

AccELLerate!

The quarterly newsletter of the National Clearinghouse for English Language Acquisition

In this issue of *AccELLerate!*

- Cook Formative Assessment: FLARE Project (page 14)
- Logan-Terry & Wright Making Thinking Visible: An Analysis of ELLs' Interactions With Access-based Science Assessment Items (page 11)
- Rasmussen Issues in the Assessment of English Language Learners (page 2)
- Wilde What Makes a "Good" Assessment? (page 18)
- Willner & Rivera State Practices for Ensuring Meaningful Participation of ELLs in State Mathematics, Reading/Language Arts, and Science Assessments (page 5)
- Wolf & Martiniello Validity and Fairness of Assessments for ELLs: The Issue of Language Demands in Content Assessments (page 9)

Also in this issue

- Information Pieces: Assessing ELLs with Learning Disabilities (page 7), Assessment for ELLs in ESEA: Understandings and Misunderstandings (page 8), NCELA Webinars on Assessment (page 6), and Upcoming Conferences (page 20)
- Teachers' Gems of Wisdom: Wells (page 16) and Whiting (page 20), and Sharing Success: Gould (page 17) and Yancy (page 19)
- askNCELA's Inbox (page 10)
- Editor's Notes (page 2)

Assessment of English Language Learners: I

Welcome to the summer issue of *AccELLerate!* The topic is assessment, an essential component of effective instruction that assists educators in making decisions about the initial placement of students in instructional programs or advancing them to next levels, in identifying their needs, and ensuring that they meet educational goals. Federal law requires that all students, including ELLs, be assessed in a valid and reliable manner. Although we all know how challenging this mandate for accountability and effective assessment has been, recent years have seen great progress in developing and using effective and appropriate assessments for ELLs. The current issue of *AccELLerate!* testifies to that fact by bringing together articles that deal with this important topic from a variety of perspectives. In fact, we had so many submissions for this newsletter that we will feature assessment in the next issue as well.

This newsletter opens with an overview of issues, current policies, and practices in the assessment of ELLs (Rasmussen), followed by a study conducted to understand state practices for ensuring ELL inclusion and accommodation in state content assessment (Willner & Rivera), and a clarification of common misunderstandings of the legislation (Rasmussen). Wolf and Martiniello identify specific language features that impact ELLs' test performance, and Logan-Terry and Wright describe promising practices for accommodating ELLs on content assessments. A discussion of the importance of ongoing formative assessment (Cook) introduces a number of stories in which teachers share their practices of implementing this approach in the classroom, and Wilde concludes the issue by providing a summary of the technical qualities of a "good" assessment for ELLs. Have a restful summer!

The National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs (NCELA) is funded by the U.S. Department of Education's Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students (OELA) and is operated under contract No. ED-04-CO-0094 by The George Washington University, Graduate School of Education and Human Development. The contents of this publication do not necessarily reflect the views or policies of the U.S. Department of Education, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This material is located in the public domain and is freely reproducible. NCELA requests that proper credit be given in the event of reproduction.

National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs (NCELA)
2011 Eye St, NW, Suite 300
Washington, DC 20006
Phone: 202.467.0867 or 1.800.321.6223
Fax: 202.467.4283 or 1.800.531.9347
E-mail: askncela@gwu.edu
www.ncela.gwu.edu

Issues in the Assessment of English Language Learners

Mari B. Rasmussen

Introduction

As the number of ELLs in our nation's schools increases, more attention is focused on their success. Accounting for their achievement is a major national priority. This article provides an overview of the complex issues involved in the assessment of ELLs, including current policies and practices. Both academic achievement and ELP assessment are addressed, but the focus is on ELP assessment.

Assessing academic achievement

Historically, ELLs have not succeeded well academically. They have performed less well than their peers on tests in math and English language arts and have been excluded from testing programs because of low scores. In response to the lack of data and low success rate, inclusion policies have been developed to ensure that state, local, and national achievement data are available for ELLs.

Including ELLs in assessment programs is a positive step since it provides information, but it also creates new issues. Most content-area assessments require high levels of English language skills. The focus may be math, science, or English language arts, but the students' facility with the language of the assessment (English) often confounds their ability to show what they know and understand [1].

Accommodating for language is a way to allow ELLs to demonstrate their content knowledge, yet accommodations can be problematic. Research on accommodations has focused primarily on students with disabilities, so there is limited information on accommodations for ELLs. Other challenges include the difficulties of accommodating students with various levels of proficiency as well as allowing for a variety of native languages and differing educational backgrounds [2].

Assessing ELP

Assessing ELP is important since it can assist educators in developing programs and reporting students' growth. However, creating ELP assessments also can be complicated, as they must be appropriate for students of different cultural, ethnic, social, and educational backgrounds and are assumed to be able to predict how well a student will do in academic classes [3].

Understanding ELP assessment also involves an understanding of language proficiency. An ELP assessment should be based on a defensible theory of language proficiency [4]; yet theories of language development and proficiency vary widely. The focus on academic success for ELLs has contributed to the development of theories or definitions of proficiency that include language skills that allow for academic success.

Editor's Notes

The following signs and abbreviations are used in the issue.



— 'Success stories' describe a successful project or an instructional approach.



— 'Teachers' Gems of Wisdom' feature teachers' professional insights or strategies.



— Information pieces

ELL, EL, or LEP—English-language learners, English learners, or limited English proficient—all refer to the same subgroup of students.

ELP—English language proficiency

ESEA—Elementary and Secondary Education Act

ESL—English as a Second Language

ESOL—English for Speakers of Other Languages

OELA—Office of English Language Acquisition, U.S. Department of Education

PD—Professional development

SEA—State education agency

USDE or ED—U.S. Department of Education.

Also, in order to improve the readability of the newsletter and maximize the use of space, we are implementing a new system for citing references. Citations in the text are in [bracketed numbers]. The reference list is in this same numerical order. Other notes are denoted by consecutively numbered superscripts. Let us know what you think!

A number of researchers [5-7] have developed definitions that identify academic English language proficiency, involving unique grammar structures and terminology used for abstract concepts and context-reduced language, as opposed to basic social language that emphasizes interpersonal communication skills.

Other researchers (e.g., 8), who believe that social language also can be abstract and complex, have developed a definition of language proficiency that not only encompasses the components of academic language, but also includes the complexity and sophistication of social situations. Bailey reminds that "schools are simply one more social context or 'community' and that language is always 'a social instrument'" [8, p. 9].

Current Legislative Requirements for ELLs

The ESEA, as reauthorized in 2001, requires a standards-based assessment system. Students must be tested in content-area subjects, and the data must be disaggregated and reported by student groups. It also requires, for the first time, that states establish ELP standards and assessments for ELLs. Table 1 shows how ELLs are included in the two accountability systems [adapted from 9, p. 7].

Current Practices in ELP Assessment

Traditionally, educators have used a variety of methods to assess ELLs. Mainstream classroom assessments have been modified for ESL learners. Standardized, norm-referenced ELP tests developed by publishing companies for ELLs also have

been available, though these tests often have been found to be inadequate.

Fundamental differences among ELP tests resulting in different levels of proficiency for the same student were found in a review of ELP tests 15 years ago. The tests not only yielded questionable results about students' language abilities, but also reflected the most "impoverished model of language testing" [4, p. 12]. To gain a more complete picture of ELLs, educators have used classroom, authentic, and performance assessments [10].

Since the commercial ELP tests had not been based on ELP standards, and did not address academic English appropriately, states found it necessary to develop new assessments to meet federal requirements. These challenges have been significant [11]. Only California had an ELP assessment in place prior to the ESEA requirements that began in 2001. Other large states also developed their own ELP assessments. Many states worked with testing companies to adapt available products. A number of states formed consortia, the most successful of which include:

- the State Collaborative on Assessment and Student Standards for ELL students (ELL-

SCASS), which developed the English Language Development Assessment (ELDA), and the World-Class Instructional Design and Assessment (WIDA) Consortium, which developed the Assessing Comprehension and Communication in English State to State (ACCESS).

More information on the assessments used by various states is available in the document, *English Language Proficiency Assessment in the Nation: Current Status and Future Practice*, edited by Jamal Abedi [12].

