

ACCESS for ELLs[®]

Language Testing
28(3) 425–431
© The Author(s) 2011
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0265532211404195
ltj.sagepub.com


Reviewed by: Janna Fox, *Carleton University, Canada* and Shelley Fairbairn, *Drake University, USA*

We want to encourage teachers to ‘teach to the standards, not to the test’... to promote the use of standards to guide students’ language learning rather than focusing instruction on preparing them for possible test items. In seeing the close relationship between the standards and the test, we believe that teachers can feel confident that by incorporating the standards into classroom instruction and assessment, they will better prepare their students for linguistic proficiency and in turn, provide access to academic success. (Boals, 2008, p. 3)

Assessing Comprehension and Communication in English State-to-State for English Language Learners (*ACCESS for ELLs[®]*), is a large-scale, high-stakes,¹ standards-based, and criterion-referenced English language proficiency test administered in the USA annually to more than 840,000 English Language Learners (ELLs), in K-12 classrooms. The test is one of the testing, curricular, professional development and teaching resources developed by the World-Class Instructional Design and Assessment (WIDA) Consortium (see www.wida.us). WIDA was formed in 2003 through a US Department of Education Enhanced Assessment Grant by the states of Wisconsin, Delaware, and Arkansas² in order to develop a standards and assessment system, to plan support for continuing English language development, and to satisfy legal requirements for assessment and accountability as spelled out in the No Child Left Behind (NCLB) Act (www.wida.us). Arguably, what distinguishes *ACCESS for ELLs[®]* from other NCLB English Language Proficiency (ELP) tests is (1) the size of the Consortium of states using the test, which has since grown to 24; (2) the number of resources available to situate the test within pedagogical practice; and (3) the care with which WIDA has attempted to systematically align the test with proficiency standards (see *WIDA English Language Proficiency Standards*, 2004, 2007) and classroom practice.

Test purpose

The test assesses social and general academic English in reading, speaking, listening, and writing, as well as language used in language arts, mathematics, science, and social studies, as exemplified in the *WIDA English Language Proficiency Standards* (2004, 2007). These standards, which have been aligned to the academic content standards of the 24 WIDA member states, have been adopted by Teachers of English to Speakers of Other Languages (TESOL), ‘to form the basis of the next national NCLB-compliant model standards for ELP Standards’ (www.wida.us). *ACCESS for ELLs[®]* is distributed to member states at a cost of \$23/student. This fee not only includes the test itself, but a large

number of pedagogical supports, resources and services. ELLs, referred to as Limited English Proficient (LEP) students by the US government, are required to take an English language proficiency test annually until they reach a proficiency level which is deemed to be English language proficient. This level varies by state. For example, according to the Illinois State Board of Education, 'As of January 1, 2010, students who obtain an overall composite proficiency level of 4.8 as well as a 4.2 composite literacy (reading/writing) proficiency level on the annually administered state approved English language proficiency test, *ACCESS for ELLs*[®], are to be considered English language proficient' (www.isbe.state.il.us/). Other states publish different requirements for their individual English language proficient designations (for full details consult the Department of Education sites of the WIDA member states at www.wida.us/states/index.aspx).

Background on development

ACCESS for ELLs[®] was launched in 2005 under the direction of the Center for Applied Linguistics (CAL), the test's principal developer. Since its launch, it has benefitted from the on-going support of a number of collaborating partners such as the Wisconsin Center for Education Research at the University of Wisconsin-Madison, currently the host institution for the WIDA Consortium. Technical reports that document the history of test development are readily available on the WIDA website. They provide evidence of the exemplary research tradition that has guided the test's development over the years, and support for the developers' claims that the test reflects current theory and research on academic language (Gottlieb & Kenyon, 2006; Gottlieb, Cranley, & Cammilleri, 2007). There is also evidence that the research undertaken by the WIDA Consortium has led to important changes in the test itself. For example, Gottlieb et al. (2007) summarize how research led to an expansion of proficiency levels accounted for by the test to include 'Reaching' – the level at which students are said to have achieved proficiency in academic English.

