Next-Generation Summative English Language Proficiency Assessments for English Learners: Priorities for Policy and Research

Mikyung Kim Wolf, Danielle Guzman-Orth, and Maurice Cogan Hauck

December 2014
Policy Recommendations to Support the Meaningful Use of English Language Proficiency Assessments

1. Federal and state education agencies should provide a clear definition of English learners and guidelines for English learner designation

The current federal definition of an EL (the definition of limited English proficient in NCLB) includes the description of having difficulties in speaking, reading, writing, or understanding the English language that may deny the individual the ability to meet the state’s proficient level of achievement on state assessments, to successfully achieve in classrooms where English is the language of instruction, or the opportunity to participate fully in society (NCLB, 2002). This definition has been operationalized with considerable variation in practice for EL designation. Varied standards and ELP assessments with differing constructs, as well as various weighting schemes for determining overall ELP, have contributed to this variation in defining the term EL in practice. Further, this federal definition is often interpreted to include students’ academic performance as one of the criteria to define an EL, a situation which has lead to the use of a range of criteria, including content assessment results, in many states. The lack of clear guidelines on how to use these various criteria has also created variability in classifying EL students’ ELP levels and exiting students from EL status, even across schools within the same state.

At present, there is a national discussion underway exploring how states and consortia might move toward a more common definition of EL (e.g., Linquanti & Cook, 2013). With the fast movement of current educational reforms in the form of CCR standards, new ELP standards and new assessments, it is possible that a more consistent concept of what English proficient means will emerge. The more streamlined correspondence between content and ELP assessments, largely due to the efforts of the consortia, and using common empirical methods, will help increase the comparability of EL classification across states. However, it is important to recognize that criteria other than results from ELP assessments will be used to make decisions about EL designation (e.g., observation, student interviews, and parental/guardian input). Allowing for local flexibility is essential, considering the different structure of decision-making processes, resources, or students’ needs across local educational agencies. Yet, this flexibility should take place in principled ways.

Therefore, federal and state education agencies must offer clear guidelines about what EL definition local districts should adopt (e.g., the existing federal definition or the consortia’s definition) and how to use ELP assessment results and other criteria to make EL designation decisions. In establishing such guidelines, federal policy guidelines for the reporting of EL students’ ELP development should be revisited to provide explicit guidance on what to report from ELP assessments (i.e., discrete four-domain scores, integrated language skill scores, and weighting schemes of different language skills for overall proficiency scores). Such policy actions would support efforts to establish a more common definition of the status of EL.
2. Federal and state policies should provide guidance on systematic approaches to linking content assessment results and English language proficiency (ELP) assessment results

As noted in previous sections, research has yielded congruent results that there is a strong, positive relationship between students’ ELP and academic performance in content areas based on the data from ELP and content assessments (e.g., Cook et al., 2012; Francis & Rivera, 2007; Parker, Louie, & O’Dwyer, 2009; Thompson, 2012). However, EL students’ ELP levels are not systematically taken into consideration when reporting their content assessment results. At the policy level, the lack of specification or guidance on the connection between Title I (Improving the Academic Achievement of the Disadvantaged) and Title III (Language Instruction for Limited English Proficient and Immigrant Students) has contributed to an inadequate picture of EL students’ performance in accountability reports (Hakuta, 2011; Hopkins, Thomson, Linquanti, Hakuta, & August, 2013). It has also contributed to the isolation of Title III, making the ELP assessment results less known to the content teachers who teach EL students.

Understanding EL students’ ELP levels and their areas of strength and needs in ELP through Title III facilitates better interpretations about EL students’ academic performance on content assessments for the Title I requirement. More importantly, a better understanding aids educators in making appropriate instructional program decisions to meet the students’ needs.

Thus, we call for policy at the federal or state level to require a systematic link between ELP and content assessment results. Federal guidance must be established, requiring the reporting of EL students’ academic content assessment performance accompanied by their ELP levels. The findings from empirical studies suggest there is a clear, positive relationship between EL students’ ELP levels and content assessment performance. This relationship confirms the importance of reporting EL students’ content performance as a function of ELP levels (e.g., Cook et al., 2012). Given the lack of progress of many ELs at intermediate levels of ELP, it is also critical to report on EL students’ ELP progress as a function of initial ELP level and time in the state school system (Hopkins et al., 2013; Working Group on ELL Policy, 2010). One of the next steps to support this action is to have a statewide database system that integrates content and ELP assessment data with the students’ background information. It is also crucial to keep track of reclassified EL students’ prior EL status and date of exit to allow for longitudinal consequential validity studies. Such an integrated database system will facilitate the generation of comprehensive score reports for EL students for both accountability and instructional services. It will also raise the awareness of content teachers about their EL students’ language skills, creating opportunities for shared understanding and collaboration among language and content teachers. These changes would support more effective uses of content and ELP assessment results in order to advance effective EL education.