Table 2 summarizes some of the types of assessments used with ELLs. These assessments can be either formal or informal. Alternative, authentic, performance, and classroom-based assessments are in the informal category, and often are used to provide formative information—such as periodically using observations and checklists. Formal assessments are administered in a defined manner, and often are used to provide summative information to determine program success or student progress. These categories are not distinct, though. Authentic assessments can be standardized to be administered in a consistent manner. Likewise, some aspects of standardized assess-

Table 1. Federal standards, assessments and annual measurable objective requirements

Students	Standards	Assessment	Annual Measurable Objectives
ELLs Only	ELP Standards	ELP Assessments	ELP Annual Measurable Objectives
All students, including ELLs	<ul style="list-style-type: none"> • Academic Content Standards • Academic Achievement Standards 	Academic Assessments	Annual Measurable Objectives in content areas of math and reading/ language arts

ments can be used informally, though validity and reliability issues come to play.

Summary

There has been progress made in the assessment of ELLs. Educators currently have a variety of assessment information available that allows them to track students, determine proficiency, and evaluate educational programs. The new assessments developed after the 2001 ESEA reauthorization not only meet requirements for accountability, but can provide better information for instructional provision and intervention [13].

Reviewers have found the new assessments to be an improvement, but they recommend more work. There needs to be rigorous analysis on an ongoing basis and partnerships between test developers, states, and researchers [14]. Reviewers also have found that, though the new tests are more comprehensive than the previous

ones, and claim to measure both academic and social language, in practice they tend to overly reflect an academic language construct [15]. Clearly the assessment of ELLs will continue to be important as researchers and policymakers discuss practices, and educators search for tools that assist them. With the approach of another reauthorization of ESEA, the tempo of this discussion will certainly increase.

References

1. Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., et al. (2008a). *Providing Validity Evidence to Improve the Assessment of English Language Learners*. (CRESST Report 738). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
2. Rivera, C., Acosta, B. & Willner, L. S. (2008). *Guide for Refining State Assessment Policies for Accommodating English Language Learners*. Arlington VA: The George Washington University.
3. Wilde, J. (2004). *Definitions for the No Child Left Behind Act of 2001: Assessment*. Washington, DC: The George Washington University, National Clearinghouse for English Language Acquisition. Retrieved April 14, 2010 from http://www.ncela.gwu.edu/files/rcd/BE021263/Definitions_for_the_NCLB.pdf.
4. Del Vecchio, A. & Guerrero, M. (1995). *Handbook of English Language Proficiency Tests*. Albuquerque, NM: Evaluation Assistance Center—West, New Mexico Highlands University.
5. Cummins, J. (2004). *BICS and CALP Explained*. Retrieved April 20, 2010 from <http://www.iteachilearn.com/cummins/bicscalp.html>.
6. Gottlieb, M. (2006). *Assessing English Language Learners: Bridges From Language Proficiency to Academic Achievement*. Thousand Oaks, CA: Sage Press.
7. Scarcella, R. (2003). *Academic English: A Conceptual Framework*. (Technical Report 2003-1). Irvine, CA: The University of California Linguistic Minority Research Institute.
8. Bailey, A. L. (2007). Introduction: Teaching and Assessing Students Learning English in School. In A. L. Bailey (Ed.) *Language Demands of Students Learning English in School: Putting academic language to the test* (pp. 1–26). New Haven, CT: Yale University.
9. U.S. Department of Education, Office of English Language Acquisition (2003). *Final Non-Regulatory Guidance on the Title III State Formula Grant Program—Standards, Assessments, and Accountability*. Washington, DC: U.S. Department of Education.
10. O'Malley, J. M. & Pierce, L. V. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. New York, NY: Longman.
11. Abedi, J. (2007). English Language Proficiency Assessment and Accountability under NCLB Title III: An Overview. In J. Abedi (Ed.), *English Language Proficiency Assessment in the Nation: Current Status and Future Practice* (pp. 3–10). Davis, CA: University of California, School of Education.
12. Abedi, J. (Ed.) (2007). *English Language Proficiency Assessment in the Nation: Current Status and Future Practice* (pp. 3–10). Davis, CA: University of California, School of Education.
13. Albers, C. A., Kenyon, D. M., & Boals, T.J. (2009). Measures for Determining English Language Proficiency and the

Table 2. Assessments used with ELLs

Informal	Formal	
Alternative, Authentic, and Classroom-based Assessments	Standardized, Commercial Assessments	Large Scale, Standards-based Assessments
<ul style="list-style-type: none"> • Anecdotal notes • Checklists • Games • Guided reading • Observations • Parent information • Portfolios • Rating scales • Running Record • Student Oral Language Observation Matrix • Student self-assessment 	<ul style="list-style-type: none"> • Assessments related to commercial instructional programs • Commercial content tests • Commercial ELP tests 	<ul style="list-style-type: none"> • State ELP tests • State academic achievement assessments

- Resulting Implications for Instructional Provision and Intervention. *Assessment for Effective Intervention*, 55 (2), 74-85.
14. Porter, S. G., & Vega, J. (2007). Appendix A: Overview of Existing English Language Proficiency Tests: Full Version. In J. Abedi (Ed.), *English Language Proficiency Assessment in the Nation: Current Status and Future Practice* (pp. 133-190). Davis, CA: University of California, School of Education.
15. Wolf, M. K., Kao, J., Herman, J., Bachman, L. F., Bailey, A., Bachman, P. L. et al. (2008b). *Issues in Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses: Practice Review (Part 2 of 3)*. (CRESST Report 731). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Mari Rasmussen, Ph.D., is the assistant director for state outreach at NCELA. E-mail: mbjr@gwu.edu.*

State Practices for Ensuring Meaningful Participation of ELLs in State Mathematics, Reading/Language Arts, and Science Assessments

Lynn Shafer Willner and Charlene Rivera

There is little information about SEA implementation practices with regard to including and accommodating ELLs in state mathematics, science, and reading/language arts assessments. SEAs vary widely in their criteria for including ELLs in state assessments, basing inclusion decisions on students' English proficiency level, recent arrival to the U.S., or other individual factors [1].

SEAs also vary widely in their practices for collecting data on the implementation of ELL accommodations. A GW-CEEE March 2009 survey of SEAs regarding their ELL accommodation monitoring practices revealed that only 37% of the 49 SEAs that responded reported regularly and systematically collecting data to examine implementation of accommodations for ELLs. Of these SEAs, still fewer (27%) reported analyzing these data. A 2006 United States Government Accountability Office report [2] suggests that part of the reason there is little or no SEA data collected on the implementation of ELL accommodation is due

to the fact that SEAs are not required to report ELL accommodation rates in their USDE Consolidated State Performance Reports; however, SEAs are required to report the rate of accommodation for students with disabilities.

This article reports on a study conducted to understand more about SEA practices for ensuring ELL inclusion and accommodation in state content assessments [3]. The data were taken from ED feedback given to SEAs as part of state standards and assessment peer review and the Title I monitoring review. Peer review provides feedback to each SEA on the adequacy of its standards and assessment system in reading/language arts, mathematics, and science [4]. Title I monitoring reviews conducted by ED examine the implementation of state standards and assessments systems to ensure SEA compliance with the *Assessment and Accountability* provisions in Title I, Part A [5].

Two research questions guided this study:

1. To what extent do SEAs receive feedback in peer review decision letters and/or Title I monitoring reports that addresses the inclusion and accommodation of ELLs?, and
2. What issues are identified for SEAs in peer review decision letters and Title I monitoring reports related to the inclusion and accommodation of ELLs?

In the investigation, the GW-CEEE research team downloaded and reviewed from the ED Web site² 273 peer review decision letters issued to all 50 SEAs plus the District of Columbia and Puerto Rico from June 17, 2005 to January 15, 2009 and the Title I monitoring reports issued to all 52 SEA-entities during the most recently-completed three-year monitoring cycle of 2006-2007, 2007-2008, and 2008-2009. Data pertaining to ELLs were sorted and counted according to the peer review components or Title I indicators prescribed for each type of review. Next, inductive analysis was used to generate the inclusion



NCELA Webinars on Assessment

NCELA continues to conduct webinars on assessment of ELLs. Currently, six such webinars are archived at <http://www.ncela.gwu.edu/webinars/>.

- Assessing K-2 English Language Proficiency: Principles, Purposes and Practices
- Meaningfully Assessing English Learners in Local and Statewide Academic Assessments: What Does It Entail?
- Participation and Performance of English Learners in the National Assessment of Educational Progress (NAEP)
- Test Translation for ELLs: Basic Concepts and Methods
- Test Development and Item Writing for Teachers of ELLs
- What Does the Research Say about Curriculum and Assessment for Young Dual Language Learners?

As we complete more webinars, they also will be archived. Watch for them!

and accommodation issues related to ELLs in state assessments.

