

Promises and Pitfalls: Implications of No Child Left Behind for Defining, Assessing, and Serving English Language Learners

Michael Kieffer, Nonie Lesaux, and Catherine Snow
Harvard Graduate School of Education

The fundamental underlying principles of the *No Child Left Behind (NCLB) Act of 2001* focus on holding *all* learners to high standards of learning and instruction, and in turn, increasing academic achievement of all identified subgroups in the K-12 population. One of these subgroups is the growing population of language minority (LM) learners—those students for whom English is not the primary language of the home (see below for technical definitions). In contrast to the historic lack of emphasis on tracking the achievement patterns of all LM learners and ensuring their academic growth, one of the significant benefits of NCLB has been an increase in awareness of the academic needs and achievement of this population; schools are now accountable for teaching English and content knowledge to these learners. There is little disagreement about the spirit of the law as it relates to LM learners, that is, to ensure that states and districts meet these students' academic needs. However, as with any law or initiative, finding an approach to implementation that ensures that intended benefits are achieved can be difficult. In the specific case of NCLB, with its presumption that test-based accountability (across subgroups of learners who differ in important ways) is the motor for educational change, the issues of valid and equitable implementation are challenging.

The policies imposed by NCLB have indeed raised awareness of the needs of LM learners and the challenges of teaching English and content knowledge to them. However, the policies currently in place fall short in ensuring that all LM students benefit, and risk disadvantaging LM students and misleading schools and districts about their accomplishments and needs. There are at least two specific ways in which NCLB has a very significant impact on the education of LM Learners. The first is through the procedures for categorizing LM students for purposes of disaggregation and achievement monitoring, and for identifying the proportion who will receive specialized support for language development. These procedures warrant considerable attention and refinement if they are to serve LM learners optimally. The second relates not to the population of LM learners itself but to the instruments that are used to assess and monitor both their language development and their academic progress in content areas such as mathematics, science and English language arts. Underlying these two issues is the basic problem of how to define the population.

Issue #1. Defining the LM learner population and monitoring their academic progress

We define LM students as those who come from a home where a language other than English is the primary language spoken. Many of these students are fully bilingual in the home language and English; some are more proficient in English than in the home language; and some speak no English at all upon school entry (August & Hakuta, 1997; August & Shanahan, 2006). In most educational policy contexts it is only that subset of LM learners who are categorized as English Language Learners (ELLs; previously called Limited English Proficient or LEP) who are attended to, because of requirements under civil rights provisions that they be provided access to meaningful learning experiences (e.g., Development Associates, 2003). Evidence suggests, though, that many LM students who do not qualify as LEP may have urgent educational needs (e.g., Gandara, Rumberger, Maxwell-Jolly, & Callahan, 2003; de Jong, 2004).

Although the exact terminology may vary, there are typically three different designations used to classify language minority learners at various stages in their schooling. An initially fluent English proficient (I-FEP) student is one who enrolls in school with English proficiency considered to be sufficient for meaningful participation in mainstream classrooms without any language learning support. A student considered to have an English proficiency level that compromises meaningful participation in mainstream classrooms is classified as ELL and receives language learning support, whereas an ELL student redesignated as fluent English proficient (R-FEP) no longer receives language support because he or she is assumed to have attained proficiency in English. A student is never designated R-FEP upon initial assessment; this designation is only assigned to a student who has qualified for reclassification from a specific ELL program to a mainstream classroom. Originally, under NCLB, no R-FEP students were included in the ELL subgroup for accountability purposes. However, in February of 2004, the U.S. Department of Education established a policy by which states had the option of including R-FEP students for up to two years after redesignation within the "ELL" subgroup (U.S. Department of Education, 2004). All other R-FEP students as well as I-FEP students are represented in the disaggregation system only to the extent that they also qualify under racial/ethnic or socio-economic disaggregation guidelines.

