal of Testing*, 23(1), 56–72. <https://doi.org/10.1080/15305058.2022.2070756>
- Shin, J., & Gierl, M. J. (2021). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 38(2), 247–272. <https://doi.org/10.1177/0265532220937830>
- Shin, J., & Gierl, M. J. (2022). Generating reading comprehension items using automated processes. *International Journal of Testing*, 22(3–4), 289–311. <https://doi.org/10.1080/15305058.2022.2070755>
- Skehan, P., Bui, G., Wang, Z., & Shum, S. (2024). Re-examining accuracy measures in second language task-based spoken performance. *Research Methods in Applied Linguistics*, 3(1), Article 100098. <https://doi.org/10.1016/j.rmal.2024.100098>
- Stoeckel, T., Ha, H. T., Nguyen, D. T. B., & Nicklin, C. (2024). Can an extended-matching second-language vocabulary test format bridge the gap between meaning-recognition and meaning-recall?. *Research Methods in Applied Linguistics*, 3(2), Article 100109. <https://doi.org/10.1016/j.rmal.2024.100109>
- Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, 43(3), 663–676. <https://doi.org/10.1017/S0272263120000704>
- Suzuki, Y., & Koizumi, R. (2021). Using equivalent test forms in SLA pretest-posttest design research. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 457–467). Routledge. <https://doi.org/10.4324/9781351034784-49>
- Tavassoli, K., & Sorat, Z. (2023). Iranian EFL teachers' oral/aural skills language assessment literacy: Instrument development and validation. *International Journal of Language Testing*, 13(2), 56–76. <https://doi.org/10.22034/IJLT.2023.376466.1219>
- Teng, M. F., & Teng, L. S. (2024). Validating the multi-dimensional structure of self-efficacy beliefs in peer feedback for L2 writing: A bifactor-exploratory structural equation modeling approach. *Research Methods in Applied Linguistics*, 3(3), Article 100136. <https://doi.org/10.1016/j.rmal.2024.100136>

- Tokunaga, H. (2021). *Moving from IBM SPSS to R and RStudio: A statistics companion*. SAGE.
- Torregrossa, J., Listanti, A., Bongartz, C., & Narinis T. (2024). Adding discourse to sentence repetition tasks: Under which conditions does bilingual children's performance improve?. *Research Methods in Applied Linguistics*, 3(2), Article 100107. <https://doi.org/10.1016/j.rmal.2024.100107>
- Turner, C. E. (2023). Assessment and evaluation: mixed methods research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). Wiley Online Library. <https://doi.org/10.1002/9781405198431.wbeal0040.pub2>
- Wind, S. A. (2022). *Exploring rating scale functioning for survey research*. SAGE.
- Winke, P., & Brunfaut, T. (2021). Perspectives on “knowing” a second language: What are we seeking to measure? In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp.1–8). Routledge. <https://doi.org/10.4324/9781351034784-1>
- Winke, P., & Brunfaut, T. (Eds.). (2021). *The Routledge handbook of second language acquisition and language testing*. Routledge. <https://doi.org/10.4324/9781351034784>
- Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2023). A multidimensional approach to assessing the effects of task complexity on L2 students' argumentative writing. *Assessing Writing*, 55, Article 100690. <https://doi.org/10.1016/j.asw.2022.100690>
- Yan, X., & Chuang, P.-L. (2023). How do raters learn to rate? Many-facet Rasch modeling of rater performance over the course of a rater certification program. *Language Testing*, 40(1), 153-179. <https://doi.org/10.1177/02655322221074913>
- Yan, X., & Fan, J. (2022). Reliability and dependability. In Fulcher, G., & Harding, L. (Eds.), *The Routledge Handbook of Language Testing* (pp. 477-494). Routledge.
- Zhang, M., & Sinharay, S. (2022). Investigating the writing performance of educationally at-risk examinees using technology. *International Journal of Testing*, 22(3-4), 312-347. <https://doi.org/10.1080/15305058.2022.2050734>
- Zhang, W., Zhao, M., & Zhu, Y. (2022). Understanding individual differences in metacognitive strategy use, task demand, and performance in integrated L2 speaking assessment tasks. *Frontiers in Psychology*, 13, Article 876208. <https://doi.org/10.3389/fpsyg.2022.876208>
- Zhang, X., & Li, W. (2021). Effects of n-grams on the rated L2 writing quality of expository essays: A conceptual replication and extension. *System*, 97, Article 102437. <https://doi.org/10.1016/j.system.2020.102437>

