

The Challenge of Assessing Language Proficiency *Aligned* to the Common Core State Standards and Some Possible Solutions

Alison L. Bailey, *University of California, Los Angeles*
Mikyung Kim Wolf, *Educational Testing Service*

For the past decade, there has been a significant change in assessing the English language proficiency (ELP) of English language learners (ELLs). Since the accountability reform efforts of the No Child Left Behind Act (NCLB, 2001), an emphasis has been placed on measuring the kind of language proficiency students presumably need in order to succeed in academic contexts. Title III of NCLB first made the suggestion that the ELP standards that are adopted by each state be “linked” – in an unspecified fashion – to the states’ academic content standards, and that the states’ ELP assessments be aligned to their respective ELP standards. This mandate has been interpreted by many states as calling for a link between ELP standards and English language arts (ELA) standards. Yet, as state ELA standards are varied in their breadth, depth, and emphasis of content, varied ELP standards have been formulated. A general consensus has emerged that ELP assessments should measure students’ academic language proficiency in order to gauge the accessibility of content instruction for ELL students. However, varied ELP standards and definitions of the language proficiency construct represent a major challenge for developing and validating the current generation of ELP assessments (e.g., Wolf, Farnsworth, & Herman, 2008).

We see both promise and new challenges in the assessment of ELP with the advent of the national movement toward Common Core State Standards (CCSS) in Mathematics, ELA and Literacy in History/Social Studies, Science, and Technical Subjects and the Next Generation Science Standards from Achieve Inc. based on the recent framework developed by the National Research Council (2011). Now that we have a common set of core content standards across most states, one potential benefit is that we can focus on identifying and measuring the linguistic knowledge and skills that students will need in order to meet the CCSS, potentially simplifying and streamlining the development of ELP assessments. Another promising aspect is that the CCSS specify the literacy skills in grades 6-12 expected for the content areas, including social studies and science, which are also mapped to skills delineated in the ELA standards. That is, the CCSS attempt to establish common language skills across the different content areas. In doing so the CCSS offer language test developers the benefits of target expectations from which to create measures of ELP needed to acquire content (and demonstrate learning) expressed in the CCSS. However, the CCSS also pose challenges for future ELP assessment. The twin goals of this paper are to discuss the significant challenges of assessing ELP in ways that are aligned to the content of the CCSS, and to offer practical suggestions for the development of next generation ELP assessments that take account of the CCSS.

Some Challenges

We organize the challenges in the development and alignment of ELP assessment to the CCSS around three sets of related issues: (1) identifying language knowledge and skills in the CCSS, (2) defining alignment in the context of ELP assessments, and (3) articulating a new ELP

standards framework that can guide states in their development of next generation ELP standards and assessments.

1. Identifying language knowledge and skills: The following example is taken from Reading Standards for Literacy in History/Social Studies, Grades 11-12.

4. Determine the meaning of words and phrases as they are used in a text, including analyzing how an author uses and refines the meaning of a key term over the course of a text (e.g., how Madison defines faction in Federalist No. 10).

The standard (and parallel standards at earlier grades and in the ELA Standards) overtly mentions what kinds of tasks students must be able to accomplish with language as they read and, in this instance, attempt to comprehend history texts. Expectations for vocabulary knowledge and analysis skill will likely be measured on future content assessments and would be obvious targets for ELP assessment. Let us take another example, this time from Reading Standards for ELA Informational Text, Grades 11-12.

7. Integrate and evaluate multiple sources of information presented in different formats and media (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem.

Despite being a reading standard, Standard 7 (and parallel standards at earlier grades and for Literacy in History/Social Studies, Science and Technical Subjects) does not overtly mention specific reading skills. In fact, this standard entails the integrated language skills of reading, writing, speaking, and listening. It is also notable that more than one task and skill is implied in one single standard.

These examples from the CCSS demonstrate the complexities of assessing ELP in a way that is aligned to the CCSS. Even if Standard 4 above overtly delineates language skill expectations (and many standards do not, as in Standard 7), it presupposes a whole host of other language skills and knowledge that are not overtly acknowledged. For instance, to demonstrate the skills covered in this standard, students must be able to use the conventions of providing formal definitions, they must be able to talk or write about word choice and semantic refinement as objects of study, and they must understand and use the language of sequencing in order to keep track of word usage across the text. Standard 7, which rather implicitly embeds a wide range of language skills, requires a thorough identification of all the language skills involved.