Identifying individual LM students as ELL, I-FEP, or R-FEP is an art practiced quite differently in different states and districts (see Ragan & Lesaux, 2006). Although states differ in their patterns of in-migration and in immigrant residential patterns, differences in classification criteria no doubt account for much of the variability across states in the ratio of ELLs to LMs (14% in New Jersey, 65% in New Mexico). If we are to protect the equal rights of all students to appropriate services and monitoring, LM students identified for ELL services in New Jersey should also be so identified in New Mexico, and vice versa. This is currently demonstrably not the case.

We argue in this paper that the entire range of LM students deserves the opportunity for special educational attention and that, as a population, their academic achievement should be monitored over the long term. Attending to the entire population of LM students

would have at least three advantages over the current practice of targeting only students classified as ELL. First, it would allow for a rational approach to monitoring progress over time by identifying a stable group of students, rather than a group with constantly shifting membership. Second, it would provide a uniform standard across states and districts for defining the population to be served. Third, it would promote the idea that all students at risk for school failure on the basis of their LM status must be served, rather than simply those that fall below an arbitrary (and often poorly measured) standard for English proficiency.

Disaggregation Categories

We argue that identifying the whole population of LM students as a disaggregation category when calculating AYP would contribute to improved achievement for LM students in general, as well as for the ELL tail of the LM population. Such a rethinking would restore interpretability to results about progress toward proficiency for this subgroup, and would increase instructional attention to LM students who are not considered ELL but still lack skills needed for academic success.

As noted above, NCLB implements a minimally nuanced system of categorizing language minority learners, distinguishing only those limited in English proficiency, those fully proficient, or a third, intermediate category of formerly limited but now redesignated. The ELL classification is designed to be temporary, unlike all the other disaggregation categories (gender, race, limited income) for which results are reported. Furthermore, exit from the ELL category is premised on performance on tests that are part of or very like the accountability assessments themselves. This creates a problematic situation when trying to estimate the size of this population, monitor its academic achievement, and determine the factors that influence its progress, in particular the factors that are most related to academic success. The ELL population brings into particularly sharp focus the problem of subgroup performance being interpreted using cross-sectional analyses, rather than longitudinal analyses that represent growth in academic achievement of fixed subgroups of learners.

A specific irony inheres in the use of mainstream accountability assessments with ELLs. The expectation is that the ELL subgroup within schools will achieve the level of performance that defines AYP, just like the other subgroups. But as soon as an individual ELL student gets to a point of scoring pretty well on the assessment, s/he is likely to be reclassified as FEP, which of course reduces the likelihood that the ELL subgroup will make progress. So, on the one hand there is pressure on schools to keep students classified as ELL so as to improve that subgroup's performance to meet the goals of NCLB's Title I. On the other hand there is incentive to reclassify as soon as possible; NCLB's Title III—the lesser discussed cousin to Title I that provides much of the funding for programs to serve ELLs—requires that states establish and enforce goals for ELLs' progress in English, including increasing the percentage who reach proficiency.

The basic flaw in logic is treating, for purposes of monitoring academic progress, a temporary category of ELLs just like a fixed category such as African-American. Membership in racially and ethnically defined subgroups is not temporary; thus, it is entirely possible for the subgroup in a school or district to show AYP. But because ELLs get reclassified on the basis of assessments like those used for mainstream accountability purposes, the schools which are most successful at moving ELLs quickly out of special programs are punished the most severely by not being able to represent the scores of the most successful learners in that subgroup. Alternately, those schools who are successful at improving the academic skills of their ELLs but do not move them out of the ELL subgroup may show AYP but face consequences under their state's implementation of Title III. In addition to the inherent tensions and dilemmas that districts face in serving this population of learners and retaining appropriate funds to do so, a practical consequence is that the statistics on the academic achievement of LM learners are based only on those students with a formal designation (LEP or ELL). They do not include those who have gained the proficiency in English language needed to participate in grade level classes without supports. Thus, for example, districts with good preschool and kindergarten programs that produce students classified as FEP early in their school careers are in effect punished by losing those high achievers from their ELL category. Thus, using ELL rather than LM as the designator almost certainly underestimates the achievement outcomes of the overall population of LM learners, and distorts the information available to districts, states, and the federal government.