The test developer has not identified a 'Monitoring and Maintaining' level of ELL proficiency, beyond the *Reaching* threshold. Thus, there are no criteria that define characteristics of ELL learners who, having reached the proficiency threshold, continue to demonstrate effective engagement with academic work in on-going classroom activity. However, information about the test is systematically related to multiple WIDA-developed pedagogical resources that support the day-to-day learning and language development of ELLs (including those deemed to have reached proficiency in academic and social English by virtue of the test). In keeping with the Boals' (2008) assessment philosophy, connecting the test with such resources is essential, if teachers are to be guided by and to incorporate the standards in classroom instruction and assessment. The integration of test information with on-going classroom assessment is also consistent with the views of many prominent assessment experts (Clark, 1998; Gottlieb, 2006; Lynch, 2001).

Test design, format, administration, and use

ACCESS for ELLs[®] has a necessarily complex design. Each of the four language domains are tested through thematic 'folders' tied to the five standards (i.e. one section for social/general academic language and for each of the four corresponding 'core' content areas). The tests operationalize the five language proficiency levels defined by the test's

construct, from ‘entering’ (level 1) to ‘bridging’ (level 5).³ Anchored in the *WIDA ELP Standards* (2004, 2007), test forms are generated annually for five grade level clusters, namely, Kindergarten, grades 1–2, 3–5, 6–8, and 9–12. The test developers describe this continuum of increasing language as the *horizontal* dimension of the test (Gottlieb et al., 2007, p. 12).

However, test designers have also introduced what they refer to as a *vertical* dimension within each of the grade level clusters above Kindergarten, by further dividing each level into three overlapping tiers – A (beginning ELLs), B (intermediate), and C (advanced) – to reflect the range of proficiency within each of the clusters and across the four domains of listening, speaking, reading, and writing. The vertical dimension (or scaling) allows test users to assess finer-grain differences in student performance as students move across *tiers* both within or across grade level clusters. Thus, within a grade level cluster, Tier A would contain easier items or tasks than Tier C. A test taker who responded correctly to 10 items in reading on a Tier A form would receive a lower scale score than a test taker who answered 10 reading items correctly on the Tier C form. Vertical scaling also makes it possible to see differences in degrees of increasing proficiency as students move across *grade clusters*.

Although the test developer provides ample information about test tasks and their evaluation in the many resources provided on the website, there is no sample test which provides students and their teachers with an idea of just what the test looks like as a whole, or what precisely a test taker will encounter in taking the test. There are warm-up items for test takers within the test, and model tasks for both teachers and students to review on the test website. There is pictorial support provided throughout these model tasks (suggesting that this is the case for test versions), and the increasing level of difficulty is evident within tiers and across test levels. Further, the Consortium provides online test administrator training courses, sample items, tasks, and rubrics in order to facilitate classroom activities tied to the standards and representative of the kinds of language production expected on the *ACCESS for ELLs*[®].

Reliability

Evidence of reliability is provided by numerous technical reports which are listed on the test’s website. For example, Technical Report no. 1 (2006), provides information on the conceptualization of the assessment, from its anchor in the ELP standards through each developmental phase, and documents procedures for standard-setting, which determined the cut-scores for the six language proficiency levels. Technical Report no. 4 (2007) explains how grade level cluster cut scores were converted to grade specific cut-scores. Technical Report no. 4 states that reliabilities were quite high for composite scores: .937 for Kindergarten; .955 for grades 1–2, .935 for grades 3–5, .937 for grades 6–8, and .942 for grades 9–12 (2007, p. iv). Reliabilities for cut scores are also provided in this Technical Report.

