

Exploring the Relationships between English Language Proficiency Assessments and English Language Arts Assessments

evea

Washington
Idaho
Indiana
Montana
Oregon



edCount, LLC
Center for Assessment
UCLA
Synergy Enterprises, Inc
PIRE

Ellen Forte, Ph.D.
edCount, LLC
Marianne Perie, Ph.D. and Pamela Paek, Ph.D.
Center for Assessment

Table of Contents

Introduction 1

Basic Distinctions between ELD and ELA 3

Operationalizing Academic English Language Proficiency 4

Relationships between ELPA scores and ELA scores..... 5

English Language Proficiency and English Language Arts 6

Connecting ELD and ELA 8

Conclusions 9

References 11

Introduction

The 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), introduced a new federal perspective on programs and services for English learners (ELs). This perspective encompassed new expectations for the development of academic English language proficiency among ELs and structured a new role for state education agencies (SEAs) in the development of such proficiency. These changes in federal policy have spurred SEA staff and researchers to explore the meanings of English language development in US schools and how English language proficiency is conceptualized and measured.

Under Title I of NCLB, SEAs are required to administer annual assessments of ELs' English language proficiency. Under Title III, these annual assessments are required to be aligned with language development standards that support ELs' access to academic content standards and to yield separate scores for proficiency in reading, writing, speaking, listening, and comprehension. Further, the scores of these assessments must be used to evaluate the degree to which ELs are making progress toward acquiring English and the degree to which they have achieved English language proficiency in each local entity that receives Title III funds. Essentially, these new Title I and Title III mandates apply the same standards, assessment, and accountability systemic reform principles that underlie content components of Title I to language acquisition services for ELs (Forte & Faulkner-Bond, 2010).

These requirements for an academic English orientation and systemic reform principles represented a new vision for the nature of ELs' English language acquisition and posed an entirely new challenge to SEAs. From the enactment of the Bilingual Education Act of 1968, which manifested as Title VII of the 1968 ESEA reauthorization, through the immediate predecessor to NCLB, federal funds to support language support services were distributed from the US Department of Education directly to local education agencies (LEAs). SEAs had no federally-imposed role in determining how these funds were used or for monitoring the consistency or effectiveness of LEAs' efforts. Thus, most states did not have language development standards or assessments and most SEAs had neither the capacity to implement such systems nor to provide guidance and support for their LEAs on how to address the new academic English focus.

Most states have responded to at least some of these new challenges by joining one of two English language proficiency assessment (ELPA) consortia. As of February 2012, 27 states are members of the World-class Instructional Design and Assessment (WIDA) consortium, which provides English language development standards and both screener and annual ELPAs in grades kindergarten through 12. An additional three states use the WIDA standards, but not the WIDA assessments. Eight states comprise the English Language Development Assessment (ELDA) consortium, which provides an annual ELPA only.

The ELDA states and the other 13 states that do not use the WIDA standards have each established their own English language development (ELD) standards. Some states informally share ELD standards, but even with some overlap there are as of February, 2012, approximately 18 to 20 sets of ELD standards operating across the 50 states and the District of Columbia. Each set of standards represents a different definition of English language proficiency and its expected patterns of development across grades and proficiency levels.

This variation across states has contributed to some confusion about the meaning of English language proficiency; in addition, a number of states have struggled to conceptualize or communicate the

distinctions between English language proficiency among students for who English is a second language and English language knowledge and skills that all students should develop through English language arts coursework. The former is meant to be defined in ELD standards and measured by ELPAs; the latter is meant to be defined and measured via ELA standards and assessments. Reading is a required component of each set of standards and assessments; writing is required ELD standards and ELPAs and often included in ELA standards and assessments. While states' ELA assessments generally do not address any aspect of speaking or listening at present, the Common Core State Standards include these components and the assessments being developed with Race to the Top federal funding will address them in some fashion.

Although ESEA will at some point be reauthorized, the general requirements for assessment are not expected to change significantly. Now and for the foreseeable future, ELs will be required to participate in annual ELPAs in grades K-12 as well as in states' annual English language arts assessments in grades 3 through 8 and in high school. Soon, both the ELPAs and the ELA assessments will address reading, writing, speaking, and listening, leading some to question the distinctions between the constructs each assessment is meant to measure. Or, more simply, why would it make sense for an EL with low English language proficiency, as evidenced by her ELPA scores, to take an ELA assessment? Would a low ELPA score in reading not suggest that the student would not be able to access the ELA test. Similarly, most current and near-term mathematics and science content assessments require participating students to have some degree of English proficiency to comprehend and process the content of the test questions. How much English proficiency is enough to take a mathematics or science test in English without a lack of English interfering with item comprehension? What is the test measuring if the student does not have that degree of proficiency.