The analysis indicates that roughly half of SEAs received feedback concerning ELL inclusion practices in Title I monitoring reports and peer-review decision letters. While SEAs were more likely to receive feedback on ELL accommodation practices in peer-review decision letters than in Title I monitoring reports, almost three-quarters of SEAs received requests for evidence or recommendations about the accommodation of ELLs in peer-review decision letters, compared with one-quarter of SEAs receiving accommodations feedback in Title I monitoring reports.

SEAs received feedback for each of the issues addressed in the areas of ELL inclusion and accommodations. As seen in the lists below, the number of SEAs receiving feedback varied greatly.

Feedback to states regarding ELL inclusion:

- Methods used to count ELLs for AYP (27 SEAs),

- How ELLs were defined in state policy for inclusion in assessments (18 SEAs),
- SEA monitoring of LEA inclusion practices (11 SEAs), and
- SEA reporting in state or district report cards of the participation of ELLs in state assessments (21 SEAs).

Feedback to states regarding ELL accommodations:

- Validity and meaningfulness of accommodated ELL scores (32 SEAs),
- Practices for monitoring the implementation of ELL accommodations (24 SEAs),
- ELL accommodations allowed in state policy (13 SEAs), and
- Alignment of assessment accommodations and instructional practices (3 SEAs).

Two findings emerged from this study. The primary finding is that issues identified in ED feedback to SEAs from both peer review and Title I monitoring revealed that the majority of SEAs have weaknesses in their policies and practices for including and accommodating

ELLs in state assessment systems. Secondly, the feedback itself is inconsistent both within and across the peer review and Title I monitoring processes. The inclusion and accommodations issues addressed in peer-review letters and Title I monitoring reports indicated a need for SEAs to reassess their policies and to find methods to implement them uniformly across a state. The wide range of ELL inclusion and accommodation feedback and the lack of similarity and detail in the feedback also suggests that SEAs are submitting different kinds of information that do not lead to coherent or consistent responses from the two review processes. This inconsistency suggests a need for ED to clarify for SEAs what is considered acceptable evidence.

The findings also have important implications for the two review processes. The inconsistencies within and across decision letters and Title I monitoring reports point to the need for greater alignment between peer review and Title I monitoring. Also, there is a need to select reviewers familiar with the assessment of ELLs and to train review teams so all members have a common understanding of issues related to the inclusion and accommodation of ELLs, even if it means increasing the size of the team.

In conclusion, because SEAs' standards and assessment systems are at different stages of development and different levels of complexity, this analysis points to the need to improve the coherence, quality, and effectiveness of both SEA and

ED practices. Efforts within ED to align the two suggestions for improving the coherence, quality, and effectiveness of both SEA and ED practices. Efforts within ED to align the two processes and to ensure reviewers are familiar with the assessment of ELLs could lead to improved compliance on the part of SEAs, more coherent SEA plans, and ultimately more meaningful inclusion and accommodation of ELLs in state assessment.

Notes

¹ The George Washington University Center for Equity and Excellence in Education

² ED Web site is <http://www.ed.gov>

References

1. Wolf, K. M., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., et al. (2008). *Issues in assessing English*

Language Learners: English language proficiency measures and accommodation uses—Practice review. CSE Report 732. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Available: www.cse.ucla.edu/products/reports/R737.pdf.

2. United States Government Accountability Office. (2006). *No Child Left Behind Act: Assistance from education could help states better measure progress of students with limited English proficiency*. Washington, DC: United States Government Accountability Office. Retrieved September 21, 2006, from <http://www.gao.gov/new.items/d06815.pdf>.

3. Shafer Willner, L., Rivera, C., & Acosta, B. (2010). *Examination of peer review and Title I monitoring feedback regarding the inclusion and accommodation of English language learners in state content assessments*. Arlington, VA: GW-CEEE. Available: <http://ceee.gwu.edu>.

4. U. S. Department of Education. (2009a). *Standards and assessment peer review (guidance and decision letters)*. Retrieved September 15, 2009, from <http://www2.ed.gov/admins/lead/account/saa.html#letters>.

5. U.S. Department of Education. (2009b). *Title I guidance and monitoring reports*. Retrieved February 9, 2010, from <http://www.ed.gov/admins/lead/account/monitoring/index.html>.

Lynn Shafer Willner, Ph.D., is a senior research scientist and Charlene Rivera, Ph.D., is the executive director of The George Washington University Center for Equity and Excellence in Education. E-mails: LWillner@ceee.gwu.edu and CRivera@ceee.gwu.edu.



Assessing ELLs with Learning Disabilities

ELLs have the double challenge of having to learn a new language and new content in the classroom. This challenge can be tripled for those children who happen to have learning disabilities (LDs). The issues surrounding identification, remediation, and assessment of these children can be very complex.

Ideally, these students should be given assessments both in the native language and in English, known as a "dual-language assessment." These assessments give a comparison of results that may show whether the child's difficulties reach across both languages or are an issue of acquiring a second language. Difficulties in both languages may be evidence of an LD, whereas difficulties in one language only (especially if it is in English) may be evidence of normal language development. A comprehensive dual-language, special-education assessment should encompass oral language, reading, writing, and math. There are several factors that need to be taken into account. It is vital that the assessment be (1) culturally sensitive and not culturally biased, (2) accurate in its measurement of the skill it is assessing, (3) reliable and valid, and (4) administered by someone who is qualified and trained to do so (and, if possible, proficient in L1 and L2).

Distinguishing normal second-language development from a disability is incredibly complex, and researchers are working to find a better understanding of the overlapping issues involved.

You may find more information on ELLs with LD at:

http://www.ncela.gwu.edu/files/uploads/17/Accelerate_1_3.pdf

http://www.ldonline.org/spearswerling/Learning_Disabilities_in_English_Language_Learners.

Submitted by Second Language Testing, Inc.

Website: <http://www.2lti.com/home2.htm>.



Assessment for ELLs in ESEA: Understandings and Misunderstandings

The ESEA addresses the assessment of ELLs in a number of areas, many of which are misunderstood. Congress now is considering a new reauthorization effort. In order to help readers to consider the suggested changes in the legislation for ELLs, we offer these brief explanations, with citations for learning more, about the current requirements. The following questions come from our askNCELA inbox and address areas of common misunderstanding.

Can ELLs be exempted from state assessment programs?

No child may be "left behind" in the academic achievement assessment program. There are no exemptions. States are allowed to establish a "flexibility" policy, allowing ELLs to take the ELP test in lieu of reading/English-language arts for one year. For more information: <http://www2.ed.gov/policy/elsec/guid/lepguidance.doc>.

What is a "subgroup"?

In the spirit of including every student in the assessment program, ESEA requires that states and school districts disaggregate the testing results by groups of students who have some similar characteristics (e.g., ELLs), many of whom historically have underachieved. The four areas that the law requires separate data for include students from different racial backgrounds, students from lower socioeconomic backgrounds, students with disabilities, and students with limited English proficiency. Each group must meet goals for making progress. ESEA requires that states and districts report on the progress of all students and subgroups in a "Report Card." For more information on report cards and subgroups: <http://www2.ed.gov/programs/titleiparta/reportcardsguidance.doc>.

What is "AYP" and how does it impact ELLs?

ESEA requires that students in each subgroup make "adequate yearly progress" (AYP) toward academic achievement goals every year. States establish their own goals and formulas for progress, and then are expected to meet them. For more information: <http://www2.ed.gov/nclb/accountability/ayp/edpicks.html?src=ln>.

Do ELLs leave the LEP subgroup when they become proficient in English?

ELLs must be exited from language support services when they reach English-language proficiency. States may count them for two more years in the LEP subgroup for purposes of calculating whether the group has made AYP. For more information: <http://www2.ed.gov/policy/elsec/guid/lepguidance.doc>.

Who decides if a student is an ELL?

For the broad definition of who may be considered as an ELL, or LEP student, as they are referred to in federal law, can be found in Title IX of ESEA: <http://www2.ed.gov/policy/elsec/leg/esea02/pg107.html>. Within this context, states and local districts define proficiency, including the criteria by which a student exits ELL status. Many states have developed guidance on identification and placement criteria for ELLs.

What are the annual measurable achievement objectives (AMAOs)?

Whereas Title I of ESEA addresses academic achievement assessment and includes AYP, Title III addresses English-language proficiency and includes requirements for students meeting objectives for progress in ELP, along with academic achievement. The AMAOs address three areas of student progress towards proficiency, attainment of proficiency, and academic achievement (AYP of Title I). For more information: <http://www2.ed.gov/policy/elsec/guid/lepguidance.doc>.

What is "peer review" and how does it impact ELLs?