Considering that the CCSS by their nature describe the expected end-goals at each grade level, identifying underlying or relevant language skills is a challenging but critical step to take for teaching and assessing ELL students. For ELL students, we need to specify the entirety of other language skills that will also be needed to meet the standards. Then, the role of ELP assessments is to measure the extent to which ELL students are able to meet the language demands of the standards in order to identify where they are in their language learning and what they need instructionally to move their learning forward. Indeed, to use a quote attributed to Winston Churchill, we are facing “*a riddle wrapped in a mystery inside an enigma.*” Nevertheless, characterizing the inherent language demands of the CCSS will be at the heart of attempts to support instructional practices and align ELP assessments to the standards. Here, we acknowledge that many students who are acquiring two or more languages simultaneously may also be receiving content instruction in two languages. Bilingual education programming will have its own set of issues concerning both the CCSS and assessment implications of the new standards, and these also deserve a full treatment by the education research community.

Such issues may include assessment in and of the two languages in order to present a more accurate profile of a student. However, these issues are outside the scope of the current paper which was tasked with the examination of ELP assessment relative to the CCSS (see the paper by Brisk & Proctor (2012), for a discussion on the CCSS in bilingual programs).

2. Defining “alignment”: Use of the term “alignment” to describe the quality of the connection between content standards and language assessments differs from that of more traditional notions of alignment between standards and assessments in the field of education. Most often, previous alignment efforts have sought to describe relationships between the content represented in subject matter standards statements and the content coverage provided by test items of the same subject matter (e.g., Webb, 1997). Alignment has also been conducted between content standards and alternative standards or assessments – but again *within* the same content area (e.g., WestEd, 2004). It has been claimed that strong alignment between standards and assessments will help ensure accurate and meaningful measures of student achievement and instructional effectiveness (e.g., Fast & Hebbler, 2004; Herman, 2004; Webb, 1997).

Following NCLB, attempts have been made to judge the degree of linkage between ELP standards and the different sets of academic standards (including ELA as a content area distinct from ELP). For example, the implicit or explicit language demands (at lexical, syntactic, and discourse levels) found to be common to both ELP standards and academic content standards can provide one such mechanism for determining linkage on a linguistic dimension also useful for instruction with ELL students (Bailey, Butler & Sato, 2005/7). Cook (2007) took a similar approach by attempting to link the content of ELP standards and the linguistic *registers* of academic content standards. He operationalized alignment in this context as both linkage (the match between standards), as well as correspondence in terms of depth of knowledge and breadth of coverage. However, the scope of these different attempts has been limited because existing state academic content standards do not overtly mention the language associated with their aspirations for mathematics, ELA, science and social studies. More recently, Chi, Garcia, Surber, and Trautman (2011) applied Cook’s approach to their studies of alignment between the CCSS and the content of the Model Performance Indicators in the 2007 WIDA ELP standards, in which it appears sufficient details in the CCSS allowed for reliable judgments of alignment to be made at most grade levels. However, this focus does not appear to attempt to link the different standards at the level of discrete linguistic features that could also be useful information for ELL instruction and ELP test development. It is also noteworthy that prior efforts have attempted to link ELP standards to academic content standards, and to align ELP assessments to ELP standards, rather than to attempt to establish some manner of alignment between ELP assessments and academic content standards directly.

Attempting to directly align ELP assessments to the CCSS without an intervening set of ELP standards may not be suitable. On the one hand, the CCSS have been written with close attention to the language demands inherent in the content areas. This should make the likelihood of aligning ELP assessments directly to the CCSS more achievable. On the other hand, a set of standards for ELP related to the CCSS is advisable for guiding both instruction and future assessment development (e.g., to avoid undesirable content drift in assessment items).

3. Articulating a new ELP standards framework: As briefly mentioned, existing ELP standards developed or adopted by states are widely varied in terms of the content and expectations at different levels of proficiency. The lack of a common framework to develop and organize ELP standards linked to academic content standards imposes significant challenges

on creating ELP assessments that are useful for ELL students' learning and instruction. It has also raised significant concern about the comparability and fairness of accountability drawn from the assessment results across states and even within states. Different ways of operationalizing the construct of academic language have surfaced in the current generation of ELP assessments. For example, some states' ELP assessments have operationalized the academic language construct by deriving it directly from the ELA standards. The WIDA consortium's ELP assessments attempt to encompass the academic language encountered in ELA, mathematics, science, and social studies. There are also other ELP assessments measuring more general language proficiency across content areas.