Even more fundamental than these assessment-related concerns is the question of how educators can know when an EL is able to fully access the academic discourse in classrooms where English is the language of instruction. What evidence would support a decision to exit a student from service?

These are among the questions that led five states to seek funding to explore validity questions related to ELPAs as they are designed and implemented in actual practice. This funding came in the form of an Enhanced Assessment Grant the US Department of Education awarded to the state of Washington to support the *Evaluating the Validity of English Language Proficiency Assessments* project (EVEA; CFDA 84.368). Washington state and its four states partners, Idaho, Indiana, Montana, and Oregon, collaborated with national language development and assessment experts to develop validity evaluation plans that would ultimately help all states address questions about the meaning of ELPA scores and how they relate to scores on other tests and to decisions about instruction and accountability.

None of the five EVEA states was a member of either the WIDA or ELDA consortium; thus, none had the support of other states or an organization to help them work through their validity questions. In addition, each of these states had to develop, adopt, or adapt ELD standards and ELPAs for use in its own unique context.

This paper is one of several documents the EVEA states and their colleagues from edCount, LLC, the National Center for the Improvement of Educational Assessment, and the University of California at Los Angeles produced to explore issues such as those described above in ways that would be generalizable to the concerns any state may have. Below, we first clarify the uses to which ELPA and ELA assessment results are put and then explore the nature of English language proficiency where academic English is the target of both instruction and assessment. We next consider possible inter-relationships between

notions of ELP and ELA and then propose a set of analyses that a state or group of states could use to test the theoretical relationships between English language proficiency and ELA.

Basic Distinctions between ELD and ELA

Although each addresses aspects of grade-relevant English language skills, ELPAs and ELA assessments are designed to measure different constructs that are delineated via different sets of standards. Further, these assessments are intended to yield scores that are used for different purposes.

Both ELA and ELD relate to four modalities: reading, writing, speaking, and listening. Reading and listening are receptive modalities, meaning that they are means by which individuals receive input. Writing and speaking are the parallel expressive modalities to reading and listening, meaning that they are the means by which individuals share information with others. Although the Common Core State Standards (CCSS) include all four modalities, many states' ELA standards exclude speaking, listening, or both. Even when states' ELA standards include one or both of the speaking and listening modalities, most state ELA assessments exclude them. Under Title III of NCLB, states' ELD standards and assessments are required to address all four modalities.

As the foundational ELA modalities, reading and, to a lesser extent, writing, are most representative of what users of current ELA assessments scores consider to be "ELA." Concern about reading skills undergirds state and federal general education policies. Although the ELA CCSS include a strong focus on writing as well as attention to speaking and listening skills, to date, writing has less frequently appeared in states' standards and one would be hard pressed to find worried mention of speaking and listening skills in any public discourse about ELA performance. Although distinct, when reading and writing are both measured on a state ELA assessment, the scores for these measures are often combined to yield a 'literacy' indicator.

Further, the reading portion of states' assessments typically focuses on literary and informational texts, and occasionally poetry. The intent of such reading measures at earlier grade levels is to make claims about students' abilities to decode text and to comprehend what they have read. In later grades, ELA standards and assessments are generally meant to support deeper inferences are made about a student's ability to understand abstract language and find underlying meaning and messages in richer text. In contrast, the reading sections of ELPAs rarely go that far at any grade level. Reading is far more about decoding, vocabulary, and understanding of grammatical structures and discourse patterns than about comprehension of complex literary, informational, or poetic ideas. Similarly, writing in ELPAs focuses more on a student's ability to express thoughts in grammatically correct and fluent sentences, while ELA writing focuses more on organization of one's ideas and the ability to persuade, describe, inform, or entertain.

More generally, ELPAs are designed to measure an academic English language proficiency construct, or group of constructs, as defined in ELD standards. As noted above, these standards vary across US states, but generally relate to the complex sets of linguistic skills that non-native speakers of English are to acquire as they gain proficiency in English. Here, we are primarily interested in English that is acquired intentionally through instruction, in a pre-kindergarten to grade 12 academic setting, and for the purpose of allowing students to access and participate fully in, academic discourse conducted in English.