The academic achievement standards and assessment program in each state is reviewed by a group of individuals organized by the USDE. The individuals, who act anonymously, come from various areas of education, including SEAs, and therefore are considered peers. They look at all aspects of the assessment program, including issues such as accommodations, inclusion, validity, reliability, and other areas that involve ELLs. The findings of the peer-review teams are summarized in decision letters sent to states by USDE. For more information on peer review: <http://www2.ed.gov/admins/lead/account/peerreview/index.html>; for decision letters sent to states: <http://www2.ed.gov/admins/lead/account/letters/index.html>.

Are the state ELP assessment systems peer reviewed?

No. The current ESEA does not address peer review for ELP assessment. The USDE recommends that the next version of ESEA include peer review for the ELP assessments. For more information: <http://www2.ed.gov/policy/elsec/leg/blueprint/english-learners-diverse-learners.pdf>.

Prepared by Mari Rasmussen, Ph.D., the assistant director for state outreach, NCELA. E-mail: mbjr@gwu.edu.

Validity and Fairness of Assessments for ELLs: The Issue of Language Demands in Content Assessments

Mikyung Kim Wolf and Maria Martiniello

Introduction

How adequately the scores from content assessments reflect ELLs' knowledge and skills has been a central validity question in assessment. For example, when a math test is administered in English, test scores may be not only a function of ELLs' math knowledge, but also of their ability to understand the language of the test. For ELLs who are still developing English-language proficiency, a significant validity and fairness issue is whether test items contain unnecessary linguistic complexity.

In this article we summarize our recent empirical studies investigating the test items that functioned differentially for ELLs, compared to non-ELLs. These studies not only attempted to identify such items in the accountability tests, but aimed to reveal specific language features that impact ELLs' test performance. Based on the findings, we offer practical recommendations for the assessment and instruction of ELLs.

Research findings

First, one may question what kind of linguistic complexity is present in high-stakes, large-scale content assessments. Two recent studies by Wolf [1] and Wolf et al. [2] analyzed the linguistic complexity of three states' math and science assessments in grades 4, 5, 7, and 8. The analytic categories included item length (the total number of

words, the number of unique words, lexical density), academic vocabulary (general academic and technical words), grammatical features (passive voice, long noun phrases using relative pronoun clauses, conditional structures, modals, and nominalization), discourse features (reference, substitution, adversative, causal, temporal, lexical cohesion), and sentence structure. Additionally, the coding scheme for the items included three holistic rating scales for the amount of language versus non-language, the amount of language to process in order to solve an item, and the amount of language in visual images such as graphs, tables, and pictures. The results indicated that the use of various linguistic features generally was limited, as the test language had a conventional script (e.g., *wh*-questions). Yet, a notable finding was that the linguistic complexity and the language demands were varied across states, even within the same content area. For example, the average number of unique words per item in grade 4 was 12, 16, and 18 in the three states' math assessments, respectively. The average number of academic words per item also yielded discrepancies (1.8 words in one state, 4.7 words in another state).

Our studies have found a statistically significant association between the linguistic characteristics

of the items and ELLs' differential performance [3,4,5,6]. Martiniello [3,4] investigated the relationship between the linguistic complexity of items and differential item functioning (DIF) for ELLs and non-ELLs in a math assessment. She examined the linguistic characteristics of math items showing large DIF, that is, showing large difficulty differences for ELLs compared to non-ELLs with equal math proficiency. Textual analysis and students' think-aloud responses to the items indicated that linguistic complexity was a source of DIF for ELLs. The identified DIF items disfavoring ELLs shared some common syntactic and lexical features such as long noun phrases, multiple embedded clauses, low-frequency words, and polysemous words.

Martiniello found that the greater the linguistic complexity in math items, the greater the item difficulty for ELLs, when compared to non-ELLs of equal math proficiency [3,5]. However, the effect of linguistic complexity on DIF is lessened when items provide certain non-linguistic visual and symbolic representations that help ELLs make meaning of the text. These are schematic representations rather than pictures. They embody mathematical relationships, either spatial relationships among objects or patterns, or numerical/mathematical relationships through mathematical symbols or algebraic expressions.

Wolf and Leon [6] found that the number of general academic vocabulary words and the holistic rating of the amount of language (vs non-language) were the most significant predictors of items functioning differentially for ELLs in math and science assessments. This relationship was more evident in the relatively easy items that were answered correctly by at least 75% of non-ELL students. That is, easy items with higher language demands tended to function unfavorably for ELL students.

Recommendations: Assessment
Our studies strongly suggest that linguistic complexity explained differentially functioning items for ELLs. Particularly, the presence of general academic words, non-technical words, and low-frequency words more likely disadvantaged ELLs. These types of vocabulary are not the intended content knowledge to be measured in math and science assessments, but rather are part of the unintended construct of language ability.

While it is important to consider the extent to which the language is part of the construct of the assessments, unnecessary linguistic complexity should be avoided in test development stages to enhance test validity and fairness for ELLs. Thus, we call for the

implementation of explicit item-writing guidelines in the development of high-stakes, accountability assessments. In addition, attention should be paid to whether visual images in math assessments are merely decorative or whether they provide mathematically meaningful representations.

Recommendations: Instruction
As noted above, general academic words, not technical academic words, were associated with the characteristics of items disfavoring ELLs. One possible explanation is that ELLs may have more explicit opportunity to learn (OTL) context-specific and technical vocabulary as part of the content to be learned in math and science classes, while they may not be taught explicitly about general academic vocabulary (such as *based on* and *substantial*) during content instruction. Technical vocabulary (such as *square root* and *geothermal*) is an aspect of content knowledge with which the students may be more familiar. This finding highlights the importance of examining ELLs' OTL. Specifically, it may aid in uncovering the ways that ELL students are exposed to, and instructed in, both general and specific academic language. Pedagogically, content teachers should be mindful of explicitly teaching

general academic vocabulary for ELLs.

References

1. Wolf, M. K. (2009, October). *Examining academic language characteristics and language demands of a state's content and ELP assessments*. Draft Deliverable. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
2. Wolf, M. K., Herman, J.L., Kim, J., Abedi, J., Leon, S., Griffin, N., et al. (2008). *Providing validity evidence to improve the assessment of English language learners* (CRESST Technical Report 738). Los Angeles: University of California, CRESST.
3. Martiniello, M. (2007). Linguistic complexity and differential item functioning (DIF) for English language learners (ELL) in math word problems. *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, 68, 2422.
4. Martiniello, M. (2008). Language and the performance of English language learners in math word problems. *Harvard Educational Review*, 78 (2), 333-368.
5. Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, 14 (3&4), 160-179.
6. Wolf, M. K. & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, 14 (3&4), 139-159.

Mikyung Kim Wolf, Ph.D., and Maria Martiniello, Ph.D., are research scientists at the Educational Testing Service (ETS). E-mails: mkwolf@ets.org and mmartiniello@ets.org.



askNCELA's Inbox

in which we highlight the answers to commonly asked questions that appear in our e-mail inbox.

Q: Where can I find information about states' standards and assessment for ELP and for the content areas?

A: You can find links to each state's ELP and content standards and assessments in NCELA's *Title III State Information System* at <http://www.ncela.gwu.edu/t3sis>. For more resources: <http://www.ncela.gwu.edu/assessment/>.

Making Thinking Visible: An Analysis of ELLs' Interactions With Access-based Science Assessment Items

Aubrey Logan-Terry and Laura J. Wright

Introduction

In recent years, the ESEA, as reauthorized in 2001, has required that all students be tested yearly to demonstrate adequate yearly progress, creating the need for new and innovative assessments, especially for ELLs. One such assessment, Obtaining Necessary Parity and Academic Rigor-Science (ONPAR), uses an *access-based framework* [1] and replaces typical linguistically heavy test items with graphics, animations, and other support features to provide ELLs with greater access to content. Using video data of students interacting with ONPAR items in cognitive lab interviews, we examined how students at varying degrees of language proficiency access and process the semantic information of the assessment. We propose that meaningfully including ELLs in large-scale testing requires taking into account the cognitive resources they draw upon when interacting with test items, and that 'one-size-fits-all' approaches to assessment cannot accommodate ELLs adequately.