Policy-makers have—to some extent—acknowledged this irony, as indicated by the 2004 rule to allow the inclusion of R-FEP students for up to two years after redesignation. Although such a stop-gap measure may allow some schools to avoid a failing label (if they happen to have a large population of these recently redesignated students performing well), it does little to address the fundamental problem of treating a temporary category as if it were fixed. Similarly, Senator Mike Crapo's recent proposal to improve NCLB would make the 2004 provision permanent and add an additional year in which R-FEP students can be counted. Although the Crapo proposal would increase flexibility for schools, the new cut-off point is no less arbitrary than previous ones. The reality of LM learners is that they approach the proficiency of English-only learners gradually, that they constitute a subgroup of importance, and that their academic development must be supported and monitored over the full span of time from initial exposure to English until full proficiency is achieved.

Moreover, a fixed LM learner category would allow for greater attention to those LM learners who may have not held the ELL designation for many years but continue to struggle with the academic demands of school. There is little—if any—evidence to suggest that redesignation as fully English proficient is a reliable and valid predictor of ability to succeed in mainstream classrooms without any language support (e.g., Linqanti, 2001). There is considerable pressure in many states to reclassify students as soon as possible, and in California, Arizona, and Massachusetts special educational settings for ELL students are available for only one year. If redesignated students' scores are reported for only two additional years, the entire span of time during which support is available and academic achievement is monitored may be as little as three years. For

many children this is almost certainly too short. Furthermore, the academic material on which redesignation decisions are based may not adequately represent the actual demands of the curriculum, in particular the demands for understanding and producing academic language. Of special note is that for a large proportion of the LM population these three years of support occur during kindergarten, first, and second grades, when exposure to academic language—in the classroom and in print—is severely limited.

Academic Language for Academic Success

Many skills are wrapped up in the notion of academic language. Vocabulary knowledge (including the multiple meanings of many English words), the ability to handle increasing word complexity and length over time, understanding complex sentence structures and formal discourse structures in English are all aspects of academic language (e.g., Scarcella, 2003). For example, among second graders being read a storybook, several LM students missed the meaning of a paragraph on account of the sentence: *The mother made him get out and he ran off.* In this case, *made* did not carry its most frequent meaning, *to create or build*, but rather the less common causative meaning *to force*. Other aspects of academic language relate to the language of text, including the organization of paragraphs, the function of transitions such as *therefore* and *in contrast*, and a wide range of academic vocabulary that appears far more often in text than in oral conversation. Consider this sentence: *John was very hungry despite having just eaten a large plate of beans and rice.* The term “despite” is key to the meaning of the sentence yet is not a term typically used in everyday conversation with school-aged children.

Academic vocabulary plays an especially prominent role in the upper elementary, middle, and high school years as students read to learn about facts, concepts, and theories in content-area classrooms such as math, science, and social studies. In these classrooms, LM learners encounter many words that are not part of everyday classroom conversation (e.g., *analyze, sustain*) yet are key to comprehension and acquisition of knowledge.

Thus, LM learners with the ELL designation may be reclassified at first grade and may indeed be “proficient” enough for the language demands of the primary grade classrooms and texts. But as those learners move through school, they are likely to be fully dependent on classroom experiences to ensure they develop sufficient sophistication with academic language to respond to demand in the upper elementary, middle, and high school years, whereas English-only classmates have resources outside school to help them develop academic language. Thus, the same learners classified as fully proficient in the primary grades may later lack the level of proficiency needed to participate meaningfully in mainstream classrooms without specialized language support.