Several studies indicate an absence of any common academic English framework, which has resulted in varying approaches to developing ELA and ELD standards (Kim & Herman, 2009; Wolf, Herman, Kim, et

al., 2008; Wolf, Kao, Griffin, et al., 2008; Wolf, Kao, Herman, et al., 2008). However, there is some level of agreement among applied linguists as to the definition of academic English being the ability of students to comprehend and analyze texts, ability to write and express themselves effectively, and acquisition of academic content in all academic areas (Anstrom, DiCerbo, Butler, Katz, Millet, & Rivera, 2010; Bailey & Heritage, 2008; Francis & Rivera, 2007; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Scarcella, 2003 Solomon & Rhodes, 1995). Francis et al. (2006) indicate that “mastery of academic language is arguably the single most important determinant of academic success for individual students” (p. 7). Picking up on both the variability in operationalization of ELD/ELP and the commonalities in expert judgments about this construct, Anstrom and colleagues (2010) indicate that:

Most likely there will never be a single definition of [academic English] AE, but a broad national framework that captures the many dimensions of AE could serve as a foundation for states to use in operationalizing AE for their own purposes. As a point of departure, the framework could be envisioned as a series of matrices for each content area based on empirical evidence that provides specific linguistic detail (vocabulary, discourse, syntax/grammar) across the dimensions of (a) grade level or grade cluster and (b) modalities/domains (listening, speaking, reading, writing). Companion documents would demonstrate the interrelated nature of the linguistic features across the dimensions and provide guidance for states in how to use the framework to review and refine their state’s system of ELD standards, curricula, and assessments. The development of such a framework would draw on the collaborative work of educators representing a range of expertise, including linguistics and the content areas as well as curriculum and test development, all sharing a common vision about the need for practical guidance based on empirical evidence. The framework would be a living document that evolves as new evidence becomes available and would be an invaluable resource in shaping a national research agenda on AE. (p. 12)

Thus, this form of English language proficiency, academic English proficiency, may best be thought of as “language ability across relevant modalities [reading, writing, speaking, and listening] used at sufficient levels of sophistication to successfully perform all language-related school tasks required of students at a specific grade level (given adequate exposure and time to acquire the second language)” (Bailey & Heritage, 2010, p. 3). In terms of the generalizability of language skills, academic language proficiency is understood to be situational, depending heavily upon context (Byrnes & Canale, 1987; Lowe & Stansfield, 1988). Students’ abilities to perform well on discrete measures of basic language proficiency, such as formulating declarative sentences or simple questions (e.g., “The sky is blue”, “When is the next bus arriving?”), may be sufficient to function in general settings, but are not sufficient to allow access to the full range of discourse in academic settings (Bailey, 2010).

Operationalizing Academic English Language Proficiency

As mentioned previously, an ELPA must measure four domains: reading, writing, speaking, and listening and states are required to report student performance scores separately for each of these modalities. State must also report a ‘comprehension’ score that is generally based on some combination of reading and listening items. Most states also report a composite score that is based on items from all four modalities.

Development of language abilities varies in nature and rate across the reading, writing, speaking, and listening modalities as well as by the student’s age and native language and native language proficiency

(Bailey, 2008; Kopriva, 2007; McKay, 2006). Second-language development theories stress the importance of both communicative (speaking and writing) and participatory language (listening and reading; Abedi, 2008; Bauman, Boals, Cranley, Gottlieb, & Kenyon, 2007; Cummins, 1981; Lara, Ferrara, Calliope, Sewell, Winter, Kopriva, Bunch, & Joldersma, 2007).

Although states are required to report separate scores for each modality and both research and theory support the notion that the modalities are somewhat distinct from one another, it also appears that a single latent higher order factor of language ability usually provides the simplest and most interpretable factor structure for language assessments (Bachman, Lynch, & Mason, 1995; Bachman & Palmer, 1981, 1982; Bae & Bachman, 1998; Byrne, 2006; Conway, Lievens, Scullen, & Lance, 2004; Kenny & Kashy, 1992; Kunnan, 1995; Lance, Noble, & Scullen, 2002; Llosa, 2007; Marsh & Grayson, 1995; Rindskopf & Rose, 1988; Sawaki, 2007; Shin, 2005). However, given that effective instruction must address each modality and evidence that a student's abilities can vary across modalities within a year and this variance can change across years (Abedi, 2004; Sawaki, Stricker, & Oranje, 2007), continuing the practice of measuring each modality separately on ELPAs remains wise.