Research Synthesis

- Agrawal, S., Oza, P., Kakkar, R., Tanwar, S., Jetani, V., Undhad, J., & Singh, A. (2024). Analysis and recommendation system-based on PRISMA checklist to write systematic review. *Assessing Writing*, 61, Article 100866. <https://doi.org/10.1016/j.asw.2024.100866>
- Allen, D., & Tahara, T. (2021). A review of washback research in Japan. *Japanese Language Testing Association Journal*, 24, 3–22. https://doi.org/10.20622/jltajournal.24.0_3
- Aryadoust, V., & Luo, L. (2023). The typology of second language listening constructs: A

- systematic review. *Language Testing*, 40(2), 375–409.
<https://doi.org/10.1177/02655322221126604>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6–40.
<https://doi.org/10.1177/0265532220927487>
- Binti Hasrol, S., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), Article 100023. <https://doi.org/10.1016/j.rmal.2022.100023>
- Bonvin, A., Brugger, L., & Berthele, R. (2023). Lexical measures as a proxy for bilingual language dominance? *International Review of Applied Linguistics in Language Teaching*, 61(2), 257–285. <https://doi.org/10.1515/iral-2020-0093>
- Brown, J. D., Panahi, A., & Mohebbi, H. (2023). James Dean Brown's 50 years of work in language testing: A systematic review. *Language Teaching Research Quarterly*, 37, 4–75. <https://doi.org/10.32038/ltrq.2023.37.02>
- Bruce, E., Dunn, K., & Clark, T. (2023). Review of recommended two-year limit on IELTS validity: A summary of preliminary internal research. *Research Reports*.
<https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/review-of-recommended-2-year-limit-on-ielts-validity-bruce-et-al-2023.pdf>
- Chen, X., Aryadoust, V., & Zhang, W. (2024). A systematic review of differential item functioning in second language assessment. *Language Testing*, 0(0).
<https://doi.org/10.1177/02655322241290188>
- Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 37(1-2), 179–221. <https://doi.org/10.1080/09588221.2022.2033787>
- Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A systematic review. *Language Teaching Research Quarterly*, 29, 20–56. <https://doi.org/10.32038/ltrq.2022.29.03>
- Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503–525.
<https://doi.org/10.1080/15434303.2022.2128802>
- Guo, J. & Cheng, L. (2024). Chinese Parent Involvement in Young Learners' English Tests: A Systematic Review of the Literature. *語言測試和評估 (Language Testing and Assessment)*, 4, 1–20.
- Han, C. (2022). Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, 39(1), 30–55. <https://doi.org/10.1177/02655322211036100>
- Han, C. (2023). Interrogating the predictive validity of aptitude testing for interpreting: A systematic methodological review. *The Interpreter and Translator Trainer*, 17 (1), 7–28. <https://doi.org/10.1080/1750399X.2023.2170049>
- Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3), 100023.
- Isbell, D. R., & Kim, J. (2023). Developer involvement and COI disclosure in high-stakes English proficiency test validation research: A systematic review. *Research Methods in Applied Linguistics*, 2(3), Article 100060.
<https://doi.org/10.1016/j.rmal.2023.100060>