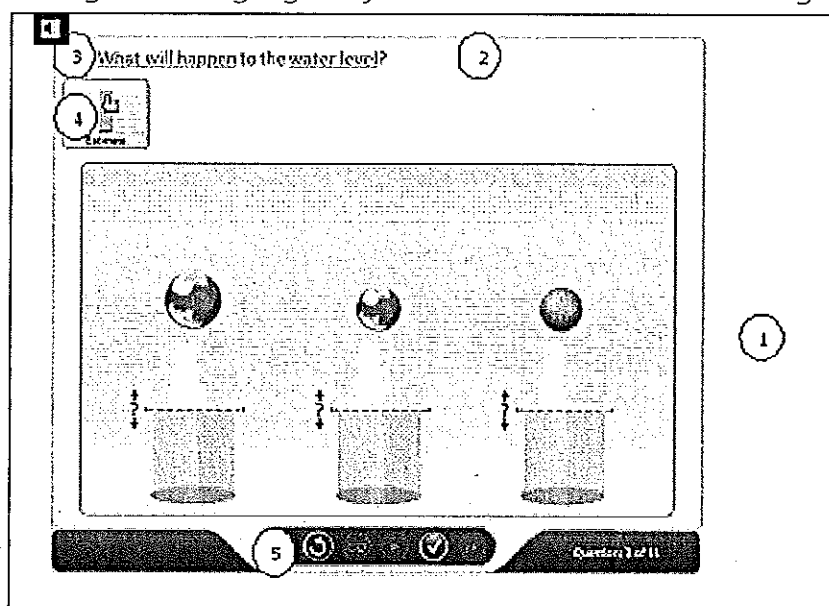
Background

Standardized test scores indicate that a significant achievement gap exists between linguistically diverse students and their native English-speaking peers across all academic areas [2]. Research suggests this may be due to tests' conflation of measuring language proficiency and content knowledge for ELLs

[e.g., 3]. Some proposed accommodations include test translations and/or simplified language, to measure student achievement more accurately. However, there are limitations to these accommodations as valid test translations and consistent simplification of language are difficult to produce and require further study [4,5]. Another promising accommodation approach that has developed recently is an access-based framework in which items are modified to make content more accessible by changing structural and contextual factors. In assessing ELLs, the goal is to minimize the challenges that language may

pose and to provide alternative means to "access meaning, solve problems, and demonstrate solutions without lessening the rigor of the item or changing the construct being measured" [6, p. 8].

ONPAR-Science is a new content assessment based on the access-based framework with items that include: (1) graphics and animations, (2) text prompts with hyperlinked vocabulary, (3) a speaker button that provides an oral Spanish translation of the text prompt, (4) an animated icon to demonstrate how to respond to an item physically, and (5) a navigation bar that allows students to go



1. Graphic: conveys primary semantic information of item
2. Text prompt: conveys task demand
3. Speaker button: provides spoken text prompt translated into Spanish (optional)
4. Icon: demonstrates physical action needed to complete task (optional)
5. Navigation bar: allows students to navigate the assessment at their own pace

Figure 1. An ONPAR-science item

forward and back and replay items (Figure 1). These features are intended to make ONPAR items more accessible to ELLs than traditional paper-and-pencil test items (see [7] for a full description of ONPAR methods go to http://www.onpar.us/sample_items.html to see the working item).

Methods

In order to investigate the ways in which students accessed and processed semantic information in the ONPAR test, 12 cognitive laboratory interviews with Spanish-speaking ELLs with a range of English proficiencies (three beginning, three intermediate, and four exited/bilingual ESOL students) from fourth and eighth grade were analyzed. Students chose the language (Spanish or English) of the approximately 45-minute interview, during which they worked through five or six ONPAR items. The students were asked to explain their answers, and an interviewer asked open-ended questions to clarify responses when needed. Thus, the cognitive laboratory interviews provide insight into comprehension difficulties and reasons for performance variation across items and languages.

The interviews were transcribed using the qualitative analysis tool Atlas.ti, which allowed for a close analysis of responses and quantitative coding across interviews, as well as analysis of visual and textual data. After the data were transcribed, three student strategies for accessing and processing

the semantic meaning of an item were identified: (1) code-switching for scientific terminology, (2) use of the speaker button for native language support, and (3) reliance on graphics as demonstrated by deictic (pointing) gestures. These strategies were coded across the interviews in order to investigate how ONPAR's features afforded ELLs access to the meaning of the test items.

Findings and discussion

Code-switching for scientific terminology

Most students (8 out of 12) chose to be interviewed in English; however, even the four students who chose Spanish as the language of the interview often (at least 25% of the time) code-switched into English when using scientific terminology. Table 1 shows the amount of code-switching into English per student.

Table 1. Percent of occurrence of code-switching into English

4 th grade	Maria	52%
	Sara	62%
8 th grade	Ana	25%
	Sofia	36%

Although these students were more proficient in Spanish than English, their knowledge of scientific terminology in Spanish was limited. This suggests that providing information in students' second language (English) may be essential for accommodating even beginning ELLs.

Native language support

In contrast, some students relied heavily on their native language to access semantic information, even when they had chosen English as the language in which to be interviewed. Students had the option of clicking on a speaker button in order to hear a translation of the text prompt into Spanish. Figure 2 shows the percentage of times students clicked on the speaker button during an interview as a function of the total number of times the students were presented with the option of using the speaker button.

For some students (e.g., Ines, Jose, and Sofia) the speaker button seems to have been an important resource for accessing semantic content of item prompts. This demonstrates how test translation and other native language support tools may be necessary, yet not sufficient (as evidenced by students' code-switching into English for scientific terminology), assessment accommodations for ELLs.

Deictic gesturing at graphics

A final point of access to semantic information in the ONPAR test was graphic information. In order to code how students relied on the graphics in lieu of language, we focused on interactions in which students appeared to understand a graphic, but did not articulate linguistic terminology to describe it. We coded examples of deictic gestures (pointing) in which students gesturally referred to graphic information and used a deictic expression (e.g., a deictic

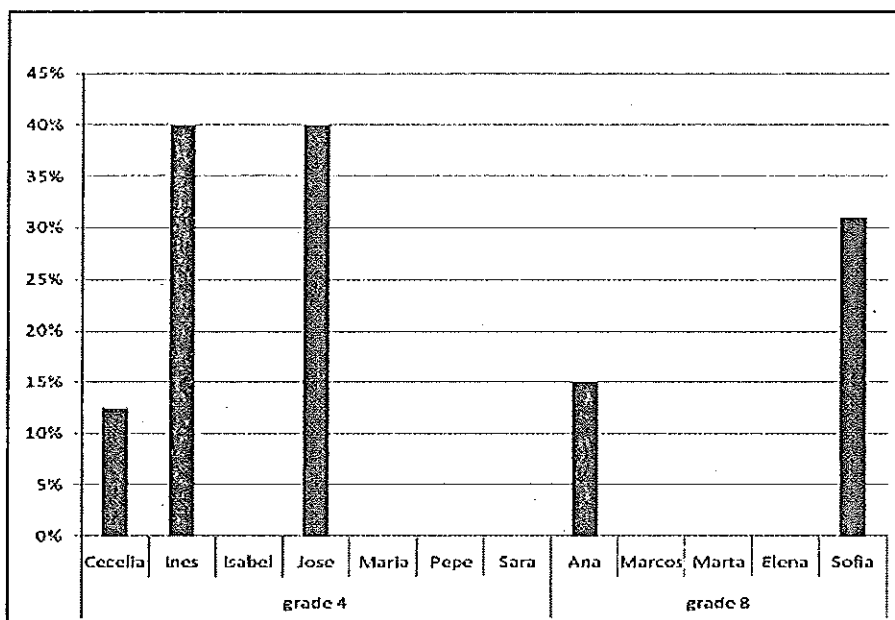


Figure 2. Total usage of audio button

pronoun, *this*, *that*, *these*, or *those*) as a proxy for linguistic terminology. Figure 3 shows the frequency of deictic gesturing per student during the interviews.

Almost all of the students used deictic gesturing as a strategy for managing language while interacting with the test items, with 4th graders using more gestures than 8th graders. This provides evidence that the multimodal features and graphics of ONPAR functioned as meaningful semantic information for students and may have helped them access and process content on the assessment. In this way, integration of visual support tools recommended by the access-based accommodation framework seems to provide students with additional routes to access and process information.

Conclusions and implications

Our findings indicate that each student has a unique profile in

terms of strategy usage: some students (e.g., Sofia) used a variety of strategies to varying degrees in order to access and process content throughout the ONPAR-

Science assessment; other students seemed to rely heavily on certain strategies (e.g., Pepe, who frequently pointed to visuals on the screen yet never utilized the audio translation button). We found that students' strategies seemed to be as diverse as the students themselves; thus, successful accommodations for ELLs on content tests must be diverse as well.

These results demonstrate the promise of an access-based accommodation framework that provides a variety of resources for students as they work through assessment items. Moreover, in the case of ONPAR, the integration of computerized multi-semiotic features may provide particularly useful tools for accommodating ELLs on content assessments.