In determining a composite ELPA score, different models for weighting the four domains have emerged across ELPAs and tend to be based on perceptions of relative importance (Bailey, 2007; Scarcella, 2003). When the guiding philosophy of the state is that all four modalities be treated as equally important, the ELPA scores across reading, writing, speaking, and listening are simply combined without relative weighting. For instance, on the California English Language Development Test (CELDT), each of the four domains has equal weight in developing the composite score. Scores in other states may reflect different weights based on theories indicating that certain modalities (e.g., reading and writing) are more essential to the acquisition of academic English or are more important skills in academic classrooms. An example of such relative weighting is the Texas English Language Proficiency Assessment System (TELPAS) composite score, which consists of 75% reading, 15% writing, and 5% each of speaking and listening. One of the EVEA states, Indiana, uses the weighting 35% reading, 25% writing, 20% listening, and 20% speaking.

The World-Class Instructional Design and Assessment (WIDA) consortium generates a composite score that consists of 35% reading, 35% writing, 15% listening, and 15% speaking for its ACCESS ELPA. This weighting scheme is supported in WIDA's Guiding Principles of Language Development with the theory- and research-based perspective that students develop language proficiency in listening, speaking, reading, and writing interdependently, but at different rates and in different ways (Gottlieb & Hamayan, 2007; Spolsky, 1989; Vygotsky, 1962).

Relationships between ELPA scores and ELA scores

As indicated above, ELA assessment yield scores that are used to make decisions about the effectiveness of schools' and school districts' approaches to teaching English language arts, typically conceptualized as reading and perhaps writing. ELPA scores are used to make decisions about school districts' (or consortia of school districts') effectiveness in supporting ELs' acquisition of the English they need to participate fully in English-speaking academic classrooms. For the purposes of EVEA, the states agreed to evaluate two primary purposes of ELPAs: determining a student's readiness to exit the ELD program and to evaluate program effectiveness. This, and the above-mentioned common understanding that ELA and ELA assessments relate to more complex and advanced English language skills, would suggest that a student would need to achieve proficiency on an ELPA before she could fully access any English-speaking academic classroom, including and perhaps particularly the ELA classroom. However, states share

anecdotes about students who were proficient on their ELA/reading assessment yet below proficient on the ELPA. While there are explanations regarding the number of domains addressed by each, the influence of Type I and Type II errors in setting cut scores on each, and the types of tasks in each assessment, it is reasonable to expect some relationship between the two scores. From a face validity perspective, it is difficult to argue why a student who is not proficient in the English language could be proficient in English language arts.

This leads us to the five EVELA states' questions about how the ELPA score are or should be related to ELA scores and how ELPA scores may be helpful in making decisions about ELs' participation in other content area assessments. As a start, this paper addresses these questions by considering:

1. How are the four domains (separately and collectively) a measure of English language proficiency?
2. What is the theoretical relationship between English language proficiency (both as a whole and by domain) and English language arts (ELA) assessments?

English Language Proficiency and English Language Arts

The fundamental validity question regarding ELPAs is whether a student who is deemed proficient by an ELPA can successfully function without language supports in academic classes taught in English. These days, academic reading/language arts and mathematics are accompanied by large-scale summative test in grades 3–8 and at specific points in high school. Test developers of state content assessments have been sensitive to the fact that these assessments must measure grade-level mastery in ways that are accessible to all students. Following best practice, item writers cannot assume students are native English speakers and need to develop items to measure the construct without allowing the English language to be a barrier to demonstrating knowledge. This issue becomes complicated in English language arts, which must measure English knowledge as related to reading and writing. Items can be written in ways that are more accessible to students who are not native English speakers, but they still must access grade-level language arts standards. Students taking an ELPA are often at a disadvantage in taking an assessment that is not in their native language because it is impossible to disentangle how their performance on the ELA is a function of their acquisition of the language or of their understanding of the academic content.

Although a common academic English framework would be helpful for determining what students need to learn in different content domains, such a framework still would not address the relationship between ELPAs and academic content assessments, such as English language arts. That may be because there has not yet been an explicit discussion of how attainment of the English language is related to academic reading and writing, there is only an acknowledgement that “students’ development of academic language and academic content knowledge are inter-related processes” (Bailey, 2007; Collier & Thomas, 2009; Echevarria, Vogt, & Short, 2008; Gee, 2007; Gibbons, 2009; Gottlieb, Katz, & Ernst-Slavit, 2009; Mohan, 1986; Zwiers, 2008).

ELPAs focus on students’ attainment of the English language by measuring how well they read, write, listen, and speak English. ELA assessments measure students’ mastery of academic content standards in reading and writing for a specific grade level. One might hypothesize that strong performance on the ELPA would be a pre-requisite for strong performance on the ELA assessment, particularly since most ELA assessments are administered in English with no language accommodation.