- Kim, M., Qiu, X., & Wang Y. (2024). Interrater agreement in genre analysis: A methodological review and a comparison of three measures, *Research Methods in Applied Linguistics*, 3(1), Article 100097. <https://doi.org/10.1016/j.rmal.2024.100097>
- Kirkpatrick, R., Vafadar, H., & Mohebbi, H. (2024). A scoping review of willingness to communicate in language education: premises to doubt, lessons to learn, and future research questions to examine. *Language Testing in Asia*, 14, 9. <https://doi.org/10.1186/s40468-024-00284-2>
- Martin, C., & Dockrell, J. E. (2024). Writing productivity development in elementary school: A systematic review. *Assessing Writing*, 60, 100834. <https://doi.org/10.1016/j.asw.2024.100834>
- Mei, H., & Chen, H. (2024). Cognitive Diagnosis in Language Assessment: A Thematic Review. *RELC Journal*, 55(2), 529–537. <https://doi.org/10.1177/00336882221122357>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68, Article 100963. <https://doi.org/10.1016/j.stueduc.2020.100963>
- Mootoosamy, K., & Aryadoust, V. (2024). Transitioning from communicative competence to multimodal and intercultural competencies: A systematic review. *Societies*, 14(115). <https://doi.org/10.3390/soc14070115>
- Nishizawa, H., Isbell, D. R., & Suzuki, Y. (2022). Review of the Japanese-language proficiency test. *Language Testing*, 39(3), 494–503. <https://doi.org/10.1177/02655322221080898>
- Oliver, R., Chen, H., & Dovchin, S. (2024). Review of selected research in applied linguistics published in Australia (2015–2022). *Language Teaching*, 57(3), 341–376. doi:10.1017/S0261444824000077
- Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly*, 39, 372–425. <https://doi.org/10.32038/ltrq.2024.39.23>
- Panahi, A., Mohebbi, H., & Oxford, R. L. (2024). Rebecca Oxford's 50 years of contributions to language education and related fields: A systematic review. *Language Teaching Research Quarterly*, 41, 169–230.
- Rahmati, T., & Sadeghi, K. (2024). Technological innovations used in researching language assessment: A scoping review. In K. Sadeghi (Ed.), *Routledge handbook of technological advances in researching language learning* (pp. 118-129). <https://doi.org/10.4324/9781003459088-11>
- Shi, G., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*. Advance online publication. <https://doi.org/10.1017/S0958344023000265>
- Shi, H., & Aryadoust, V. (2023). A systematic review of automated writing evaluation within the argument-based validation framework. *Education and Information Technology*, 28, 771–795. <https://link.springer.com/article/10.1007/s10639-022-11200-7>
- Shin, D., Park, S., & Cho, E. (2023). A review study on discourse-analytical approaches to language testing policy in the South Korean context. *Language Testing in Asia*, 13, Article 44. <https://doi.org/10.1186/s40468-023-00257-x>

- Weisleder, A., Friend, M., Tsui, A. S. M., & Marchman, V. A. (2024). Using parent report to measure vocabulary in young bilingual children: A scoping review. *Language Learning*, 74(2), 468-505. <https://doi.org/10.1111/lang.12617>
- Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. *Educational Science*, 12(10), Article 723. <https://doi.org/10.3390/educsci12100723>

Open Science and Academic Publishing

- Chapelle, C. A., & Ockey, G. J. (2024). Open Science in language assessment research contexts: A reply to Winke. *Language Testing*, 41(4), 882–885. <https://doi.org/10.1177/02655322241239377>
- Clark, T., & Bruce, E. (2024). Open Science should be welcomed by test providers but grounded in pragmatic caution: A response to Winke. *Language Testing*, 41(4), 872–876. <https://doi.org/10.1177/02655322231223105>
- Gebril, A., & Bali, M. (2024). A Global South perspective on Open Science in language assessment: A response to Winke. *Language Testing*, 41(4), 886–891. <https://doi.org/10.1177/02655322241260121>
- Isaacs, T., & Winke, P. M. (2024). Purposeful turns for more equitable and transparent publishing in language testing and assessment. *Language Testing*, 41(1), 3-8. [doi:10.1177/02655322231203234](https://doi.org/10.1177/02655322231203234)
- Jin, Y., & Fan, J. (2024). Open Science for language assessment research and practice in China: A response to Winke. *Language Testing*, 41(4), 877–881. <https://doi.org/10.1177/02655322231223100>
- Koizumi, R., Maie, R., Yanagisawa, A., & In'nami, Y. (2024). Considerations to promote and accelerate Open Science: A response to Winke. *Language Testing*, 41(4), 892–897. <https://doi.org/10.1177/02655322241239379>
- Papageorgiou, S. (2024). Can language test providers do more to support Open Science? A response to Winke. *Language Testing*, 41(4), 860–864. <https://doi.org/10.1177/02655322241232361>

