

COLLOQUIUM ON THE VALIDATION OF ORAL PROFICIENCY TESTS

TITLES OF PAPERS

1. Bachman, L. Testing oral production in the individualized language learning project, Thailand.
2. Canale, M. French as a second language: Ontario assessment instrument pool project.
3. Clark, J. Convergent-divergent test validation at Educational Testing Service.
4. Clifford, R. The "what" and "why" of convergent-divergent validation.
5. Davies, A. To be announced.
6. Groot, P. On the need for construct validation of oral proficiency tests.
7. Hinofotis, F., Bailey, K, and Stern, S. The development of a performance test of oral proficiency for foreign teaching assistants.
8. Lowe, P. Structure of the oral interview and content validity.
9. Madsen, H. and Jones, R. Classification of oral proficiency tests.
10. Engelskirchen, Cottrell, and Oller. The reliability and validity of the Ilyin Oral Interview.
11. Palmer, A. Measurements of reliability and concurrent validity of two picture description tests of oral communication.
12. Hendricks, Scholz, Spurling, Johnson, and Vandenberg. Three pragmatic tests of oral proficiency and the FSI Oral Interview: an evaluation.
13. Shohamy, E. The concurrent validity of the oral interview with the cloze procedure in Hebrew as a second language.
14. Stevenson, D. Beyond faith and face validity: the multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests.
15. Upstur, J. A sociolinguistic analysis of Albert Camus' "The Myth of Sisyphus": implications for oral proficiency testing.

LTRC 1
1979

REPORT ON THE TWO-DAY COLLOQUIUM ON THE CONSTRUCT
VALIDATION OF ORAL PROFICIENCY TESTS

Adrian S. Palmer
English Department
University of Utah
Salt Lake City, Utah 84112

For the second year in a row, a small group of researchers in language testing have met for two days during the TESOL national convention. The first colloquium was held in Boston during TESOL '79 and was chaired by Peter J.M. Groot (Institute of Applied Linguistics, University of Utrecht, Wilhelminapark 11, Utrecht, Holland) and Adrian Palmer. It was organized to bring researchers together to discuss not only the problems of testing oral proficiency but also test validation procedures in general and construct validation procedures in particular. One of the outcomes of this colloquium was the development of a design (using a model proposed by Campbell and Fiske) for a pilot multitrait-multimethod convergent-divergent construct validation study of tests of the traits "communicative competence in speaking" and "communicative competence in reading." Another outcome was the preparation of a volume of papers on the validation of oral proficiency tests (Palmer and Groot, editors, forthcoming). A third outcome was the decision to continue to meet to consider the results of the pilot study as well as other advances in oral testing.

The second colloquium is the one reported on here. Organized by Adrian Palmer and Lyle F. Bachman (Division of ESL, 3070 Foreign Languages Bldg., University of Illinois, Urbana, Ill. 61801), the colloquium met March 4-5, 1980 at the TESOL national convention in San Francisco. Papers were prepared and circulated in advance to the 20 participants. The colloquium itself was devoted to brief summaries of the papers followed by extensive discussion. The final working session was devoted to planning coordinated follow-up studies.

The Papers

Lyle Bachman and Adrian Palmer presented a paper entitled "The Construct Validation of the Constructs 'Communicative Competence in Speaking' and 'Communicative Competence in Reading': A Pilot Study." Competence in speaking and reading were tested by three methods: interview, translation, and self-ratings. The subjects were 75 native speakers of Mandarin Chinese at the University of Illinois. The researchers found strong evidence for convergent validity and weak evidence for discriminant validity of tests of speaking and reading using correlational analysis and analysis of variance. Subsequent to the colloquium, they used confirmatory factor analysis and found that a two trait model (speaking and reading) accounted for the data significantly better than a one trait (unitary language factor) model.

Jack Upshur was asked prior to the colloquium to critique the pilot (Bachman-Palmer) study. Because Upshur was not directly involved in planning the pilot study, it was felt that he could provide an unbiased evaluation of the design, implementation, and conclusions. In his paper "Critique of Bachman-Palmer Pilot Study on Communicative Competence in Speaking and Reading," Upshur pointed out that the number of traits (2) and methods (3) used resulted in an underidentified model. This, in conjunction with the subject selection procedure, which may have involved choosing subjects who had been affected by a highly similar set of variables, may have made discriminant validity very difficult to demonstrate. He expressed reservations about reaching conclusions about convergent and discriminant validity based upon satisfaction of the Campbell-Fiske criteria.

John Oller (Department of Linguistics, University of New Mexico, Albuquerque, NM 87131) presented a paper (co-authored with Ryuichi Yorozuya) entitled "Oral Proficiency Scales: Construct Validity and the Halo Effect." He investigated the construct validity of four 10 point scales of oral proficiency (grammar, vocabulary, pronunciation, and fluency) in a study using interviews with 10 foreign students, evaluated by 15 native speakers of English. Oller found that there was no unique reliable variance which could be attributed to the separate constructs of grammar, vocabulary, pronunciation, and fluency.

Pardee Lowe and Ray Clifford (C.I.A. Language School, P.O. Box 9523, Rosslyn Station, VA 22209) described a recorded oral proficiency test developed at the CIA for administration under conditions where a face-to-face interview was impractical. The test incorporates many of the features of the live interview used at the CIA (the systematic use of Lowe's question types for eliciting performance at different levels).