In the few studies that have examined the relationships between state ELA and ELP assessments, reading and writing scores from ELPAs have been shown to be significant predictors of performance on ELA assessments (Bailey & Butler, 2007; Cook et al., 2009; Crane, Barrat, & Huang, 2011; Kato et al., 2004; Parker, Louie, & O'Dwyer, 2009). Parker and colleagues (2009) further argue that the domains of reading and writing (literacy skills) are more closely associated with academic ELA performance than are the English language domains of speaking and listening (oral skills).

These studies show moderate to moderately high correlations between the composite scores for ELP and performance on ELA assessments. However, ELA and ELP assessments were not developed to have any explicit connection to each other, so the meaning of one score is unclear in the context of the student's score on the other (Crane et al., 2011). Furthermore, because test scores on content assessments are affected by both ELP and content knowledge, we could be underestimating English learners' actual content achievement by design on more than just the ELA assessment (Cook, 2010; Winter, 2011).

To further explore current relationships among ELPA and ELA assessments, Oregon, one of the states participating in the EVEA project, gathered some data comparing the two assessments. Assessment and language experts in Oregon wanted to analyze the rigor of their cut scores on the state reading assessment. As is shown in Table 1 below, the state reading results of EL students in grades 3 and 4 were summarized by those students' ELPA proficiency level. The Oregon ELPA consists of five levels, with the highest level (level 5) defined as the English language proficiency level, indicative of readiness to exit the ELD program. The higher the ELPA proficiency level, the higher the scale score on the reading assessment, as seen in other studies (see Cook, 2011). However, using the current cut scores for reading, the average score reading score for students at Level 3 on the ELPA was at or above the proficient cut score. However, students at level 3 on the ELPA are considered emerging, meaning they are just beginning to use English for academic purposes. They still require visual organizers for responding in written form. They are not yet at the developing stage where they could respond to context-embedded or reduced situations, meaning they still have not developed the English language skills they need for performing more complex academic tasks. Thus, there appears to be a mismatch between the two tests on the interpretation of a student's English language ability.

The state proposed increasing the cut scores for both grades 3 and 4. The proposed increase is between the average score of students scoring at the developing level (level 4) and the proficient level (level 5). A question for this state and others doing such analyses is around the assumptions or beliefs that underlie the assessments. Oregon may have interpreted this information as evidence that their reading assessment cut scores were too low. Or they may have interpreted this data as indicating that their ELPA levels and descriptors underestimate what students can do at those levels. If these assessments were not developed to have a similar foundation or continuum for reading, it will be impossible for anyone to truly understand what the relationship of these scores mean (Francis & Rivera, 2007). Thus, interpreting the relationship of scores on these assessments will be more meaningful in disentangling English language acquisition and mastery of academic content when the underlying relationships have been clearly articulated.

Exhibit 1. Relationship of ELPA and Reading Results in Oregon for ELD students in Grades 3 & 4

Grade	2009-10 ELPA Performance Level	OAKS Reading 2009-10 Results		OAKS Cut Scores	
		# of Students	Mean Scale Score	Current Reading	Proposed Reading
03	1	391	190.2	204	212
	2	1941	199.1		
	3	1895	206.1		
	4	1497	210.2		
	5	557	214.3		
04	1	184	196.9	211	217
	2	918	203.3		
	3	1238	209.2		
	4	1814	213.9		
	5	1278	218.8		

Connecting ELD and ELA

To support further exploration of relationships between ELD and ELA, this paper proposes a set of steps that could be conducted to empirically test basic connections between the ELD and ELA theoretical frameworks.

As the Oregon example makes clear, it can be difficult to interpret relationships between performance on ELPAs and performance on ELA assessments. We suggest states consider the following questions as they revisit their ELD standards:

- How do you conceptualize the relationship between ELD and ELA standards? What about with other academic content areas?
- How do you conceptualize the importance of each of the four domains as measures of English language development? How will these be further articulated by grade level and with academic content?
- How will you develop ELD standards that can disentangle students’ acquisition of language from their understanding of the content?

Likewise, states should consider these questions as they revisit and revise their ELPAs:

- What other analyses can you use to better study the relationship of data from ELP with ELA assessments?
- How can you use data from both ELP and ELA assessments when setting achievement standards and defining cut scores on each test?