Reporting scores

The test produces three types of scores: (1) *raw scores* indicate the number of items/tasks a test taker responded to correctly out of the total number of items or tasks on the test; (2)

psychometrically derived *scale scores* from 100 to 600, which report on grade levels in relation to the continuum of language development, allowing test users to compare student progress across the grades from Kindergarten to grade 12; (3) *proficiency level scores* which interpret scale scores. They describe test performance in a single digit in terms of the six WIDA language proficiency levels (1 – Entering, 2 – Beginning, 3 – Developing, 4 – Expanding, 5 – Bridging, and 6 – Reaching), followed by a decimal to the tenth. The number to the left of the decimal indicates the test taker’s proficiency level; the number to the right of the decimal indicates to the nearest tenth, the test taker’s degree of proficiency in relation to the range between cut scores. Confidence levels are also reported by the test developer to encourage test users to interpret scores cautiously and to take into account standard errors of measurement.

A consistent message to test users throughout the documentation describing the test and WIDA Consortium resources is that the information provided by the test is only one source of evidence in consideration of the language development of linguistically and culturally diverse learners. This message may not be fully appreciated, however, as the definitive statement on the Illinois Department of Education website above suggests. Such test users appear to define ‘fully English proficient’ solely in relation to cut-scores on the test and without consideration of additional evidence.

Overall evaluation: ELL-specific considerations

How well does *ACCESS for ELLs*[®] meet the eight requirements set out by Fairbairn and Fox (2009, p. 2) for the development of tests for ELLs?

1. Ensure that test language is accessible

The language used throughout the test appears to be systematically aligned to the WIDA ELP Standards. Further, students are given tests that are in keeping with their individual English language proficiency levels through the tiered system. This consideration is clearly met, as it well should be in a test designed for English language learners.

2. Include graphic/visual support

As indicated above, the model tasks available on the WIDA website suggest there is ample use of graphic/visual support throughout the levels of *ACCESS for ELLs*[®] in the form of color illustrations. Some photos and maps are being introduced into new items.

3. Adjust the test development process to include local and situated perspectives

One key way in which this consideration is addressed is through the development of items by educators (both teachers and administrators). These items are refined by staff at the Center for Applied Linguistics and are then reviewed by a different group of educators prior to field testing. Following field testing, a third group of educators reviews the

items once more. Representatives of WIDA-member state departments of education also participate in item writing and reviews.

4. Increase research on ELL test processing and feedback on tests

Demographic data is collected on students who take the test, such as first language, length of time in an English language development program, and ethnicity. However, collecting some of this information is somewhat cumbersome for WIDA since different states define various demographic indicators (e.g. first language or ethnicity) in different ways. Investigation into student language development growth patterns is being conducted by WIDA staff (e.g. Cook, 2008). Research specifically linking a range of subsets of students would be an excellent expansion of this important work.

5. Recognize the centrality of test taker feedback and response in testing processes

Feedback from test takers is collected using ‘cognitive labs’ on the writing and speaking items during two rounds of ‘try outs’ prior to field testing and as part of WIDA’s field testing process. This is an excellent start; it is recommended that this sort of test taker feedback be collected for listening and reading items, as well. Further, data regarding test impact should be collected to validate the ‘Reaching’ level. A follow-up monitoring and support process should be elaborated based on feedback from former *ACCESS for ELLs*[®] test takers and their teachers. An on-going feedback process would provide essential test validation evidence and support the validity of inferences drawn from test performance.

6. Norm tests and report scores in relation to different student populations

Score reports are tailored in relation to different stakeholder populations. As stated above, different states’ unique definitions of first language or ethnicity impact the possibility of addressing this consideration. However, it is recommended that, given the large number of ELLs who take *ACCESS for ELLs*[®], score reporting should be expanded to address the texture of the ELL population served by WIDA. For instance, given the large number of ELLs who take the *ACCESS for ELLs*[®] annually, normative data related to demographic descriptors could be provided (e.g. relationships between number of years in English language development programming and scores in different language domains). Such data would be a very valuable addition to the educational research literature in terms of understanding ELLs’ learning trajectories.

7. Develop ELL-specific test preparation support as part of the test development process

As discussed above, warm-up items and tasks are included in each section of the test. This is a noteworthy strength of the test. However, the WIDA website does not appear to offer specific preparatory materials (beyond sample items/tasks and guidance for test

administrators), which teachers might use in familiarizing ELLs with the experience of taking the test. Typically, high-stakes test developers provide for such experience by publishing a sample test. Provision of such support would expand the test's fairness, as defined by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999).