The majority of states have elected to adopt the CCSS with augmentation from their own content standards. These states now need to incorporate the language skills delineated or embedded in the CCSS into their ELP standards. While we acknowledge that the concept of alignment itself for ELP assessments has yet to be clarified, establishing a framework for the creation of ELP standards aligned with the CCSS is an area of pressing need. Such a framework can be construed as a set of principles and approaches to guide the formulation of ELP standards.¹ The framework can also guide states throughout their decision-making in the adoption of ELP standards and aligned assessments.

In terms of principles for the development of an ELP standards framework, we propose that:

1. The formulation of an ELP framework attend to its systematic uses across curriculum, instruction, and assessment;
2. The ELP construct be articulated with a view to language as a social practice and action to enhance meaningful language learning for students (see the paper by van Lier & Walqui (2012) for discussion on this topic);
3. The ELP construct be elaborated in ways that can help teachers and students to understand general and specific language demands associated with various school tasks across content areas;
4. Content and language be acknowledged as largely intertwined to help understand the link between ELP and content standards including the CCSS;
5. Macro- and micro-level of details of expected proficiency be delineated at each proficiency level;

¹ The necessity for ELP standards has been questioned. Their quality and usability have been critiqued in the educational field (e.g., Bailey & Huang, 2011), and teachers may simply not consult them. Outright elimination of ELP standards, however, requires that both ESL and content teachers can and will be sufficiently prepared to be able to extract the necessary language skills and knowledge from ELA standards and other content area standards on their own. Teachers will no longer have the option of consulting a document (albeit considered flawed in some current forms) describing expectations for language development and that can serve as guides to the kinds of learning that should occur on route to proficiency in the English language. Without such documents to rely on, the education profession has the opportunity to respond – indeed will likely be impelled to respond – with alternate forms of support for teachers attempting to meet the linguistic needs of ELLs.

6. Clear language be used to describe the expected performance, avoiding vague words and phrases that may be interpreted in different ways.

To execute these principles, new models to help define content and language dimensions for the purpose of instruction, curriculum, and assessment may be created. These principles will also help us understand the unique role of ELP standards and differentiate ELP standards from ELA standards. Although there would be natural overlap between ELP and ELA standards, ELP standards formulated under these principles would contain foundational language skills as well as more complex language skills that are required across multiple content areas in addition to ELA content. An explicit articulation of different levels of proficiency in ELP standards also plays an important role in establishing developmental progressions to guide teaching and student learning.

We identify at least two different approaches regarding the creation of next generation ELP standards:

1. Creation of new standards for all levels of ELP. This approach is a continuation of the current practice, but following the principles outlined above could lead to improvement of the ELP standards. The improved ELP standards should build closer ties to instruction and curriculum of different content areas and articulate the ELP construct to reduce redundancies with ELA standards. This approach assumes a parallel development of language proficiency and content, highlighting a distinctive ELP construct underlying the language demands across all content areas.
2. Creation of standards up to a threshold level of ELP only. This approach acknowledges the need for ELP standards to delineate foundational and basic language functions and skills up to a certain level of proficiency in English. That is, standards can be created that identify precursor skills for achieving the skills articulated in the CCSS or other content standards. Beyond that threshold the language skills students need for successful achievement on the CCSS may overlap sufficiently with the language skills in the CCSS for acquiring and expressing content learning that these may be indistinguishable from the CCSS (although one challenge will be to determine an adequate ELP threshold). Moreover, at upper levels of ELP, language and academic content may be intertwined and difficult to meaningfully assess as separate constructs (see also the paper by Abedi & Linqanti (2012) for a discussion of assessing the content learning of ELL students). As the CCSS explicitly articulate language and literacy skills across content areas, these skills may need to be assessed for all students as part of academic content assessments. Rather than a separate ELP assessment for the upper levels of proficiency, the education field should consider developing and piloting novel options of assessing language related to the CCSS within the new academic achievement assessments (e.g., the Race To The Top consortia assessments). In doing so, it is crucial for the combined expertise of ESL and content teachers to be utilized to develop and implement local-level performance and formative assessments of language and content knowledge during authentic content instruction, activities and practices. This approach highlights the notion that ELL students' English language proficiency must be developed in conjunction with content learning in content classes.