- How do you operationalize the importance of each of the four domains as measures of English language development in scoring of individual domains and for an overall composite score? How do you then set achievement standards? How should you define separate proficiency levels for different language skill domains?
- What analyses can you conduct to disentangle English language development across the domains as they relate to learning academic content?
- Do you believe native language assessment of content is the only way a student can show complex understanding of academic content? If yes, how will you develop assessments in the native languages of all your ELs? If only a majority of languages will be served, what will be done for measuring the academic knowledge of ELs whose native languages are not included in that majority?
- What data will you recommend be used to identify, classify, and reclassify students for ELD programs?

Some questions district and/or school administrators may consider in their development and support of EL programs:

- Do you see English language development and ELA as separate subjects? If yes, how do you support English learners in learning the academic content? If no, how do you support ELs' development of English within the academic context?
- What resources do you need to better prepare and support your teachers to best serve ELs?
- In what ways can the school be organized to expose students to both language development and content? How will this design be extended into classrooms?
- What data will be used to identify, classify, and reclassify students for ELD programs?

Teachers may provide insights if they are asked to provide responses to questions such as:

- What type of information do I need to better understand my ELs' acquisition of English?
- What type of information do I need for identifying, classifying, and reclassifying students for ELD programs?

Conclusions

As of June 2010, the Blueprint of Education proposes to “establish new criteria to ensure consistent statewide identification of students as English Learners, and to determine eligibility, placement, and duration of programs and services, based on the state’s valid and reliable English language proficiency assessment” (USDE, 2010, p. 20). To do this, states are being called upon to make such new criteria actionable, with ELPAs that better measure and serve English learners. While practitioners may ask themselves the questions listed above as part of their renewed efforts for better supporting ELs, some general recommendations for next steps are included here for consideration:

1. Make an explicit connection with academic and ELD standards. Recommendations from research (e.g., Wolf, Herman, Kim, et al., 2008; Wolf, Kao, Griffin, et al., 2008; Wolf, Kao, Herman, et al., 2008) have proposed using the current academic English frameworks as a foundation for directly linking academic and ELD standards. Further, the information from the assessments that measure these connected standards needs to provide a good understanding of student learning to be useful to educators (Abedi & Gandara, 2006; August & Shanahan, 2006).

2. Ensure the ELPA is well aligned to ELP standards that are demonstrably connected to the academic content standards. This will ensure that the assessment measures what is expected and will provide teachers the expectations they need to teach (Johnson, 2005; La Marca, 2001; Näsström & Henriksson, 2008). In addition, formative and interim assessments for measuring ELP should be aligned to these same ELP standards if progress is to be measured within a school year (Cawthon, 2004; Mohamed & Fleck, 2010).
3. Test developers continue their work on using Universal Design and bias/sensitivity rules when developing items to ensure the assessments that are developed do not have high language loads that will impede English learners from demonstrating their content knowledge. Some of the techniques currently used that should be continued include reducing the use of low frequency words and complex language structures, as well as simplifying language (not the content), as long as this does not disrupt the construct-relevant language aspects of cognitive processing. These methods should be used from the outset of item development, not only later as a part of item reviews (Abedi & Hejri, 2004; Abedi, Hofstetter, & Lord, 2004).
4. Examine the way the four domains (reading, writing, listening, and speaking) contribute to a measure of English language proficiency. This information will be used for score reporting and may have implications for how standards are set; for instance, a state may want to consider setting separate cut scores by domain. It will be important to define how these four domains represent multiple constructs or a single construct before deciding on developing a composite score and potentially weighting different domains. A continued effort to explore and research score reporting is needed for validation of different weighting and reporting methods (Bailey, 2007; Sawaki, Stricker, & Oranje, 2007; Scarcella, 2003; Wolf, Herman, & Diete, 2010).
5. Examine the policies that are used for identifying, placing, and exiting students from English language programs. ELPAs should be one of multiple measures that are used for such purposes. If involved in a consortium, agreement should be made on the common standards, assessments, cut scores, and classification policies for ELs, to ensure consistency in operationalizing the theories into practice.
6. Be cautious in concluding anything from the correlational studies between ELPAs and academic content assessments. Until the two types of assessments are based on an integrated set of standards that will support their relationship, any inferences made from these analyses will lead to an incomplete and, perhaps, erroneous interpretation of what these results mean.