3. States should provide guidance and professional support to local education agencies for the use of ELP and content assessments to support EL instruction

To support the implementation of the next generation of content and ELP assessments based on CCR standards, professional development and support is more important than ever (Hakuta, 2014). Teachers need to understand the expectations and changes in the new CCR and ELP standards as well as
in new assessments to provide appropriate instruction for EL students. To make appropriate interpretations of assessment results and make valid academic decisions about EL students based on these assessment results, teachers must also understand the construct of the assessments and polices around the assessment uses. Despite the critical importance of professional support for the success of any educational reform, relatively little attention has been paid to this area at the policy level.

As for policy changes to support improvements in EL instruction and assessments, systematic and ongoing professional development around the use of ELP assessments should be in place for both content and language teachers. Clear documents to explain new policies, standards, and assessments are essential to support professional development. An effective channel to communicate the guidelines should also be established at the state, district, and school levels. In particular, professional development to increase teachers’ assessment literacy is crucial considering the current shifts in ELP constructs and the various ways of defining ELP levels across different assessments. Teachers and other stakeholders who use ELP assessment results should have a clear understanding of the ELP construct and assessment scores in order to view them in light of the standards and their instruction across content areas. Further, educators (teachers and administrators) must receive professional development regarding the appropriate and valid uses of assessment results to make informed decisions about EL placement and reclassification.

While funding for and consistent implementation of professional development activities are often a challenge, setting appropriate polices and funding to allow for systematic, continuous professional development should be a high priority for states, because, in practice, teachers are the ones who implement new reform efforts.

4. The federal department of education should monitor the validity and effectiveness of the use of ELP assessments

ELP assessment results involve relatively high-stakes decisions for program evaluation, funding eligibility, and resource allocation, as well as individual students’ academic paths. Hence, it is crucial to ensure that the quality of the assessments and the uses of the assessment results are valid. NCLB requires states to submit evidence to demonstrate the technical qualities of the states’ standards and assessment systems for adequate accountability. To evaluate collective evidence, a peer review program has been utilized with the federal guidelines as in the Standards and Assessments Peer Review Guidance (U.S. Department of Education, 2009). This guidance lays out important types of evidence to collect for technical quality and alignment of content assessments.

Federal guidance for states to examine and demonstrate the qualities of their ELP assessments is also needed and is currently under development. Many of the elements listed in Standards and Assessments Peer Review Guidance (U.S. Department of Education, 2009) are in fact applicable to ELP assessments. However, ELP assessments also require unique qualities, or different priorities in examining their qualities, compared to content assessments. For instance, alignment tools used for content assessments cannot be directly applied to ELP assessments considering the central aspect of linguistic complexity in addition to the cognitive aspect (Cook, 2005). The provision of evidence for rigorous standard setting to determine ELP levels is also crucial for ELP assessments because the levels
determined from ELP assessments influence EL students’ reclassification. Further, states should continuously monitor the uses of ELP assessments, including their consequences, as part of important validation efforts. Particularly, the use and consequences of the assessments for EL designation is a critical area where much guidance is needed. As in the current *Standards and Assessments Peer Review Guidance*, it will be useful to have examples of the evidence to be collected in the areas of ELP assessment uses and consequences. Such guidance will facilitate states’ active investigation and documentation of empirical evidence to support or refine their accountability assessment systems, including those for ELP assessments.

5. **Federal and state education agencies should offer sustained support for collaboration among state stakeholders, practitioners, and researchers**

States have been charged with various tasks to ensure that they have appropriate assessment systems for accountability. The development of summative ELP assessments, the appropriate uses of the assessment results, and the provision of validity evidence entail considerable amounts of specific resources and expertise. It may be unrealistic to expect states alone to comply with all these requirements. While we call for federal support through the provision of guidance for states, we also recommend that federal support include funding specifically for states to collaborate with researchers to investigate the validity of ELP assessment uses. A systematic connection among state stakeholders, researchers, and practitioners will be beneficial for improving the assessment system to support EL education. The sustainability of this effort can be achieved with federal-level guidance and policy enactment.

