ILTA Guidelines for Practice in English

These Guidelines were presented as a draft at the 2005 ILTA meeting in Ottawa and then circulated among members for further consideration. They were finally adopted at the 2007 ILTA meeting in Barcelona and were reviewed and revised between 2018 and 2020.

Those involved in the original drafting were: Charles Alderson, Alan Davies (chair), Glenn Fulcher, Liz Hamp-Lyons, Antony Kunnan, Charles Stansfield and Randy Thrasher. The 2018-2020 revisions were drafted by Vivien Berry, Benjamin Kremmel (chair) and India Plough.

(Grateful acknowledgement is made to the Japanese Language Testing Association for their pioneer work on these Guidelines. Much of Part 1 is based on their work. Acknowledgement is also made to the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education, whose Standards for Educational and Psychological Testing informs Part 2 and to the International Test Commision, whose Guidelines on Computer-Based and Internet-Delivered Testing informs this update to the ILTA Guidelines.)

International Language Testing Association
Guidelines for Practice

(Adopted at the annual meeting of ILTA, held in Barcelona, June 2007)
(Updated version to be ratified by ABM in 2020)

Part 1 covers basic considerations for good testing practice in all test/assessment situations. Part 2 covers the rights and responsibilities of test takers. Part 3 outlines specific considerations for classroom-based language assessment. The document has been written for different audiences and therefore repetition is inevitable. Readers may refer initially to the part that best reflects their own context if they so wish.Unless otherwise noted, these guidelines also apply to the development and use of new technologies in language assessment. If unique guidelines are required for these modes of delivery, they are provided in the appropriate section.

Part 1

A. Basic Considerations for good testing practice in all test/assessment situations

1. The test developer's understanding of what the test, and each sub-part of it, is supposed to measure (its construct) must be clearly stated.
2. All tests, regardless of their purpose or use, must provide information which allows valid inferences to be made. Validity refers to the accuracy of the inferences and uses that are made on the basis of the test's scores. If, for example, the test is intended to measure the ability to use English in business communication, the inferences based on the test score are valid to the degree that the test does in fact measure that ability. However, since the ability to use English in business communication is a construct, the test developer must spell out just what that construct is or what it consists of. The test score inference or interpretation can be valid only if the test offers as accurate as possible a picture of the skill or ability it is supposed to measure.
3. All tests, regardless of their purpose or use, must be reliable. Reliability generally refers to the consistency of the results of a particular test, to what extent they are generalizable and therefore comparable across time and across settings.

B. Responsibilities of test designers and test writers

1. Test design should include a determination and explicit statement of the test's intended purpose(s).
2. A test designer must decide on the construct to be measured and state explicitly how that construct is to be operationalized.
3. The specifications of the test and the test tasks should be spelled out in detail and in language that is understandable to intended test users.
4. The work of the task and item writers needs to be edited before pretesting. If pretesting is not possible, the tasks and items should be analysed after the test has been administered but before the results are reported. Malfunctioning or misfitting

tasks and items should not be included in the calculation of individual test takers' reported scores.

5. Information guides on scoring (also known as grading or marking schemes) must be prepared for test tasks requiring human rating. These guides must be tried out to demonstrate that they permit reliable evaluation of the test takers' performance.
6. Those doing the scoring should be trained for the task and both inter- and intra-rater reliability should be calculated and published. If automated scoring systems are employed, they must be demonstrated to function reliably and validly.
7. Test materials should be kept in a safe place and handled in such a way that no test taker is allowed to gain an unfair advantage over other test takers.
8. Care must be taken to ensure that all test takers are treated in the same way in the administration of the test.
9. Scoring procedures must be carefully followed and score processing routines checked to make certain that no mistakes have been made.
10. Reports of the test results should be presented in such a way that they can be easily understood by test takers and other stakeholders.

Additional responsibilities for developers of computer-based tests (CBT) and internet-delivered tests

11. Test developers must provide evidence that CBT and internet-delivered tests are fail-safe, to the extent possible and, in the event of system failure, do not disadvantage test takers unduly.
12. Test developers should keep their testing system continuously functional without undue interruptions, for instance by minimizing, to the extent possible, the number of software and hardware updates that must be made.
13. All hardware, software, and browser requirements needed to administer and take the test must be clearly documented in lay language to the extent possible.
14. Hardware and software that can be modified for accomodations (e.g., font size; response format) must be used.
15. To increase accessibility, developers should use the minimum software and hardware features necessary to measure the construct.
16. Test developers must provide information on the procedures to follow in the event of software or hardware failures.
17. Practice tests and tutorials should be created so that test takers can become familiar with the functionalities of the test.
18. If the CBT or internet-delivered test is an adaptation of a pencil and paper test and the tests are used for the same purpose, developers must provide evidence of equivalence.