In the next section, we make suggestions for addressing the challenges discussed above and the reality of advancing toward next generation ELP assessments.

Suggestions for Next Generation ELP Assessments

The ultimate goal of assessing ELL students' English language proficiency is to gauge the extent to which they have acquired the necessary language skills to access content learning. By doing so, assessments should provide meaningful information about the link between content learning and language proficiency to support academic achievement. The purpose of current ELP assessments has been to measure annual growth and ultimate attainment, information used summatively for accountability purposes under Title III. While NCLB made a significant impact on paying special attention to ELL students' English language proficiency and the assessment of their proficiency levels and language development, the emphasis on accountability assessment has been criticized for lack of connection to and support for instruction and student learning. For next generation ELP assessments aligned to the CCSS, it is time to consider the purposes of assessment more widely, not only for accountability but for meaningful feedback to teachers and students. We look first at assessment purpose and use and then at determining language constructs to be aligned to the CCSS.

1. Assessment purpose and use: A single assessment cannot serve multiple purposes unless it is deliberately designed to do so. As the purpose of any test drives its design, including the definition of the construct to be measured, the articulation of the intended use of the test is the first and foremost consideration. Under an overarching construct of English language proficiency necessary to meet the academic content standards (e.g., the CCSS), more concrete and operationalized constructs can be defined for ELP assessments serving different purposes. Coverage of the overarching construct may differ for macro- and micro-level assessments (Black, Wilson, & Yao, 2011). Measurement-driven or macro-level assessments are familiar to us as those most often used summatively for large-scale accountability purposes and for covering broad intervals of learning (e.g., a school year). The strengths of these assessments include such factors as economy of scale, uniformity (i.e., standardized administration procedures and scoring), and consistency (i.e., normed for comparison over time/across test-takers). Their weaknesses can include omission of certain important features of the ELP construct (e.g., dialogic interactions, although creative item development using digital technology should ameliorate this), assumed homogeneity of test-takers that raises issues of fairness and bias if assumptions are violated with ELL students, and a high level of assessment literacy required by teachers to interpret test scores.

Performance data-driven or micro-level assessments can also play a role in next generation ELP assessment in providing formative feedback to inform teaching and learning on a continuing basis. This approach to assessment can capture authentic uses of language in the classroom and can account for the local context (small-scale) but can also archive and collate student performances for summative purposes. Weaknesses in this approach include the difficulty of establishing standardization and validity, issues of fairness and bias, and demands on teacher assessment literacy for both implementation of assessments and interpretation of assessment results.

2. Determining the language construct(s) to be aligned to the CCSS: Few may realize that when facing his enigma (Russia at the start of World War II), Churchill went on to say, "but perhaps there is a key." The key to unlocking the enigma that is alignment in this context may involve the adoption and integration of more than one notion of what it means to align an assessment to standards. This includes establishing relationships between standards and assessments and between sets of standards on an abstract dimension such as language

demands in addition to content and cognitive demands. Discrete linguistic and discourse features (both explicit and implicit) in the CCSS might be identified and organized, drawing on existing language competence models. This can serve as a base to define the construct of ELP assessments. A systematic correspondence among language, content, and cognitive demands should then be examined between ELP assessments and the kinds of tasks or activities that the CCSS express. This might be comparable to “connections” to the CCSS, the approach adopted by the WIDA Consortium 2012 draft standards for ELP. Scenarios suggested by the CCSS might serve as model instantiations of the language students will likely encounter in the classroom during different content tasks.

For example, to meet CCSS Reading Standard 7 given above, some language demands include understanding a given question, comprehending the content of multiple oral and written materials, comparing and contrasting the information in the materials, integrating the information in a similar or different theme, and evaluating the relevance of the information to the question. Depending on the topic of the question and materials, knowledge of domain-specific linguistic features (e.g., technical vocabulary, certain grammatical structures) may also be needed. Some of the language skills may overlap with those in other standards. Once an array of language demands and skills is identified, the next step is to organize those skills holistically and systematically at a higher level. This organization can also help in the creation of a core ELP construct or an ELP standards framework. If ELP assessments are built employing this mechanism, alignment to the CCSS may be inherently incorporated into the assessments.