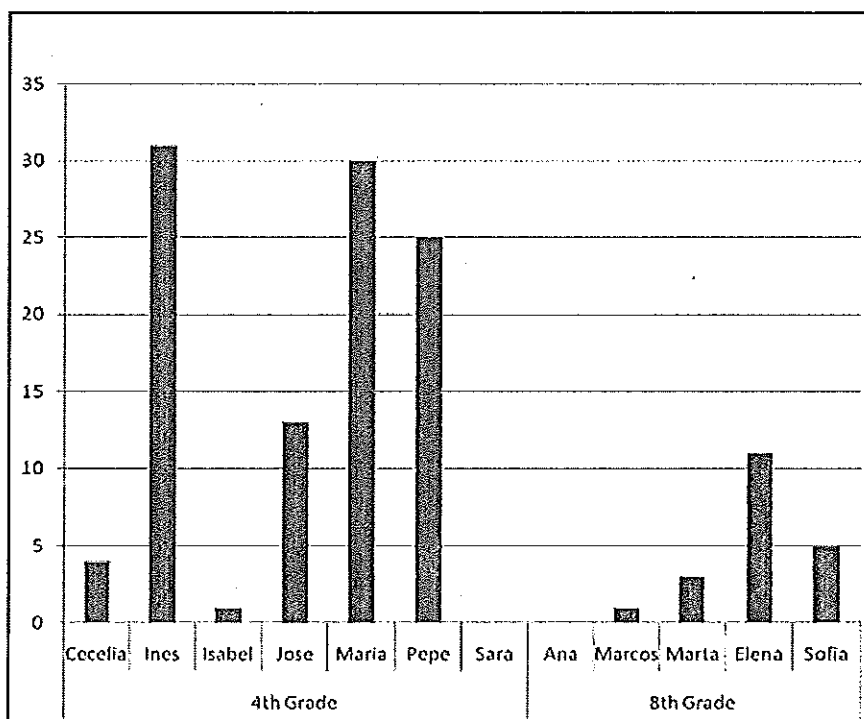


Figure 3. Number of deictic gestures

References

1. Kopriva, R. (2008). *Improving testing for English language learners*. New York: Routledge.
2. Snyder, T. & Dillow, S. (2010). *Digest of Education Statistics 2009*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
3. Abedi, J. & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practices*, 26(5), 36-46.
4. Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy based research. *Review of Educational Research*, 74(1), 1-28.
5. Shafer Willner, L., Rivera, C., & Acosta, B. (2009, November). *An overview of approaches for accommodating English language learners in state content assessments*. Presented at the East Coast Association of Language Testers, Washington, DC.
6. Carr, T. & Kopriva, R. (2007). *Handbook on developing accessible assessment items for English language learners and students with learning disabilities*. Columbia, SC: South Carolina Department of Education.
7. Kopriva, R. (2009). Assessing the skills and abilities in math and science of ELLs with low English proficiency: A promising new method. *AccELLerate!*, 2(1), 7-10.

Aubrey Logan-Terry is a Ph.D. candidate in linguistics at Georgetown University and a consultant for the ONPAR math and science projects at the Center for Applied Linguistics. Laura J. Wright, Ph.D., is an educational linguist who conducts research for the ONPAR math and science projects at the Center for Applied Linguistics. Corresponding author's email: ael34@georgetown.edu.

Formative Assessment: FLARE Project

H. Gary Cook

Introduction

The reauthorization of ESEA in 2001 as the No Child Left Behind Act (NCLB) has led to enormous debate. Many praise its requirements. Many decry its intrusive mandates. It is not difficult to identify points to argue about in this law, but most, if not all, who praise or criticize, acknowledge that previously underserved students, i.e., racial and ethnic minorities, special education students, and ELLs, now matter.

Prior to 2002, few states had assessments for ELLs created from state-adopted English language proficiency standards. Now, all states have assessments that are in some fashion related to state language proficiency standards. Substantial progress has been made in the development of large-scale ELL summative assessments. However, although it has long been accepted that formative (classroom) assessments are important,

they have garnered little attention by the measurement community due to issues of validity and reliability. To address this deficit, a recent issue of *Educational Measurement Issues and Practices*, Volume 22:4 (2003) was dedicated to this subject exclusively. Investigation of classroom assessments is now "[a] cutting edge area of assessment research" [1, p. 150].

Background

Definitions of formative assessment have varied widely [2-6]. We find Brookhart's conception of formative assessment [7] the most useful. Her definition includes three important features:

- Students' focus on meaningful learning goals, supported by exemplars or models of "good work;"
- A mechanism or method of identifying where students are and where they need to be, relative to their learning goals; and

- A system that allows both teachers and students to act upon "gaps" between students' current status and desired learning goals.

Implicit in this definition is the notion of continuous or cyclical measurement. "Assessment needs to be conceptualized as an ongoing activity that involves gathering, interpreting, and evaluating information, and action, based on results, rather than mere documentation of student performance [i.e., measurement]" [8, p. 39]. Good formative assessments need to incorporate all of these elements, and their use "must become an integral part of the learning process" [7, p. 6].

The technical quality of these integral tools, however, often is limited. Teachers' own classroom assessments do not mention technical measurement concepts or principles, but emphasize "fairness

... in relation to providing opportunities for students to demonstrate success" [8, p. 36]. It is important to refine and improve teachers' formative assessments as well as to hone their decision-making skills to use these instruments to promote student learning effectively.

In the current educational context, the development of a useful formative assessment system must integrate with mandated state and district academic standards and assessments, and "[i]t is imperative to investigate ways in which teachers can meet the accountability requirements but at the same time successfully assess for formative purposes in the classroom" [3, p. 255].

Taken together, a good formative assessment system should:

- be of adequate technical quality;
- be an ongoing, classroom-based process that integrates with instruction;
- focus students on learning goals;
- provide examples of good work;
- identify students' current skills and abilities;
- highlight students' gaps in learning goals and provide methods to address gaps;
- integrate seamlessly with external standards and assessments;
- be dynamic enough to accommodate for classroom realities (e.g., easily administered, account for disruptions, adjust to student heterogeneity); and
- incorporate a rigorous PD program for teachers.

The FLARE project

This section briefly describes a project funded by the Carnegie Corporation of New York that seeks to apply the aforementioned elements to create a replicable formative assessment system as well as to contribute to a continuing dialog on how to support ELL student achievement best. The project is called FLARE, or Formative Language Assessment Records for ELLs, and it is designed to assist teachers in measuring student progress as they develop the essential language needed for success in academic classes at middle and high school.

Figure 1 displays FLARE's assessment model based on the current understanding of research on formative assessment. Language Learning Targets are the foundation of this model and are used to support teachers and students in setting language learning goals for instruction and learning. These targets are academic language learning progressions

framed by linguistic components (language functions, vocabulary, grammar, and genre/language discourse) that span four academic disciplines—English-language arts, mathematics, science, and social studies. The heart of this model is its assessments: student self-assessments, an assessment toolbox for teachers, and benchmark tests. All assessments are used in concert to support students' acquisition of academic English, which in turn supports student achievement. Once information about student progress is collected, meaningful feedback can be provided to teachers and students.

The FLARE project is currently in its second year. The language learning targets are being field tested and FLARE's assessments are being developed. The targets and assessments can be adapted to fit into a variety of instructional contexts. They are generative in nature, not static. Thus, one critical lesson learned so far is the need

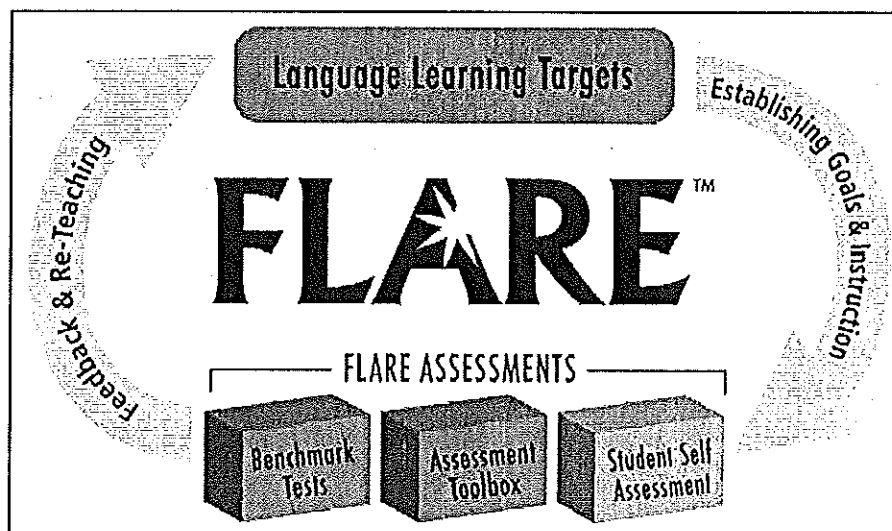


Figure 1. FLARE's assessment model

for deep and sustaining teacher PD and support. Three school districts are participating in the FLARE project: Garden Grove Unified School District, CA; Chicago Public Schools; and Charlotte-Mecklenburg Public Schools, NC. Since each district has a unique ELL context, each district adds to our understanding of how formative assessments can be implemented effectively.