C. Obligations of institutions preparing or administering high stakes examinations

Institutions (colleges, schools, certification bodies, etc.) developing and administering entrance, certification or other high-stakes examinations must use test designers and item writers who are knowledgeable about current language

testing theory and practice. Items must be checked by someone with a high level of competence in the language being assessed.

Responsibilities to test takers and related stakeholders

(Before the test is administered)

The institution should provide all potential test takers with adequate information about the purposes of the test, the construct (or constructs) the test is attempting to measure and the extent to which that has been achieved. Information should also be provided as to how the scores/grades will be allocated and how the results will be reported.

In the case of CBT and internet-delivered tests, the institution must ensure that technical requirements are met, must train administrators and test takers in the functionalities of the test, and must provide testtakers time to practice and to become familiar with those functionalities.

(At the time of administration)

The institution shall provide facilities for the administration of the test that do not disadvantage any test taker. Test administration materials should be carefully prepared and proctors trained and supervised so that each administration of the test can be uniform, ensuring that all test takers receive the same instructions, time to do the test, and access to any permitted aids. If something occurs that calls into question the uniformity of the administration of the test, the problem should be identified and any remedial action to be taken to offset the negative impact on the affected test takers should be promptly announced.

In the case of speaking tests, the facilities shall be capable of proper invigilation and oversight, providing a safe and secure environment in professional surroundings for raters/interlocutors and for test takers.

In the case of CBT and internet-delivered tests, the institution must ensure that technical support is available before, during, and immediately after the time of testing. Plans should be in place to respond to technical difficulties (e.g., disconnections, system crashes).

(At the time of scoring)

The institution shall take the steps necessary to see that each test-taker's performance is scored/graded accurately and the result correctly placed in the database used in the assessment. There should be ongoing quality control checks to ensure that the scoring process is working as intended.

(Other considerations)

If a decision is to be made on candidates who did not all take the same test or the same form of a test, care must be taken to ensure that the different measures used are in fact comparable.

If more than one form of the test is used, inter-form reliability estimates should be published as soon as they are available.
If test data are hand-entered into computers, procedures must be in place to check the accuracy of input.

D. Obligations of those preparing and administering publicly available tests

They should:
1. Make a clear statement as to what groups the test is appropriate for and for which groups it is not appropriate.
2. Make a clear statement of the construct the test is designed to measure in terms a layperson can understand.
3. Publish validity and reliability estimates and bias reports for the test along with sufficient explanation in accessible language to allow potential test takers and test users to decide if the test is suitable in their situation.
4. Report the results in a form that will allow test users to draw the appropriate and meaningful inferences from them.
5. Refrain from making any false or misleading claims about the test.
6. Speak up publicly, consider legal measures or even withdrawing a test when a test is misused.
7. Publish a document for test takers which:

   7.1. Explains the relevant measurement concepts so that they can be understood by non-specialists.
   7.2. Reports evidence of the reliability and validity of the test for the purpose for which it was designed.
   7.3. Describes the scoring procedure and, if multiple forms exist, the steps taken to ensure consistency of results across forms.
   7.4. Explains the proper interpretation of test results and any limitation on their accuracy, including the maximum length of time that test scores will be valid for.

E. Responsibilities of users of test results

Persons who use test results for decision making must:

1. Use results from a test that is sufficiently reliable and valid to allow fair decisions to be made.
2. Make certain that the test construct is relevant to the decision to be made.

3. Clearly understand the limitations of the test results on which they will base their decision.
4. Take into consideration the standard error of measurement (SEM) of the test that provides the data for their decision.
5. Be prepared to explain and provide evidence of the fairness and accuracy of their decision making process.
6. Take into consideration that test providers cannot give an indeterminate guarantee of score validity as there is the possibility of language attrition after the test date.

F. Special considerations

In norm-referenced testing:

The characteristics of the population on which the test was normed must be reported so that test users can determine if this group is appropriate as a standard to which their test takers can be compared.