The WIDA leadership points out that the provision of sample tests has been purposefully avoided in an effort to ensure that educators are not *teaching to the test*. Still, there is no substitute for *the experience* of taking a high-stakes test, and ELLs may need this experience with test booklets, time limits, rubrics and tasks if they are to demonstrate what they know and can do on a 'live' test (Fox & Cheng, 2007). Failing to provide this experience to a test taker is tantamount to asking a tennis player to prepare all of the required skills for tennis without playing a practice match before the tournament game. Like tennis players, test takers may need a practice test in order to give their best performance when it counts. The practice test experience can reveal to test takers, their teachers and test developers alike, features of interaction between test takers and tasks that are not evident by other means. With 840,000 test administrations annually and the high-stakes of the test, the test developer may want to reconsider, out of fairness, the production of sample test booklets. The developer could clearly do so without threatening the security of the test. Given that one third to one-half of the items are retired each year, there is an ever-larger collection of items and tasks that might contribute to sample test booklets – and support learning.

8. Continue to investigate test accommodations

WIDA offers extensive guidance in this regard online, although this guidance is currently being revised. Given that each state is free to set its own guidance regarding testing accommodations, this is somewhat beyond the control of the WIDA Consortium. However, on-going research, based on test taker performance should continue to inform WIDA's advice in this regard.

Notes

1. The stakes for the *ACCESS for ELLs*[®] are, indeed, high; the test results are used for making decisions about student English language development programming and are part of individual states' accountability for the learning of ELLs under NCLB.
2. Interestingly, Arkansas withdrew after the initial development of the testing instruments.
3. Once students master Level 5, they enter Level 6 ('Reaching'), in which they are considered to be proficient in social and academic English.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Boals, T. (2008). Foreword: IDAACCESS for ELLs[®] Sample Items. *ACCESS for ELLs[®] Listening, Reading, Writing, and Speaking: Sample Items, 2008*. Madison, WI: WIDA Consortium.

- Clarke, S. (1998). *Targeting assessment in the primary classroom*. Bristol, UK: Hodder & Stoughton.
- Cook, H. G. (2008, June). Effects of native language and context factors on the learning trajectory of English language learners: A growth model approach. Retrieved December 11, 2009, from www.ccsso.org/content/PDFs/106_Cook.pdf.
- Fairbairn, S., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice*, 28(1), 10–24.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first (L1) and second (L2) language test takers. *Assessment in Education*, 14(1), 9–26.
- Gottlieb, M. (2004). *WIDA English Language Proficiency Standards, Kindergarten to Grade 12: Overview*. Madison: State of Wisconsin.
- Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, CA: Sage.
- Gottlieb, M., Cranley, M. E., & Cammilleri, A. (2007). *Understanding the WIDA English language proficiency standards: A resource guide*. Madison, WI: WIDA Consortium.
- Gottlieb, M., & Kenyon, D. (2006). *History of research intensive test development: The Bridge Study between Tests of English Language Proficiency and ACCESS for ELLs*. WIDA Consortium Technical Report No. 2.
- Kenyon, D. M. (2006). *Development and Field Test of ACCESS for ELLs®*. WIDA Consortium Technical Report No. 1.
- Kenyon, D. M., Ryu, J. R., & MacGregor, D. (2007). *Setting Grade Level Cut Scores for ACCESS for ELLs®*. WIDA Consortium Technical Report No. 4.
- Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351–372.
- WIDA Consortium. (2007). *WIDA English Language Proficiency Standards and Resource Guide: Prekindergarten-Grade 12*. Madison, WI: WIDA Consortium.
- WIDA Consortium. (2010). *Assessing Comprehension and Communication in English State to State for English Language Learners: ACCESS for ELLs® Interpretive Guide for Score Reports*. Madison, WI: WIDA Consortium.