We are just beginning to understand how formative assessments can enhance student learning and support language acquisition and academic achievement for ELLs. For more information on this project and what we're learning, see: www.flareassessment.org.

References

1. McKay, P. (2005). Research into assessment of school-age language learners. *Annual Review of Applied Linguistics*, 25, 243-263.
2. Nitko, A.J. (1989). Designing tests that are integrated with instruction. In Robert L. Linn (Ed.), *Educational Measurement* Third Edition. New York: MacMillan.
3. Oosterhof, A. (1994). *Classroom applications of Educational Measurement*, Second Edition. Edlewood Cliffs, CA: Macmillan.
4. Popham, W.J. (1995). *Classroom Assessment: What teachers need to know*. Needham Heights, MA: Allyn and Bacon.
5. Shepherd, L.A. (2006). Classroom Assessment. In Robert L. Brennen (Ed.), *Educational Measurement*, Fourth Edition. Westport, CT: Praeger.
6. Stiggins, R. (2005, December). From Formative Assessment to Assessment FOR Learning: A Path to Success in Standards-Based Schools. *Phi Delta Kappan*, 87(4), 324-328.
7. Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement Issues and Practices*, 22(4), 5-12.
8. McMillan, J.H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement Issues and Practices*, 22(4), 34-43.

H. Gary Cook, Ph.D. is an associate research scientist at the Wisconsin Center for Education Research.
E-mail: hcook@wisc.edu.



Preassess without the Stress

Pre-assessment is vital when planning upcoming units, and helps provide meaningful and effective instruction for all students. By using pre-assessment effectively, teachers are making the best use of their valuable instructional minutes. Yet students, especially ELLs, can get stressed out by traditional paper-and-pencil pre-assessment. If they do, try a Total Physical Response (TPR) approach. All it takes is some enthusiasm, creativity, and an open mind.

"Walking the Line" is one method for pre-assessing all students in your class simultaneously. Have your students line up parallel to any line in your class (tile or flooring often make up this line, or masking tape, yarn, etc.). Explain to your students that you will be making a series of statements, and that they will not respond verbally. If they agree with the statement, they will "walk the line" or walk up to the line. If they do not agree, or are unsure, they will stay where they are. If they agree a little bit, they may walk halfway. I often begin by using common phrases such as, "I like pizza," as a way for students to become familiar with the process. After that, I will have students model particular concepts using their bodies, such as types of angles (arms, legs, angles of their bodies), mathematical symbols, and strategies they may use in class. It also has been effective to include phrases from prior grades or units that all students know, as a way for all students to feel successful.

As a way to ensure that ELLs are understanding the questions or prompts, I often front load the sentence structure and wording of how I present the tasks. For example, I will begin by discussing the statement, "I feel comfortable working with ____." We will talk about concepts that they are comfortable with (addition, subtraction, 2-D shapes, etc.) and practice saying, "I feel comfortable working with addition." Therefore, when the actual pre-assessment begins, students will know exactly what is being asked. Other useful wordings include, "I know a lot about ____," "I have worked with ____ before," and many others. Almost any phrase will work, as long as it is explained beforehand.

This strategy has proven to be a fun and effective way for students to show their prior knowledge of any subject and at any age.

Submitted by Matthew Wells, math focus teacher at Georgian Forest Elementary School, Silver Spring, MD (Montgomery County Public Schools), and a student in The George Washington University's Bilingual Special Education Certificate Program. E-mail: mwells84@gvmail.gwu.edu.



Incorporating Informal Assessment Techniques

Not particularly a fan of grades and testing, I constantly seek ways to lessen the anxiety and concerns that so often go hand-in-hand with assessing students, particularly ELLs, who are very aware that their performance on assessments frequently does not display what they actually have learned or are capable of. There are a number of ways to make the assessment process less agonizing while at the same time yielding an accurate picture of what our students actually are able to do. The following are a few practical reminders/considerations for making the most of assessments.

Use a Variety of Assessments. Students benefit from opportunities to practice with different kinds of assessments, such as multiple choice, essay, cloze, presentations, etc., and even the language of different types of tests (e.g., math, science). Sheltered instructional environments in particular are usually less threatening for them and build their confidence.

Offer Choices/Options. It is rewarding to watch the levels of anxiety and concern diminish when students are offered a choice of how they wish to be assessed with specific topics. It is even more intriguing to involve them in the process of deciding the criteria by which they should be assessed, thereby eliminating the mystery and fear so often associated with assessment. If teachers are encouraged to offer students choices with classroom activities, then why not do the same when it comes to assessment? The same material assessed in a written essay or journal also can be assessed in an oral presentation, debate, or even a game.

Provide Clear Expectations. Students should know at the onset of a lesson/unit what they are expected to know and demonstrate by the end. Unit/lesson activities should connect to the assessment—in other words, there should be no surprises.

Allow Students to Self-Assess. Students tend to be remarkably on target when assessing their own performance (e.g., using a rubric or scale to indicate learning or performance). It is vitally important that not only I as a teacher know where students are, but that they themselves are aware of their progress or lack thereof. Further, it is amazing how interested students are in their progress and how capable they are of working together with the teacher to take the pulse of their growth and development.

Involve Students in Assessing Each Other. Students are somehow less offended and brutally honest when evaluating one another and they tend not to negotiate how they have been graded. As with self-assessment, in some cases they seem to feel they have received an “easy grade” or a gift, not entirely realizing that they actually have earned the grade received and evaluated each other just as the teacher would, or has. Peer assessment with simultaneous feedback from the teacher (using the same criteria) on assignments can provide useful insight.

The following examples of activities may work well in your classroom.

- Students work together to formulate answers to an oral quiz on a particular topic. They answer questions orally as a class with no one student answering more than one question. This provides a nice picture of how well the class as a whole absorbed the material, and allows students to use their strengths collectively. This scenario may spark discussion if, for example, students disagree on a response.
- After reading a story, small groups are formed and each group is provided with a portion of the story to act out, while the other groups guess which portion of the story is being acted out. For lower-level classes, students are provided with sentence strips/descriptions (and/or pictures) of all of the scenes to refer to. Following this, students write a summary of the story on their own or as a class retell the story to the teacher as he/she records it. Those students who are unable to write a complete summary can arrange pictures in the correct order of the story, match simple sentences with the correct pictures, and then place them in the proper sequential order.
- When introducing a new unit, students are provided with a learning survey to complete. They think about and indicate the things they already know about the unit topic and then those things that they wish to learn. During or at the end of the unit, students revisit the survey and fill in/check off what they have learned and can do. Students share from their logs in class discussion. These learning logs also can serve as a ready-made unit review for an end-of-unit exam.

Submitted by Seantele Gould, M.Ed., currently a research associate at NCELA, also has been a classroom teacher working with ELL students. E-mail: smrg@gwu.edu.

What Makes A "Good" Assessment For ELLs?

Judith Wilde

ESEA, as reauthorized in 2001, calls for ensuring that student assessments are high quality [1], improved (presumably on a regular basis) [2], and are both valid and reliable measures of what students know and can do [3].

Meaningful assessment is essential. To ensure that an assessment is meaningful, the content must be carefully considered—does the content of the assessment match what has been taught, are the items developed in a way that allows students to demonstrate their knowledge, are the items clearly written in a way that is understandable to students, and so on. Underlying the issues related to content are three additional factors: reliability, validity, and fairness—the technical qualities of an assessment. While psychometricians still argue about the relative importance of each of these concepts and what constitutes “good” reliability, validity, and fairness, we can provide some guidance to clarify these test qualities.

Reliability is the stability or consistency of an assessment. For instance, two assessments of a student, performed at the same time, should show similar results; two reviews of a teacher’s qualifications should result in similar conclusions about her/his skills as an instructor. Think of it this way: if you weigh yourself each morning, but each morning the scale varies wildly, then the scale is not a reliable measure of your weight. An instru-

ment *must* be reliable if it is to be used to make decisions about how well a student is performing. As a general rule, the more items on an assessment, the greater the reliability. A test with 50 items will be more reliable than an assessment with 10 items; however, an assessment with 300 items may fatigue the test-takers and be very unreliable.