Other

Reflections and Perspectives on the Field

- Baker, B. & Gebril, A. (2024). Editorial: The assessment of writing in languages other than English. *Assessing Writing*, 60, 1-2. <https://doi.org/10.1016/j.asw.2024.100840>.
- Barkaoui, K., Shi, L., & Riazi, M. (2021). “Follow your interests because those will motivate you to excel”: An interview with Alister Cumming. *Language Assessment Quarterly*, 18(1), 56–75. <https://doi.org/10.1080/15434303.2020.1846190>
- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Chapelle, C. A., & Ockey, G. J. (2024). Open Science in language assessment research contexts: A reply to Winke. *Language Testing*, 41(4), 882–885. <https://doi.org/10.1177/02655322241239377>
- Clark, T., & Bruce, E. (2024). Open Science should be welcomed by test providers but grounded in pragmatic caution: A response to Winke. *Language Testing*, 41(4), 872–876. <https://doi.org/10.1177/02655322231223105>

- Gebril, A., & Bali, M. (2024). A Global South perspective on Open Science in language assessment: A response to Winke. *Language Testing*, 41(4), 886–891.
<https://doi.org/10.1177/02655322241260121>
- Jin, Y., & Fan, J. (2024). Open Science for language assessment research and practice in China: A response to Winke. *Language Testing*, 41(4), 877–881.
<https://doi.org/10.1177/02655322231223100>
- Kazemian, M. & Khonamri, F. (2022). Past, present, and future of language assessment: An interview with Dr. Hossein Farhady, *MEXTESOL Journal*, 46(4), 1–8.
https://www.mextesol.net/journal/index.php?page=journal&id_article=46395
- Kim, P., Beltrán, J., & Joo, S. H. (2022). An interview with APPLE Lecture speaker Dr. Paul D. Deane. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University*, 22(1), 47–55. <https://doi.org/10.52214/salt.v22i1.9901>
- Koizumi, R., Maie, R., Yanagisawa, A., & In'nami, Y. (2024). Considerations to promote and accelerate Open Science: A response to Winke. *Language Testing*, 41(4), 892–897.
<https://doi.org/10.1177/02655322241239379>
- Kwon, O., & Lee, W. (2021). A life for English language education: An interview with Oryang Kwon. *Language Assessment Quarterly*, 18(2), 188–198.
<https://doi.org/10.1080/15434303.2020.1859512>
- Liao, R. J. T. (2021). Advancing the international recognition of the locally-produced GEPT: An interview with Jessica Wu. *Language Assessment Quarterly*, 18(4), 446–455.
<https://doi.org/10.1080/15434303.2021.1919116>
- Mohebbi, H. (2021). 25 years on, the correction debate continues: An interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), Article 3. <https://doi.org/10.1186/s40862-021-00110-9>
- Papageorgiou, S. (2024). Can language test providers do more to support Open Science? A response to Winke. *Language Testing*, 41(4), 860–864.
<https://doi.org/10.1177/02655322241232361>
- Purpura, J. E. (2024). An Interview with Timothy Francis McNamara. *Language Assessment Quarterly*, 21(2). 123-124.
- Purpura, J. E. (2024). In Memoriam: Timothy Francis McNamara. *Language Assessment Quarterly*. 21(2). 125-140.
- Purpura, J. E., & Kunnan, A. J. (Eds.) (2024). *The writings of L. F. Bachman: "Assuring that what we count counts."* New York: Routledge.

Bibliography of Material and Datasets

As a new section in the bibliography, we encourage authors to submit reference entries to their open material and datasets. Datasets predating 2021 are welcome and will be added on a yearly basis.