References

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4–14.
- Abedi, J. (2008b). Measuring students' level of English proficiency: Educational significance and assessment requirements. *Educational Assessment*, 13, 193-214.
- Abedi, J., & Gandara, 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.
- Abedi, J., & Hejri, F. (2004). Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Applied Measurement in Education*, (17)4, 371-392.
- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1-28.
- Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on Academic English: Implications for K-12 English language learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.
- August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (Executive summary). Mahwah, NJ: Erlbaum.
- Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgments in a performance test of foreign language speaking. *Language Testing*, 12, 238-257.
- Bachman, L. F., & Palmer, A. S. (1981). The construct validity of the FSI oral interview. *Language Learning*, 31(3), 67-86.
- Bachman, L. F., & Palmer, A.S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16, 449-465.
- Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English Two-way immersion program. *Language Testing*, 15(3), 380-414.
- Bailey, A. L. (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.
- Bailey, A. L. (2008). Assessing the language of young learners. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education, Vol. 7: Language testing and assessment* (pp. 379-398). Berlin: Springer.
- Bailey, A. L. (2010). Implications for assessment and instruction. In M. Schatz, & L. Wilkinson (Eds.). *The education of English language learners: Research to practice*. Guilford Press.

- Bailey, A. L., & Butler, F. A. (2007). A conceptual framework of academic English language for broad application to education. In A. Bailey (Ed.), *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.
- Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy, grades K-6: Building reading and academic language skills across the curriculum*. Thousand Oaks, CA: Corwin/Sage Press.
- Bailey, A.L. and Heritage, M. (2010). *English Language Proficiency Assessment Foundations: External Judgments of Adequacy*. Evaluating the Validity of English Language Proficiency Assessments. (An Enhanced Assessment Grant). Available at www.eveaproject.com.
- Bauman, J., Boals, T., Cranley, E., Gottlieb, M., & Kenyon, D. (2007). Assessing comprehension and communication in English state to state for English language learners (ACCESS for ELLs®). In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 81-92). Davis, CA: UC Davis School of Education.
- Byrne, B. (2006). *Structural equation modeling with EQS: basic concepts, applications, and programming*. Lawrence Erlbaum Associates.
- Byrnes, H., & Canale, M. (Eds). (1987). *Defining and developing proficiency: Guidelines, implementations, and concepts*. Lincolnwood, IL: National Textbook Company.
- Cawthon, S. (2004). How will No Child Left Behind improve student achievement? The necessity of classroom-based research in accountability reform. *Essays in Education, 11*. Retrieved March 31, 2009 from <http://www.usca.edu/essays>.
- Collier, V., & Thomas, W. (2009). *Educating English learners for a transformed world*. Albuquerque, NM: Fuente Press.
- Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling, 11*(4), 535-559.
- Cook, H. G., Hicks, E., Lee, S., & Freshwater, R. (2009). Methods for establishing English language proficiency using state content and language proficiency assessments: A white paper. Madison, WI: WIDA Consortium.
- Cook, G. (2011). *Modalities, formats, and English language proficiency assessments*. Presentation to the Reidy Interactive Lecture Series (RILS), Boston, MA.
- Cook, H. G. (2010). *What do we know about English language proficiency assessments? What more should we know?* Paper presented at the meeting of the National Conference on Student Assessment, Detroit, MI.
- Crane, E. W., Barrat, V. X., & Huang, M. (2011). *The relationship between English proficiency and content knowledge for English language learner students in grades 10 and 11 in Utah*. (Issues & Answers Rep. REL 2011-No. 110). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from <http://ies.ed.gov/ncee/edlabs>

- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 1, 132-149.
- Echevarria, J., Vogt, M., & Short, D. J. (2008). (3rd Ed.). *Making content comprehensible for English learners: The SIOP model*. Columbus, Ohio: Allyn & Bacon/ Merrill.
- Forte, E. & Faulkner-Bond, M. (2010). *The administrator's guide to federal programs for English learners*. Washington, DC: Thompson.
- Francis, D. J. & Rivera, M. O. (2007). Principles underlying English language proficiency tests and academic accountability for ELLs. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 13-31). University of California, Davis, School of Education.
- Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved March 17, 2009, from www.centeroninstruction.org/files/ELL3-Assessments.pdf
- Gee, J. P. (2007). *Social linguistics and literacies: Ideology in discourses*. New York: Taylor & Francis.
- Gibbons, P. (2009). *English learners, academic literacy, and thinking*. Portsmouth, NH: Heinemann.
- Gottlieb, M., & Hamayan, E. (2007). Assessing oral and written language proficiency: A guide for psychologists and teachers. In G. B. Esquivel, E. C. Lopez, & S. G. Nahari (Eds.), *Handbook of multicultural school psychology: An interdisciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum, Inc.
- Gottlieb, M., Katz, A., & Ernst-Slavit, G. (2009). *Paper to practice: Using the English language proficiency standards in PreK-12 classrooms*. Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Johnson, D. (2005). *Aligning ELP assessments and ELP standards*. Pearson Assessment Report. San Antonio, TX. Pearson Education.
- Kato, K., Albus, D., Liu, K., Guven, K., & Thurlow, M. (2004). *Relationships between a statewide language proficiency test and academic achievement assessments* (LEP Projects Rep. 4). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved 4/17/2008, from <http://education.umn.edu/NCEO/OnlinePubs/LEP4.html>
- Kenny, D. A., & Kashy, D. A. (1992). The analysis of the multitrait-multimethod matrix using confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. *Educational Assessment*, 14, 212-231.
- Kopriva, R. (2008). *Improving testing for English language learners*. New York: Routledge.