In criterion-referenced testing:

The appropriateness of the criterion must be confirmed by experts in the area being tested.

Since correlation is not in itself a sufficient way of determining the reliability and validity of criterion-referenced tests. other methods appropriate for analysing such test data (e.g. forms of qualitative analyses) must also be used.

Part 2 - Rights and Responsibilities of test takers

A. As a test taker you have the right to:

1. Be informed of your rights and responsibilities as a test taker.
2. Be treated with courtesy, respect and impartiality, regardless of your age, disability, ethnicity, gender, national origin, religion, sexual orientation or other personal characteristics.
3. Be tested with measures that meet professional standards and that are appropriate, given the manner in which the test results will be used.
4. Receive a brief oral or written explanation prior to testing about the purpose(s) for testing, the kind(s) of tests to be used, whether the results will be reported to you or to others, and the planned use(s) of the results. If you have a disability, you have the right to inquire and receive information about testing accommodations (special arrangements). You have a right to know in advance of testing whether any such special arrangements may be available to you.
5. Know in advance of testing when the test will be administered, if and when test results will be available to you and if there is a fee for testing services that you are expected to pay.
6. Have your test administered and your test results interpreted by appropriately trained individuals who follow professional codes of ethics.
7. Know if a test is optional and learn of the consequences of taking or not taking the test, fully completing the test, or cancelling the scores. You may need to ask questions to learn about these consequences.
8. Receive a written or oral explanation of your test results within a reasonable amount of time after testing and in commonly understood terms.
9. Have your test results kept confidential to the extent allowed by law.
10. Present concerns about the testing process or your results and receive information about procedures that will be used to address such concerns.
11. In the case of CBT and internet-delivered tests, be informed of the minimum system requirements in order to take the test.
12. In the case of CBT and internet-delivered tests, be provided ample time to become familiar with the technology and the functionalities of the system.

B. As a test taker, you have the responsibility to:

1. Read and/or listen to your rights and responsibilities as a test taker.
2. Treat others with courtesy and respect during the testing process, and do nothing to prevent other test takers from performing to the best of their abilities.
3. Ask questions prior to testing if you are uncertain about why the test is being given, how it will be given, what you will be asked to do, and what will be done with the results.
4. Read or listen to descriptive information in advance of testing and listen carefully to all test instructions. If you have difficulty comprehending the language of the test instructions, it is your responsibility to inform an examiner. You should inform an examiner in advance of testing if you wish to receive a testing

accommodation or if you have a physical condition or illness that may interfere with your performance on the test.

5. Know when and where the test will be given, pay for the test if required, arrive on time with any required materials and be ready to be tested.
6. Maintain the integrity of the test by not seeking or promoting the assistance of other individuals during the test.
7. Maintain the integrity of the test by keeping the content of the test confidential (not releasing information on the test content).
8. Be familiar with and accept the consequences of not taking the test, should you choose not to take the test.
9. Inform appropriate person(s), as specified to you by the organization responsible for testing, if you believe that testing conditions affected your results.
10. Ask about the confidentiality of your test results, if this aspect concerns you.
11. Present any concerns you may have about the testing process or results in a timely, respectful way.

Part 3 - Guidelines for Practice in Classroom-based Language Assessment

In this section we will first of all differentiate between *tests* and *assessment* and between *formal* and *informal* classroom testing and assessment.

Tests can be understood to mean instruments for the measurement of given knowledge at a specific point in time, such as at the end of a unit of learning, or at mid-term or end-of-year, or at any specified point in time determined by either an institution or external authority. Tests (sometimes known as examinations or exams) often collect information in numerical form such as scores that allow for hierarchical ranking in percentiles. *Formal testing* refers to situations where the results of tests have specific consequences such as moving up to a higher year level, repeating a unit or course of learning or gaining entrance to a higher educational institution.

Assessment refers to any of the procedures teachers use in their classrooms for diagnostic or dynamic purposes, to relate learning to the curriculum, as well as assessment of and for learning. This includes determining what students' potential for learning is and how effectively they are learning. It also includes both formative and summative assessments. It can also be used to determine the effectiveness of teaching methods and the materials being used. Procedures can include both *formal assessments* such as tests, and *informal assessments* such as observations, interviews, questionnaires, checklists, performance and portfolio assessments, peer and self assessment, and discussions between teacher and student. Informal assessments do not rank or judge students and the information gained from them should generally only be used to improve learning outcomes.