- Burton, J. D. (2021). *Question complexity, nonunderstanding, and nonverbal behavior* (OSF F2JXK, Version V1) [Data set]. Open Science Framework. <https://doi.org/10.17605/OSF.IO/F2JXK>
- Burton, J.D. (2020). *Raters' perceptions and operationalization of authentic engagement in oral proficiency tests* (OSF TSDH9, Version V1) [Data set]. Open Science Framework. <https://doi.org/10.17605/OSF.IO/TSDH9>
- Burton, J. D. (2023). The impact of nonverbal behavior on ratings of second language proficiency (OSF SK726, Version V1) [Data set]. Open Science Framework. <https://osf.io/sk726/>
- Brunfaut, T., & Kormos, J. (2024, August 7). Assessing multimodal viewing-to-write constructs: Task design, performance, processing, and rating. Retrieved from osf.io/xude7
- Wilkens, R., Pintard, A., Alfter, D., Folny, V., & François, T. (2023). TCFLE-8: a Corpus of Learner Written Productions for French as a Foreign Language and its Application to Automated Essay Scoring
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2022). *Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach* [Data set and materials]. Open Science Framework. <https://osf.io/xvnf2/>
- In'nami, Y., Cheung, M. W.-L., Koizumi, R., & Wallace, M. P. (2023). *Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach* [Materials]. *Language Learning*, 73(3), 759–798. <https://doi.org/10.1111/lang.12548>
- In'nami, Y., Hijikata, Y., & Koizumi, R. (2021). Working memory capacity and L2 reading: A meta-analysis [Data set and codes]. Open Science Framework. <https://osf.io/x34h8/>
- In'nami, Y., Hijikata, Y., & Koizumi, R. (2022). *Working memory capacity and L2 reading: A meta-analysis* [Materials]. *Studies in Second Language Acquisition*, 44(2), 381–406. <https://doi.org/10.1017/S02722263121000267>
- In'nami, Y., & Koizumi, R. (2022). *The relationship between L2 listening and metacognitive awareness across listening tests and learner samples* [R codes]. Open Science Framework. <https://osf.io/hnvr8/>
- In'nami, Y., & Koizumi, R. (2022). *The relationship between L2 listening and metacognitive awareness across listening tests and learner samples* [Materials]. *International Journal of Listening*, 36(2), 100–117. <http://dx.doi.org/10.1080/10904018.2021.1955683>
- In'nami, Y., Koizumi, R., Kojima, M., & Jeon, E. H. (2023). *Understanding L2 proficiency: Theoretical and meta-analytic investigations* [Data set]. Open Science Framework. <https://osf.io/edzqa/>
- In'nami, Y., Mizumoto, A., Plonsky, L., & Koizumi, R. (2022). *Promoting computationally reproducible research in applied linguistics: Recommended practices and considerations* [Data set and codes]. Open Science Framework. <https://osf.io/6jkms/>
- In'nami, Y., Mizumoto, A., Plonsky, L., & Koizumi, R. (2022). *Promoting computationally reproducible research in applied linguistics: Recommended practices and considerations* [Materials]. *Research Methods in Applied Linguistics*, 1(3), 100030, 1–9. <https://doi.org/10.1016/j.rmal.2022.100030>
- Jeon, E. H., & In'nami, Y. (Eds.). (2022). *Understanding L2 proficiency: Theoretical and meta-analytic investigations* [Data set]. John Benjamins. <https://doi.org/10.1075/bpa.13>
- Koizumi, R., & In'nami, Y. (2022). *Assessing functional adequacy using picture description tasks in classroom-based L2 speaking assessment* [Data set]. Open Science Framework. <https://osf.io/ubwnr/>

- Segura, M., & Knight, J. (2024). Creating a school-child-family evaluation rubric for evaluating children's plurilingual and intercultural competence in a Catalan primary school (v.1) [Data set]. Open Science Framework. <https://doi.org/10.17605/OSF.IO/WE9C2>
- Winke, P., & Zhang, X. (2022). *Data and codebook for SSLA article: "A closer look at a marginalized test method: Self-assessment as a measure of speaking proficiency."* (Version V1) [Data set] Inter-university Consortium for Political and Social Research (ICPSR). <https://doi.org/10.3886/E164981V1>