Assessment and Testing: Overview

Carol Chapelle

Iowa State University, carolc@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/engl_pubs

Part of the Language and Literacy Education Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/engl_pubs/206. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Book Chapter is brought to you for free and open access by the English at Iowa State University Digital Repository. It has been accepted for inclusion in English Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Assessment and Testing: Overview

Abstract
What language testing does is to compel attention to meaning of ideas in linguistics and applied linguistics. Until they are put into operation, described, and explained, ideas remain ambiguous and fugitive. A test forces choice, removes ambiguity, and reveals what has been elusive: thus a test is the most explicit form of description, on the basis of which the tester comes clean about his/her ideas. (Davies, 1990, p. 2)

These words express what many applied linguists recognize: that language assessment and testing intersects almost all language-related issues that applied linguists study...

Disciplines
Language and Literacy Education

Comments
ASSESSMENT AND TESTING: OVERVIEW


Suggested Readings


Assessment and Testing: Overview

CAROL A. CHAPELLE AND LIA PLAKANS

What language testing does is to compel attention to meaning of ideas in linguistics and applied linguistics. Until they are put into operation, described, and explained, ideas remain ambiguous and fugitive. A test forces choice, removes ambiguity, and reveals what has been elusive: thus a test is the most explicit form of description, on the basis of which the tester comes clean about his/her ideas. (Davies, 1990, p. 2)

These words express what many applied linguists recognize: that language assessment and testing intersects almost all language-related issues that applied linguists study. These
intersections take place at the level of ideas, but typically also in many of the everyday practices in language teaching, certification, and gate keeping, where language tests are used. In an intensive French program, students are tested for their level of French-language ability for placement into an appropriate level of a class. Applicants hoping to be admitted to a university or hired for a job requiring spoken English have to attain a particular score on designated language tests. In some cases, would-be immigrants have to attain a particular score to have their application for immigration to a particular country accepted. In research on second language acquisition, scores on language tests are often used as evidence for learners’ development. In the classroom, assessment results provide information to teachers and to students about their learning. In educational testing, language ability is assessed for both native speakers and non-native speakers, but the researchers and practitioners that align themselves with applied linguistics typically work with second language assessment, and therefore so do the entries in the encyclopedia.

Across the many different testing contexts, both the terms “test” and “assessment” are used to denote the systematic gathering of language-related behavior in order to make inferences about language ability and capacity for language use on other occasions. This core meaning of the two terms links them even though each is more strongly associated with certain collocates and contexts of use. The expression “high-stakes testing” is more frequent than “high-stakes assessment” (with a ratio of approximately 3 : 2 based on a Google search of the Web), and the expression “classroom assessment” is more frequent than “classroom testing” (with a ratio of approximately 3 : 1). Throughout the encyclopedia both of these terms are used, sometimes interchangeably, but typically with their preferred collocates. The title of this section contains both terms because of the different meanings that they hold for some people. Although “language testing” is the preferred expression on the Web (with over a million instances), language assessment (with not even a half a million instances) denotes for many people a more varied process of data gathering and interpretation than language testing.

Assessment Uses

The selection of the term—assessment or testing—is often governed by the use of the test, and accordingly assessment uses are described in several entries in the encyclopedia. One entry explains the basic approaches that help users to understand the requirements of different uses of language assessments. Another entry, high-stakes language testing, considers test uses with serious consequences, in education or elsewhere. Test uses in the educational context are described in three entries: assessment in the classroom, language assessment in program evaluation, and language testing and accountability. Two entries describe test uses connected to government, language testing and immigration, and language testing in the government and military. One entry describes language testing in second language research. The language testing needs across these domains differ, and therefore the specific approaches to and aspects of language tested differ as well.

What Is Tested?