> A Responsibilities of teachers as test writers of both formal and informal tests, quizzes and assessment activities (including when selecting or adapting from different test materials)
>
> 1. An explicit statement of the test's or assessment's intended purpose (such as diagnostic, summative evaluation, formative assessment, achievement) should be articulated.
> 2. When selecting or adapting tests or assessment activities from existing materials, the teacher/test writer should ensure that the original materials are suitable in terms of purpose, construct and intended target population for the current test takers.
> 3. Tests and assessment activities must include clear, unambiguous instructions for test takers/participants.
> 4. In the case of achievement tests for summative purposes, the teacher/test writer must decide on the knowledge and skills to be measured and state explicitly how the test measures these.
> 5. Scoring guides must be prepared in advance for those tests requiring human scoring.
> 6. Those doing the scoring should be trained for the task with the scoring guide developed for this purpose.

7. Scoring procedures must be carefully followed and score processing routines checked to make certain that no mistakes have been made.
8. Test materials should be kept in a safe place and handled in such a way that no test taker is allowed to gain an unfair advantage over other test takers.
9. Before administration, the test should be reviewed by colleagues to ensure clarity and to avoid ambiguity of items.
10. Care must be taken to ensure that all test takers are treated in the same way in the administration of the test.
11. Reports of the test results should be presented in such a way that they can be easily understood by test takers and other stakeholders.
12. When developing implicit and instruction-embedded tests internal to the classroom, including both planned and spontaneous assessments, the teacher/test writer should ensure that the test-takers understand the intended purpose of the test, how and by whom the test will be scored, and the use that will be made of the results.

B. Basic considerations for good testing practice in all formal situations

1. The purpose and content of the test are clearly stated.
2. All tests, regardless of their purpose or use, must provide information which allows valid inferences to be made. Validity refers to the accuracy of the inferences and uses that are made on the basis of the test's scores. If, for example, a reading test purports to be measuring the ability to transfer information presented in a text to a chart, the inferences based on the test score are valid to the degree that the test does in fact measure that ability. However, since the ability to transfer information from one medium to another is a construct, the test developer must spell out just what that construct is or what it consists of. The test score inference or interpretation can be valid only if the test construct offers as accurate as possible a picture of the skill or ability it is supposed to measure.
3. All tests, regardless of their purpose or use, must be reliable. Reliability refers to the consistency of the test results, to what extent they are generalizable and therefore comparable across time and across settings and from one rater/grader to another if there are multiple raters/graders.
4. Teachers responsible for creating and administering tests should be familiar with the testing procedures and codes of conduct of their local institution. Where possible, external monitoring of test writers and administrators should take place in order to handle complaints procedures effectively.

C. Obligations of teachers preparing and administering formal tests

Responsibilities to test takers and related stakeholders

(Before a formal test is administered for summative purposes)

1. Test takers must be provided with clear information regarding the day, time, and location of the test.
2. Test takers should be told the length of time allowed to take the test.
3. Test takers should be provided with information about the purpose of the test, the content of the test, and the scoring procedures.
4. The scoring procedure must be explained in terms that each test taker can understand.

(At the time of administration)
1. The location of the test should not disadvantage any test taker. All test takers should receive the same instructions and be provided access to any permitted aids.
2. If something occurs that calls into question the equity of the administration of the test, the problem should be identified and any remedial action to be taken to offset the negative impact on the affected test takers should be promptly announced. Remedial action may include a change of accommodation for particular test takers, such as those with differing physical disabilities.
3. The test administrator is responsible for monitoring and maintaining the integrity of the testing environment. Any violations must be documented, discussed with the test taker(s) at the end of the test, and reported to appropriate individuals prior to scoring.
4. The test administrator has the right and responsibility to remove any individual whose behavior disrupts and interferes with the ability of other individuals to take the test.
5. The test writer and administrator have an obligation to intervene when there is misuse of the test.

(At the time of scoring)
1. Necessary steps should be taken to ensure that each test is scored/graded accurately and the result correctly recorded.
2. There should be ongoing quality control checks to ensure that the scoring process is working as intended.
3. Results must be reported in a form that will allow test users (e.g. test takers, parents, principals,…) to draw the correct inferences from them.
4. No false or misleading claims should be made about the test.
5. The proper interpretation of test results and any limitations in the interpretation of the results must be explained.
6. There should be clear complaint procedures and policies in place, preferably with external oversight (if possible).