**Research Areas Needed to Support Policy Recommendations**

The policy recommendations discussed above demand much empirical research to establish appropriate guidance and policies around the use of ELP assessments. Based on the discussion at the Working Meeting and policy areas recommended above, we select the following research areas as the most pressing and relevant to help with implementing policy guidance.

1. **Validation of criteria used in English learner designation**

Much empirical research is needed to produce refined understanding and guidance on best practices in determining ELP levels and reclassifying ELs. As discussed earlier, schools use various combinations of multiple criteria to determine the exit of EL students from EL designation. The primary criterion, results of ELP assessments, requires in-depth investigation. For instance, investigation of the constructs, alignment with standards, composite scores, and standard setting to determine various levels of proficiency for ELP assessments is critical.

In particular, research to help determine the appropriate ELP level at which to reclassify EL students is urgently needed. While defining what it means to be proficient in English in school contexts can be done conceptually, based on the theories and principles, such definitions must be accompanied by empirical data to validate the conceptual definitions and inform the setting of proficiency levels from ELP assessments. To fulfill this need, some specific research areas for further examination include the
relationship between students’ ELP levels and their academic performance in content areas, the academic performance of reclassified EL students over time, and the longitudinal trajectories of current and reclassified EL students’ performance in both ELP and content assessments.

Thus far, only a handful of studies have been completed, providing important insights into determining the levels for EL designation (Cook et al., 2012; Parker et al., 2009; Thompson, 2012). Further research should be conducted in order to generalize and strengthen the findings so that they may inform appropriate policy and practical decisions with confidence. Research in this area also needs to be expanded to examine the consequences of EL reclassification to ensure students do not remain in the EL status longer than needed or exit prematurely despite the need for ongoing instructional support.

2. School language characteristics and English language proficiency developmental stages

The field has made considerable progress in understanding the concept of school language, including both academic and social language. Although there are competing conceptualizations of what constitutes academic language, a general consensus has emerged that academic language needs to be explicitly taught to EL students and assessed through summative ELP assessments (DiCerbo, Anstrom, Baker, & Rivera, 2014). As discussed throughout this paper, the CCR standards and new ELP standards require certain language skills for students to participate successfully in content-area learning. Moreover, disciplinary literacy at the secondary level in history/social studies, science, and technology is specifically explicated in the CCSS. While these disciplinary literacy skills will be embedded in content assessment for all students, it is unclear what language skills need to be assessed in ELP assessments. For instance, should we measure discipline-specific language in summative ELP assessments? If so, how? How can the line between general academic language and discipline-specific language be defined? More research is needed to identify the language characteristics specific to and common across the various content areas, as well as the foundational language skills with which EL students should be equipped in school settings.

With the new types of language demands expected of students in new CCR standards, further empirical research is necessary to help educators and stakeholders better understand the characteristics of the school language EL students need and the ways to operationalize this school language in ELP assessments. This line of research will be beneficial not only for assessment design and development but also for the instruction of ELP to ensure that students are provided opportunities to develop appropriate language skills. The recent national initiative called Understanding Language is an example of research that delves into understanding the language and literacy skills contained in the CCSS and NGSS.

Another important area of research is concerned with collecting empirical evidence to validate the ELP levels described in standards and assessments. A close and systematic examination of actual school discourse and the materials with which both EL students and mainstream monolingual English-speaking students are engaged is a much needed area of research (e.g., Bailey & Heritage, 2014; Hiebert, 2013; Schleppegrell, 2012). Such research will also help to define what being English proficient.

---

2 For more information about Understanding Language, see http://ell.stanford.edu
means for EL students and to set reasonable expectations for the development of ELP. Further, it will advance knowledge about operationalizing academic or school language that would be assessed in ELP assessments.