As the breadth of this encyclopedia attests, language is a complex phenomenon that can be conceptualized in many different ways. This complexity is apparent in the many aspects of language that are assessed. Four of the entries describe issues and practices in assessing systems of knowledge: assessment of grammar, assessment of vocabulary, assessment of pragmatics, and assessment of cultural knowledge. Entries cover assessment of each one of “the four skills,” as they have been traditionally called: assessment of
LISTENING, ASSESSMENT OF READING, ASSESSMENT OF SPEAKING, and ASSESSMENT OF WRITING. Recognition of limitations with such a four-skill view of language, however, has prompted test developers to construct tests of combined skills, such as a test that would require the examinee to read a text and then speak about it. Such ASSESSMENT OF INTEGRATED SKILLS has proven to be an important avenue of research and development, particularly in tests of academic language ability whose scores are intended to reflect how well examinees' language will allow them to perform on academic tasks, which typically involve a combination of skills. Regardless of the particular aspect of language tested, the conception of language that underlies a particular test is expressed in a model of language, and the ideas behind construction of models in language assessment are described in MODELING LANGUAGE FOR ASSESSMENT.

Assessment Methods

Language assessment methods have been developed to prompt examinees in a manner that will elicit relevant performance. The entry on LANGUAGE ASSESSMENT METHODS describes the work that has gone into devising appropriate methods to capture the detail of the linguistic knowledge that underlies examinees' performance as well as the performance itself. Whether the examinees' performance is intended to be used as a sign of their language knowledge or as a sample of their potential performances on a similar range of tasks, the methods used to gather the performance are critical in determining what the resulting test score can mean. As a consequence, schemes have been developed to describe the particular facets of language test methods, most notably by Bachman (1990). In addition, new combinations of test method facets have resulted in considerable interest and research into such testing methods as PAIRED AND GROUP ORAL ASSESSMENT, SELF-ASSESSMENT, and TASK-BASED ASSESSMENT. The encyclopedia therefore contains an entry on each of these topics.

One facet of the test methods—evaluation of examinees' language—is an area of central importance. Some aspects of language knowledge can be evaluated through the use of selected-response items (such as multiple choice); however, most language-test uses require a sample of language performance that the examinee produces, whether it be a word, a sentence, a turn in an interactive dialogue, a written essay, or one of many other possibilities. Whereas responses on a selected-response test can be scored as correct or incorrect, rating of linguistic performance is based on a more complex set of considerations. As a result, research into RATING SCALES FOR LANGUAGE TESTS is an ongoing area of activity in language assessment. Rating of written language is a challenge that is common to both native language and second language assessment, but in second language assessment, tests of oral language ability are also widely used, and therefore RATING ORAL LANGUAGE is an area of particular attention in applied linguistics. With advances of knowledge and practice in computational linguistics, a promising area of inquiry is the study of AUTOMATED ESSAY EVALUATION AND SCORING and COMPUTER SCORING OF SPOKEN RESPONSES.

Developing Assessments

Many applied linguists work in areas where their expertise is in language test development. In this context, knowledge of language, models of language ability, and testing methods are needed in addition to process knowledge about how to design tests using TEST SPECIFICATIONS. Test design often takes place in a context where a test already exists, and therefore one entry examines together the related processes of TEST DESIGN AND
It is commonplace today for language tests to be computer delivered even if the language performance is not automatically scored by computer. Therefore, an important part of test development is an understanding of AUTHORING TOOLS FOR LANGUAGE ASSESSMENT. A final process in moving a new test from development to use is setting CUT SCORES ON LANGUAGE TESTS. This process makes the test results usable because it allows test users to assign meaning to them, typically in terms of what interpretations and actions one should be able to make on the basis of a specific test score. All of these test development processes are critical for developing high-quality tests, but a complementary set of processes are also needed for evaluating assessments.