Reliability is a correlation across items that is measured on a scale from 0.0 to 1.0, with higher numbers being better (i.e., more reliable), although it is virtually impossible to achieve a rating of 0 or 1. If a test is used to make decisions about test-takers, then most psychometricians agree that a reliability coefficient, or value, of at least

- .80 is needed to make decisions about an individual (how well is Maria doing in math?);
- .65 is needed to make decisions about a group of individuals (are the ELL students ready to proceed to the next math unit?); and
- .50 is needed if a test will be used to provide some general information about a group of test-takers (how literate are students in the third year of ELD classes?).

Validity is more difficult to describe, in part because psychometricians continue to refine the definition of validity. Basically, validity asks whether the interpretation, uses, and actions based on assessment results are appropriate. Think of it this way: if you know you weigh

110 pounds, but your scale, day after day, says you weigh 90—the scale is reliable, but not valid.

The communicative competence of learners must be considered when creating a valid test. In addition, the specific purpose of the assessment must be considered. An assessment may be valid for one purpose, but not for another (for instance, an assessment of English language achievement should not be used to measure English language proficiency). Basic questions when considering validity are “Does this test measure what it purports to measure?”, “Do I believe what this test tells me about my learners?”, and “Are the results of this assessment similar to results from other assessments of the same topic?”

Fairness issues are especially important for linguistically and culturally diverse students. These issues include considering the language(s), gender, culture, and overall abilities of the test-takers. For instance, if it is known that a group of test-takers have difficulties responding to a test written in English, then a fair test will include response options that allow the students to create pictures or graphs to show their answers or may allow them to dictate answers to another person.

Fairness is affected by how items are developed, the scoring procedures used, training of scorers, calibration of scores, access to good

good instruction and classroom materials, and so on.

Fairness also should ensure that biases are not evident in the testing procedures or test items. Biases generally fall into three areas:

- item development or scoring procedures that promote or oppose an individual's race/ethnicity, culture, language, or physical ability;
- stereotyping within items or reading passages based on race or ethnicity, language, culture, or physical ability through under- or over-representing or ridiculing certain groups; and
- illustrations that negate the impact of certain individuals, or

groups, typically by not including them.

In many cases, biases can be quite subtle. For instance, if items on an assessment only use names that are typically associated with white students, there is a subtle bias for this group and against others (e.g., Asians, Hispanics) who choose to maintain culturally-appropriate names. These issues can impact students' interest in subject matter as well as their interest in and ability to achieve well on a test.

There are many more issues related to good assessment for all students, and especially for ELL

students. Described here are those that affect the technical qualities of the assessment, and their importance cannot be underestimated when developing an assessment for any group of students, but especially those who may be less resilient, such as ELLs.

References

1. ESEA, as reauthorized in 2001, Title I §1001(1), §1111(b)(3)(A), & §1112(b)(1)(A).
2. ESEA, as reauthorized in 2001, Title III §3115(c)(2)(A).
3. ESEA, as reauthorized in 2001, Title I §1111(b)(2)(D)(1), §1111(b)(3)(C)(ix)(III), & Title III §3121(a)(3)(B).

Judith Wilde, Ph.D., is the executive director of NCELA. E-mail: jwilde@gwu.edu.



Effects of Ongoing Assessment in the Writing Classroom

As a teacher of ESL in Puerto Rico for over 10 years, I encountered much resistance to language learning. This led me to advocate for performance-based assessment and student involvement in language learning. One particularly successful approach was using *ongoing assessment techniques*. I differentiated between *language proficiency* and *content knowledge* when assessing students' writing. Engagement was key, and in the initial stages I accepted responses in either English or Spanish during topic selection and the organization process, as well as in classroom discussion throughout. This allowed me to gauge students' language development from the outset. I also included assignments that looked simple in terms of language skills, but allowed the students to scaffold and reinforce their writing skills (e.g., providing supporting details, writing solid concluding sentences). Taking incremental steps from structured brainstorming worksheets to guided questions in plain language for internet research allowed me not only to assess their organizational skills and logic application, but to assist them in making corrections or focusing their writing throughout the writing process, rather than just in the final stage. When the final stage of assessment arrived, I used rubrics and portfolios as my foundation for grading. Providing rubrics beforehand increased the students' motivation to learn and guided their learning; they knew exactly how the final product would be evaluated and prepared accordingly. In addition, the rubrics allowed me to assess content skills gradually over time at various stages of the project. Each writing task was included in their portfolios to show the depth and breadth of their development and provide documentation of their academic improvement. As students saw their progress from unstructured paragraphs to accurate essays, their confidence in writing increased. There is no better reward for an educator than seeing their students succeed!

Keys to success:

- Allow the organization process and verbal discussion of writing to take place in either English or the native language, so that the focus remains on the *writing*.
- Incorporate technology as much as possible throughout the process to keep students engaged.
- Provide a grading rubric at the beginning of the project so that students always have the learning outcome in mind.
- "Chunk" the project into smaller pieces to enable you to evaluate each step and provide guidance throughout the process rather than just at the end.
- Involve students in the assessment process through peer-editing.
- Use portfolios to provide tangible evidence of growth.

Submitted by Carlos A. Yancy, M.A., an ESL test developer at SLTI. E-mail: cyancy@2lti.com.



Student Podcasts: Oral Assessment and e-Portfolios

A student podcast is a digital audio recording that is created, stored, and retrieved as a file on a computer. Easy-to-use software available from sites such as <http://www.podomatic.com> allows ELLs to control the creation of podcasts, and because students enjoy using technology, they view these activities positively.

Podcasts are easily adaptable for classroom use and they can be created at home. Their flexibility is similar in some respects to writing:

- Podcasting can be controlled or free, of varying lengths, timed and untimed, and can make use of student- or teacher-directed prompts;
- Podcasts can be longer or shorter as needs, abilities, and goals change; and
- Podcast files can be transferred from one teacher to another.

Podcasts provide an authentic, easily scorable, and accessible way to evaluate student speaking abilities in a stand-alone assessment, or as one piece in an e-portfolio of student speech that allows students and teachers to evaluate students' oral language development over time. These e-portfolios are similar to traditional portfolios that have focused on the improvement of students' writing [1]. Depending on the goals, evaluation rubrics could focus on pronunciation, fluency, grammatical accuracy, vocabulary knowledge, or overall speech comprehensibility and complexity. Podcasts can be combined with reading and writing to create assessments of more than one skill and can be used in assessing content knowledge. Teachers can determine how often students are allowed to record and delete their recordings. Finished podcasts can be placed in an electronic dropbox or sent to the teacher as an attachment to an email. The podcasts can be stored and arranged in a folder with other podcasts from that student as part of an e-portfolio.

There has been considerable excitement about the use of podcasts in the language classroom [2, 3], and some school districts already are taking advantage of the possibilities of using iPods and iTunes to enhance language instruction [4]. However, the benefits of using student-created podcasts for oral language assessment have been largely ignored. Their flexibility and inherent attractiveness could be a useful alternative oral assessment tool for ELLs.

References

1. O'Malley, J.M., Valdez Pierce, L. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. New York: Addison-Wesley.
2. Diem, R. (2005). Podcasting: a new way to reach students. *The Language Teacher*, 29(8), 45–46.
3. Rosell-Aguilar, F. (2007). Top of the Pods—In Search of a Podcasting 'Podagogy' for Language Learning. *Computer Assisted Language Learning*, 20(5), 471–492.
4. The Apple Education Profiles series website. Retrieved May 4, 2010 from <http://itunes.apple.com/us/podcast/apple-education/id334296685>.

Submitted by James Whiting, Ph.D., an assistant professor in applied linguistics at Plymouth State University, NH.
E-mail: jcwhiting@plymouth.edu



Upcoming Conferences and Workshops



- Direct Strategies Institutes, Center for Applied Linguistics (CAL)
What's Different About Teaching Reading to Students Learning English?
June 22–24, 2010 and July 20–22, 2010
Savoy Suites Hotel, Washington, D.C.
<http://www.cal.org/solutions/profdev/workshops/strategies.html>
- Second Language Research Forum (SLRF) 2010
Reconsidering SLA Research: Dimensions and Directions
October 14–17, 2010
University of Maryland, College Park, MD
<http://www.webspaces.umd.edu/SLRF2010/>
- East Coast Organization of Language Testers (ECOLT) 2010
Innovations in Language Assessment
October 29–30, 2010
Georgetown University, Washington, DC
<http://www.cal.org/ecolt2010/>