- Kunnan, A. J. (1995). *Test taker characteristics and test performance: a structural modeling approach*. Cambridge: Cambridge University Press.
- La Marca, P. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research, and Evaluation*, 7(21), 1-6.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 7, 228-244.
- Lara, J., Ferrara, S., Calliope, M., Sewell, D., Winter, P., Kopriva, R., Bunch, M., & Joldersma, K. The English Language Development Assessment (ELDA) (2007). In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 47-62). Davis, CA: UC Davis School of Education.
- Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*, 24(4), 489-515.
- Lowe, P., Jr., & Stansfield, C. W. (Eds). (1988). *Second language proficiency assessment: Current issues*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Marsh, H.W., & Grayson, D. (1995) Latent variable models of multitrait-multimethod data. In Hoyle, R. H. (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- Mohamud, A., & Fleck, D. (2010). Alignment of standards, assessment and instruction: Implications for English language learners in Ohio. *Theory Into Practice*, 49, 129-136.
- Mohan, B. (1986). *Language and content*. Reading, MA: Addison-Wesley.
- Näsström, G., and Henriksson, W. (2008). Alignment of standards and assessment: a theoretical and empirical study of methods for alignment. *Electronic Journal of Research in Educational Psychology*, 16(6), 667-690.
- Parker, C. E., Louie, J., & O'Dwyer, L. (2009). *New measures of English language proficiency and their relationship to performance on large-scale content assessments* (Issues & Answers Rep. REL 2009-No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51-67.
- Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing*, 24(3), 355-390.

- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2007). *Factor structure of an ESL test with tasks that integrate modalities*. Paper presented at the Annual Meeting of Educational Research Association. New Jersey: Educational Testing Service.
- Scarcella, R. (2003). *Academic English: A conceptual framework* (Linguistic Minority Research Institute Technical Report 2003-1). Santa Barbara, CA: University of California. Retrieved July 3, 2007, from http://www.lmri.ucsb.edu/publications/03_scarcella.pdf
- Shin, S. K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, 22(1), 31-57.
- Solomon, J., & Rhodes, N. C. (1995). *Conceptualizing academic language*. National Center for Research on Cultural Diversity and Second Language Learning. Washington, DC: Center for Applied Linguistics.
- Spolsky, B. (1989). *Conditions for second language learning*. Oxford, UK: Oxford University Press.
- United States Department of Education. (2010, March). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Alexandria, VA: Author.
- Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Winter, P. C. (2011). Building on what we know—Some next steps in assessing ELP. *AccELLerate*, 3(2), 9-11.
- Wolf, M. K., Herman, J. L., & Diete, R. (2010). *Improving the validity of English language learner assessment systems* (National Center for Research on Evaluation, Standards, and Student Testing Policy Brief 10). Los Angeles: Graduate School of Education & Information Studies, University of California.
- 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*. (National Center for Research on Evaluation, Standards, and Student Testing Rep. 738). Los Angeles: Graduate School of Education & Information Studies, University of California.
- Wolf, M. K., Kao, J. C., Griffin, N., Herman, J., Bachman, L. F., Chang, S. M., et al. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses: Part 2. Practice review* (National Center for Research on Evaluation, Standards, and Student Testing Rep. 732). Los Angeles: Graduate School of Education & Information Studies, University of California.
- Wolf, M. K., Kao, J. C., Herman, J., Bachman, L. F., Bailey, A. L., Bachman, P. L., et al. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses: Part 1. Literature review* (National Center for Research on Evaluation, Standards, and Student Testing Rep. 731). Los Angeles: Graduate School of Education & Information Studies, University of California.
- Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms*. San Francisco: Josey-Bass.