Evaluation of Language Assessments

Evaluation takes place through validation research which is conducted to support the interpretations and uses of test scores. VALIDATION OF LANGUAGE ASSESSMENTS is described in one entry as the organizing framework for all of the research that supports test score use. This research consists of a vast array of types, not all of which are covered in the encyclopedia. The encyclopedia aims to provide a conceptual overview of the approaches to research to demonstrate how they fit into the process of validation. STATISTICAL ANALYSIS OF TEST RESULTS outlines the fundamental item-level statistics that are typically done to examine the empirical performance of the test and the individual items for each examinee group. RELIABILITY AND GENERALIZABILITY OF LANGUAGE TESTS explains the importance of reliability, or consistency, as a characteristic of test scores and the reliability issues that researchers face in language assessment. Examining another approach to analysis of consistency, GENERALIZABILITY THEORY IN LANGUAGE TESTING, explains an analytic method for estimating consistency that yields an estimate of the amount of error contributing to scores by sources such as raters and tasks.

Two entries provide a basis for understanding some of the other quantitative methods used in validation research: CORRELATIONAL RESEARCH IN LANGUAGE ASSESSMENT and STRUCTURAL EQUATION MODELING IN LANGUAGE ASSESSMENT. Two entries introduce qualitative methods used in validation research: DISCOURSE ANALYSIS IN LANGUAGE ASSESSMENT and QUALITATIVE RESEARCH IN LANGUAGE ASSESSMENT. Finally, two entries explain types of inquiry that require critical and empirical examination of the ideological and consequential basis for test use in a particular context: VALUES IN LANGUAGE ASSESSMENT and WASHBACK IN LANGUAGE ASSESSMENT.

Issues in Language Assessment

The extensive use of assessment in all areas of research and education raises a number of issues that are constant sources of concern for language testing researchers. One such issue is explained in an entry on ENGLISH-LANGUAGE LEARNERS AND SUBJECT AREA TESTS, which describes the difficulty in interpreting scores on school achievement tests of math or science, for example, as indicators of English-language learners' knowledge of those content areas. This issue arises in the context of school-age learners, as does the issue of ASSESSMENT OF YOUNG ENGLISH-LANGUAGE LEARNERS for obtaining scores that reflect their language ability.

The fact that performance on content tests assumes proficiency in English creates a potential for test bias. Another entry describes test BIAS IN LANGUAGE ASSESSMENT, which can also arise in language tests when one group of test takers systematically performs differently than another for reasons unrelated to the ability that the test is measuring. Bias is one of the issues pertaining to FAIRNESS IN LANGUAGE ASSESSMENT, which is discussed more fully in another entry.
Another issue of growing importance related to fairness is *standard setting on language tests*, which entails the concepts and procedures used to make the link between a test score and a particular score meaning. In an environment where educational bodies are generating proficiency level descriptors and test developers are required to demonstrate that test scores can be interpreted on those scales, standard-setting activities are in the spotlight. Research in this area is particularly interesting when researchers investigate *assessment across languages*.

In a context where language assessments are so pervasive and consequential for test takers, the title of Shohamy’s (2001) book, *The Power of Tests*, sets the stage for the final two issues. One is *language assessment literacy*, the knowledge that applied linguists need to have in order to construct, analyze, and interpret language tests. The second is *ethics in language assessment*, the responsibility to use that knowledge with an awareness of who may benefit and who may be harmed from it.

Research and practice in language testing is constituted from the work of a large and diverse community, whose members span a range of professions from academia to government, to business. The following distinguished scholars were chosen to feature in biographical entries from a much larger set of distinguished members of the profession: Charles J. Alderson, Lyle F. Bachman, John B. Carroll, John L. D. Clark, Alan Davies, Robert Lado, John W. Oller, Jr.; and Charles Stansfield. Their biographies reveal a diverse set of backgrounds and different paths that brought them to language assessment, but what they share is a pragmatic motivation driving their scholarship and practice—a motivation to improve professional practices in language assessment.

**SEE ALSO:** Shohamy, Elana; Spolsky, Bernard

**References**


**Suggested Readings**