Susanna Brüttsch (224 Peik Hall, University of Minnesota, Minneapolis, Minnesota 55455) presented a paper entitled "Convergent/Discriminant Validation of Proficiency in Oral and Written Production of French." In this study, using 82 subjects, two traits, and three methods, she found evidence for convergent validity, but not discriminant validity.

Harold Madsen and Randy Jones (Departments of Linguistics and German, Brigham Young University, Provo, UT 84602) presented a paper entitled "A Survey of Oral Proficiency Tests: Phase II." This was an updated report of their large scale study of over 180 different oral tests they have collected over the last two years. They summarized their findings as related to test description and scoring procedures.

Meredith Pike (Center for Developing English Language Teaching, Faculty of Education, Ain Shams University, Cairo, Egypt) presented a paper entitled "An Investigation of the Interviewer's Role in Oral Proficiency Testing." She analyzed one interviewer's behavior in a series of oral interviews with foreign graduate students at UCLA. The behavior was analyzed on the basis of form and function, and these analyses contributed to a preliminary definition of interviewer consistency.

Brendan Carroll (English Teaching Information Centre, British Council, 10 Spring Gardens, London, England SW1A 2BN) presented a paper entitled "Measuring the Communicative Value of an Oral Performance." He examined

the adequacy of current criteria for assessing oral performance and outlined a model which incorporates communicative factors from the Munby model and which subsumes the language aspects (pronunciation, vocabulary, grammar, and fluency) commonly used as the basis for descriptions of spoken performance.

Arthur Hughes (Department of Linguistic Science, University of Reading, Reading, England RG6 2AA) presented a paper entitled "A Closer Look at Conversational Cloze." He described the features which distinguish the conversational cloze from the prose cloze and attempted to discover how the presence of any or all of them accounted for the greater success of conversational cloze in predicting oral ability.

Marianne Adams (Testing and Publications Office, School of Language Studies, Room 906 SA-3, Foreign Service Institute, Department of State, Washington, DC 20520) presented a paper entitled "The FSI Oral Interview: Test/Conversation." She described some of the basic features of the FSI oral interview (one of the tests used in the Bachman-Palmer study) and explained how the test is used at the FSI.

Elana Shohamy (School of Education, Stanford University, Stanford, CA 94305) presented a paper entitled "The Construct Validity of the Oral Interview Rating Scale." She evaluated the extent to which speaking proficiency as rated by teachers or linguists trained in using the FSI oral interview rating scales correlated with the ratings of speaking proficiency by lay native speakers using their own evaluation criteria. She found high inter-rater reliability for both lay and trained raters and a high correlation between the ratings assigned by both groups.

Helmut Vollmer (Universität Osnabrück, 45 Osnabrück, Postfach 4469, Federal Republic of West Germany) presented a paper entitled "On the Psycholinguistic Construct of an Internalized Expectancy Grammar." He reviewed some of the evidence for a general language proficiency (one factor) model and criticized it as a methodological artifact. He argued against the use of principle components factor analysis and proposed the use of designs that would allow investigators to apply confirmatory factor analytic techniques.

Andrew Cohen (The Hebrew University of Jerusalem, Israel) presented a paper entitled "Developing a Rating Scale for Testing Functional Speaking Ability." He described the development of a rating scale for assessing sociocultural competence based on contrastive analysis of sociocultural patterns.

Discussion of Directions for Future Research

In addition to the presentation and discussion of the papers, an entire afternoon was devoted to a discussion of possible directions for follow-up studies. The participants seemed to agree that there was sufficient evidence for discriminant validity of tests of communicative competence in speaking and reading to warrant a continued investigation into the internal structure of the construct "communicative competence." It was decided that the primary focus of the study should be to attempt

to assess whether raters could reliably distinguish between aspects of communicative competence such as those proposed by Canale and Swain in their framework. It was also suggested that actual tests used in the study should be as "natural" as possible in the hope that the results of the study would then be of greater interest to researchers in Europe interested in functional language tests. Finally, it was suggested that future colloquia allow for the discussion not only of construct validation studies but also of criterion related validation.

Other Participants in the Colloquia

In addition to the individuals cited above, a number of other researchers have contributed to the construct validation project, both by presenting papers at the first colloquium in Boston and by participating in the discussion sessions at the colloquia. These individuals are named below.

Dr. Michael Canale
Project Director, FSL Project
The Ontario Institute for Studies
in Education
252 Bloor Street West
Toronto, Ontario
Canada M5S 1V6

Dr. Francis Cartier
1029 Forest Avenue
Pacific Grove, CA 93950

Dr. John L.D. Clark
Educational Testing Service
Princeton, NJ 08540

Dr. Alan Davies
Department of Linguistics
University of Edinburgh
15 Buccleuch Place
Edinburgh 8, Scotland

Dr. Frances B. Hinofotis
Department of English ESL
UCLA
Los Angeles, CA 90024

Donna Ilyin
Placement Officer
San Francisco Community College
76 Sixth Ave.
San Francisco, CA 94118

Dr. Dale Lange
224 Peik Hall
University of Minnesota
Minneapolis, MN 55455

Stephen B. Ross
American Language Program
California State University
Long Beach, CA 90840

George E. Scholz
Director of Studies
American Language Institute
Rua Jose Falcao 15-5
Porto, 4000, Portugal

Dr. Charles Stansfield
Dept. of Spanish and Portuguese
University of Colorado
Boulder, CO 80302

Dr. Douglas K. Stevenson
R12R0 A14
Gesamthochschule Essen
Fachbereich 3 - Postfach 6843
Universitätsstr. 12
4300 Essen 1, West Germany