D. Responsibilities of teachers as users of <u>formal</u> test results

Persons who use test results for decision-making must:

1. Use results from a test that is sufficiently reliable and valid to allow fair decisions to be made.
2. Make certain that the knowledge and skills tested are relevant to the decisions to be made.
3. Clearly understand the limitations of the test results on which they will base their decisions.
4. Be prepared to explain and provide evidence of the fairness and accuracy of their decision-making processes.
5. Outline a transparent and fair appeal or re-marking process.
6. Provide feedback to students with suggestions for enhancing future learning.

E. Rights and Responsibilities of test takers in <u>formal</u> contexts

As a test taker, you and your parent or guardian if you are a minor, have the right to:
1. Be informed of your rights and responsibilities as a test taker.
2. Be treated with courtesy, respect and impartiality, regardless of your age, disability, ethnicity, gender identity, national origin, religion, sexual orientation or other personal characteristics.
3. Be tested with measures that meet professional standards and that are appropriate, given the manner in which the test results will be used.
4. Receive a brief oral or written explanation prior to testing about the purpose(s) for testing, the kind(s) of tests to be used, whether the results will be reported to you or to others, and the planned use(s) of the results. If you have a disability, you have the right to inquire and receive information about testing accommodations (special arrangements). You have a right to know in advance of testing whether any such special arrangements may be available to you.
5. Know in advance of testing when the test will be administered, if and when test results will be available to you
6. Have your test administered and your test results interpreted by appropriately trained individuals who follow professional codes of ethics.
7. Know if a test is optional and learn of the consequences of taking or not taking the test, fully completing the test, or cancelling the scores. You may need to ask questions to learn about these consequences.
8. Receive a written or oral explanation of your test results within a reasonable amount of time after testing and in commonly understood terms.
9. Have your test results kept confidential to the extent allowed by law.

10. Present concerns about the testing process or your results and receive information about procedures that will be used to address such concerns.

As a test taker, you and your parent or guardian if you are a minor, have the responsibility to:

1. Read and/or listen to your rights and responsibilities as a test taker.
2. Treat others with courtesy and respect during the testing process.
3. Ask questions prior to testing if you are uncertain about why the test is being given, how it will be given, what you will be asked to do, and what will be done with the results.
4. Read or listen to descriptive information in advance of testing and listen carefully to all test instructions. You should inform an examiner in advance of testing if you wish to receive a testing accommodation or if you have a physical condition or illness that may interfere with your performance on the test. If you have difficulty comprehending the language of the test, it is your responsibility to inform an examiner.
5. Know when and where the test will be given, appear on time with any required materials and be ready to be tested.
6. Follow the test instructions you are given and represent yourself honestly during the testing.
7. Maintain the integrity of the test by not seeking or promoting the assistance of other test takers during the test and keep the content of the test confidential (do not release information on the test content).
8. Be familiar with and accept the consequences of not taking the test, should you choose not to take the test.
9. Inform appropriate person(s), as specified to you by the teacher or other staff members responsible for testing, if you believe that testing conditions affected your results.
10. Ask about the confidentiality of your test results, if this aspect concerns you.
11. Present any concerns you may have about the testing process or results in a timely, respectful way.

F. Obligations of teachers conducting <u>informal or spontaneous</u> classroom assessment activities

1. Teachers must ensure learners are informed that they will be assessed during the activity, even if this has not been announced in advance.
2. Care must be taken to ensure that students understand how they will be assessed and by whom.
3. Teachers must inform learners how scores from self, peer and group assessments will be used.
4. Students should be informed how individual scores and group scores will be allocated when performing group assessment tasks.

5. Necessary care should be taken to ensure all learners have appropriate and equitable opportunity to participate in an activity to the best of their ability.
6. Outcomes from informal or spontaneous classroom assessments should only be used to inform further instructional purposes and should not be used to demean students in future classroom activities.
7. Assessment of classroom activities should not be presented publicly in the form of scores and participants should not be ranked publicly, unless required by school policy, law and/or responsible authorities'.
8. Feedback from informal or spontaneous classroom activities should be provided to students individually or in pairs/groups as appropriate to enhance learning.
9. Opportunities for students to review outcomes and discuss in pairs/groups their understanding of feedback should be provided, so students are enabled to take responsibility for their own learning and understand how to move forward using the feedback.