The ELP construct can be further specified by proficiency levels within grade or grade span. These should describe the extent to which students are able to listen, speak, read and write within each standard or, at minimum, key or recurring standards. Specifically, language acquisition theory and empirical evidence of language learning progressions can be used to articulate the four modalities, capturing the subskills or prerequisite skills that delineate proficiency levels (including any skills not articulated in the CCSS). These can serve to organize ELP associated with the CCSS for both instructional and all assessment purposes. Moreover, specific language skills for content areas may also need to be assessed, and the results will need to be available to content area teachers (and, in the case of performance data-driven approaches, content area teachers can be encouraged to adopt these as part of instructional practice).

Implications

We have discussed possible ways to develop next generation ELP assessments aligned to the CCSS. This unique opportunity has a number of implications for policy makers, practitioners, and test developers. New ELP assessments will be used in state and district classification systems for ELLs and will play a role in establishing a “common definition” of ELL. The efforts of developing ELP assessments in alignment with the CCSS could lead to greater ELP data use by both content and ESL teachers; state agencies should establish a system to link content and ELP test data for ease of sharing by content and ESL teachers, which in turn may lead to changes in the instructional strategies and linguistic pedagogies of all teachers. Furthermore, what and how teachers teach is influenced by assessment (i.e., washback effects) (Cheng, 2008). We need to guard against washback that could lead to negative impacts on teaching by carefully defining the ELP construct to be measured. One risk is to de-emphasize the standards that require inter- and intra-personal uses of language necessary for successful engagement in school but are not readily assessable on a traditional large-scale assessment. Another is to ignore entirely the language used during social (i.e., non-scholastic) experiences that is

necessary for becoming fully functional in English across all aspects of a student's life. We should consider ways in which these diverse interactions can also be systematically included in ELP assessment. Finally, teachers will need professional development support in the area of assessment literacy to take advantage of this new opportunity for broadening and improving ELL assessment practices.

References

- Abedi, J. & Linqunti, R. (2012). *Issues and opportunities in strengthening large scale assessment Systems for ELLs*. Paper presented at the Understanding Language Conference, Stanford University, CA.
- Bailey, A. L. & Huang, B. H. (2011). Do current English language development/ proficiency standards reflect the English needed for success in school? *Language Testing*, 28, 343-365.
- Bailey, A., Butler, F., & Sato, E. (2007). Standards-to-standards linkage under Title III: Exploring common language demands in ELD and science standards. *Applied Measurement of Education*, 20(1), 53-78. Previously CSE Tech. Rep. 667 (Dec. 2005). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Black, P. J., Wilson, M., & Yao, S-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement*, 9(2-3), 71-123.
- Brisk, M. E. & Proctor, C. P. (2012). *Challenges and supports for English language learners in bilingual programs*. Paper presented at the Understanding Language Conference, Stanford University, CA.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education, Vol. 7: Language testing and assessment*. (pp. 349-364). Berlin: Springer Press.
- Chi, Y, Garcia, R.B. Surber, C, & Trautman, L. (2011). *Alignment Study between the Common Core State Standards in English Language Arts and Mathematics and the WIDA English Language Proficiency Standards, 2007 Edition, PreKindergarten through Grade 12*. University of Oklahoma, College of Education eTeam.
- Cook, H.G. (2007). Some thoughts on English Language Proficiency Standards to Academic Content Standards Alignment. Working Draft. Retrieved 12/21/11. http://www.nciea.org/publications/RILS_3_GC07.pdf.
- Fast, E. F., & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers.
- Herman, J. L. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education*. New York: Teachers College Press.
- National Research Council. (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- No Child Left Behind (2001). *No Child Left Behind. Title III: Language instruction for limited English proficient and immigrant students*. 107th Congress, 1st Session, December 13, 2001.
- van Lier, L. & Walqui, A. (2012). *How teachers and educators can most usefully and deliberately consider language*. Paper presented at the Understanding Language Conference, Stanford University, CA.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education). Monograph No. 6. Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.
- WestEd. (2004). *Alignment study: Louisiana content standards and the Louisiana Alternate Assessment (LAA)*. San Francisco, CA: WestEd.

Wolf, M. K., Farnsworth, T., & Herman, J.L. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment* 13(2), 80-107.

The Understanding Language Initiative would like to thank the Carnegie Corporation of New York and the Bill and Melinda Gates Foundation for making this work possible. For more information about this paper, please contact **UnderstandingLanguage@stanford.edu**

Understanding Language

Stanford University School of Education

485 Lasuen Mall

Stanford, CA 94305-3096

ell.stanford.edu