3. Alignment among standards, curricula, instruction, and assessments

The challenges and issues discussed regarding the development of new ELP assessments and appropriate use of ELP assessments for EL designation point to the importance of empirical investigation of alignment among new standards, curricula, instruction, and assessment. Research into the alignment of ELP standards, curricula, instruction, and assessment is relatively scarce compared to similar research in the context of academic content areas. With the current transition to new ELP standards and assessments, research on the alignment across these four areas is particularly pressing to ensure that ELP assessment results are used validly for EL students. Alignment evidence between the assessments and standards should be gathered via research studies conducted not only by the test developers (or others with direct involvement in the design and development process) but also by independent researchers. In the course of such studies, examination of the correspondence between the language skills represented in ELP standards and assessments and those in content standards and assessments will be necessary.

ELP standards should be understood in the same way by test developers and practitioners, as intended by standards developers. Empirical investigation of a common understanding across standards, curricular materials, instructional practice, and assessments needs to take place. Thus, research is needed to examine practitioners’ actual interpretations of the assessment results and instructional practices based on the standards and assessment results. An investigation of the types of professional support to implement the standards and understand the assessment results will be an important piece of alignment research. This line of research will not only help to ensure fairness and equity in the use of the assessment results but will also offer useful guidance to support the teachers’ understanding of the relationships among the standards, curricula, instruction, and assessments.

Another area of alignment issues lies in the assessment reporting practice. As described earlier, the current accountability requirement has been interpreted by many states and test developers to report four discrete domain scores (listening, speaking, reading, and writing). This reporting brings up challenges for states that are moving toward implementing new ELP standards reflecting the integrated nature of language skills as they are actually used in the classroom. That is, the examination of alignment will become more complicated given the discrepancy between the reporting requirements and the language skills as described in the standards. Research to offer guidance regarding alignment practices will be useful to inform policies on the reporting requirement.

4. Subgroups of English learners and contextual factors

Thus far, we have described the research areas that were explicitly related to the topics presented and discussed during the Working Meeting. One additional area of research underlying the discussions that we the authors would like to raise is examination of variation within an EL group. Despite a wide recognition of the EL group’s heterogeneity, students’ background characteristics and
the relationship between these characteristics and performance on ELP and content assessments have not been well-researched. For example, despite EL students speaking more than 400 languages and coming from various ethnic and racial backgrounds, all the students are grouped similarly and their academic results are reported in the aggregate as EL or non-EL. EL students differ with respect to home language use and proficiency in the home language as well as English (e.g., balanced bilinguals or dominant bilinguals), formal schooling experiences, length of stay in the United States (e.g., U.S.-born, or newcomers), socioeconomic status, types of instructional program (e.g., English immersion, dual language immersion, transitional bilingual education), and more.

Investigating the existing data may provide further distinguishable classes within the EL group (Guzman-Orth & Nylund-Gibson, 2013). Research has shown that EL students tend to be densely clustered in economically disadvantaged areas or schools (Capps et al., 2005). Yet, it is not clear to what extent this contextual factor impacts student performance. Students’ background information can be used by score users to interpret score information to arrive at a more meaningful understanding. This understanding can inform subsequent decision making to support the teaching and learning of students. Thus, research to delve into EL students’ background characteristics and other contextual factors and their relation to student performance would be of considerable value.

Along this line of research to examine the background characteristics within the EL group, it is important to examine the students’ opportunity to learn the language being assessed in ELP and content assessments. For example, some students’ low scores on ELP assessments can be related to their length of stay in the United States, their formal schooling experience, or the instructional services they receive. Hence, it is critical to examine such important background variables and their association with EL students’ progress and performance over time to inform adequate policy guidelines.

5. Further validation of English language proficiency assessment uses

Validation of assessments involves a continual process of accumulating various strands of evidence and making integrated arguments supporting the validity of the intended uses of assessments. The appearance of new assessments and the constant changes in the students taking the assessments warrant the importance of this continual validation investigation. The aforementioned research areas already encompass various sources of validity evidence. Considering that summative ELP assessments are primarily intended to be used for accountability purposes and for EL designation purposes, validity research to support these uses is critical. As mentioned above, empirical research on the constructs of the assessments; the alignment of the assessments with the standards, curricula, and instruction; and the consequences of assessment uses for EL students’ designation needs to be continuously conducted.

Additionally, further validity research on the methods of standard setting and creating composite scores needs to be conducted; this is one of the pivotal issues related to the impact of assessment uses for schools and students that was raised in the Working Meeting. Comparability across consortia assessments and various ELP assessments is also an important validity research area. This area of research will provide important insights into EL classification and the definition of EL status.