Indeed, many LM learners who face academic challenges in the years beyond the primary grades have been enrolled in US schools since kindergarten and do not have a formal designation licensing support services for language development. These learners typically have good conversational English skills, but may lack much of the academic language that is central to text and school success. Several studies of elementary/middle

grade minority learners--whether formally designated ELL or not—revealed that their vocabulary levels were between the 20th and 30th percentiles (Carlo et al., 2004; Francis et al., 2006; Kieffer & Lesaux, in press; Proctor, Carlo, August, & Snow, 2005; Tabors, Paez, & Lopez, 2003). Yet, the role that language plays in determining students' success with academic content cannot be overstated. Proficient use of—and control over—academic language is the key to reading comprehension and content area learning. Lack of proficiency in academic language affects LM learners' ability to comprehend and analyze texts in the years beyond the primary grades, limits their ability to write and express themselves effectively, and can hinder their acquisition of academic content in all academic areas, including mathematics. For example, in the domains of math and science the sentences below may appear on worksheets or in problem sets. For each sentence, students need a specific understanding of at least two words that are considered academic in nature and are only likely to be encountered in print.

- 1) *Directions: Make and record three observations.*
- 2) *Calculate the sum of the first eight terms of the sequence.*

Thus, many LM students—whether designated ELL or not—are likely to struggle with academic language, especially in the years beyond the primary grades. However, at this time there is no mechanism in place to systematically monitor the progress of the population as a whole. Monitoring of the subset typically lasts only a few years. If a longitudinal mindset were adopted to monitoring the progress of all students who were ever subject to designation as ELL, then in fact the category would be more properly thought of as LM. We argue that this would be the optimal approach to ensuring both attention to and equity for students who arrive at school needing to acquire English proficiency.

We do not argue that the importance of English proficiency for LM learners should be downplayed, nor do we argue that the temporary ELL designation should be abolished. The careful and valid identification of those students most in need of additional language support for accessing the curriculum at any given grade level is essential for providing them equal opportunities to learn. In fact, we assert that a uniform and psychometrically valid standard for classifying students as ELL imposed nationally would have enormous benefits for ensuring that learners are provided with more equitable support services across states and districts. However, for the reasons described above, such a standard should *not* define the category by which a school's success in promoting the learning of the LM population is judged.

Creating a fixed category for LM status is, as previously noted, only one of two issues that needs to be addressed in order to increase the positive impact of NCLB on LM learners. The second (related) issue, discussed in the section that follows, focuses on the actual assessments used to evaluate and monitor LM learners' language development and academic progress (in science and mathematics as well as English language arts). Standardized and standards-based measures play a central role in defining this population and monitoring its academic achievement, as well as in making decisions about the placement of individual LM learners (e.g., reclassification). Several questions about

content and the predictive validity of these assessments must be addressed in order to establish that the system is indeed serving these learners effectively.

Issue #2. Assessments of Language Development and Academic Progress for LM Learners

Language Proficiency Assessments

NCLB had the beneficial effect of initiating the development of language proficiency assessments for use in the required monitoring of ELL students nation-wide. Although many states had previously been assessing some of their LM learners' proficiency and academic progress, at various levels and for varying purposes, NCLB has ensured that these data are systematically collected and reported, using valid instruments. The potential benefits here include increased breadth and improved psychometric properties of the tests being used to evaluate and monitor the development of LM students' English proficiency – whether in speaking, listening, writing or reading. These also include the potential to shift instruction for ELLs to reflect standards for English language development. The dangers, however, relate to a persistent lack of emphasis within the language proficiency assessments on the complex academic language needed for success in content area classes, and thus the risk that this key domain be neglected in instruction for ELL students.

Though U.S. schools have been classifying incoming students as LEP since the early 1970s, the criteria for those decisions and the tools available to make them were generally unsatisfactory. Indeed, there was considerable ambiguity about the basis for providing services to LM students, i.e., whether simply a low level of English knowledge or discrepancy between knowledge of English and knowledge of another language should trigger access. More serious, though, was the absence of well-designed and rigorous tests that reflected the skills that would be a sensible basis for sending students to mainstream classrooms. Probably the most widely used assessment between 1980 and 2000 was the Language Assessment Scales (LAS; de Avila & Duncan, 2005), a test that, although appropriate for assessing basic proficiencies, is not linked to any particular academic outcomes, and has insufficient alternate forms or psychometric sensitivity to be used to monitor progress (e.g., Pray, 2005; Del Vecchio & Guerrero, 1995). The test measures speaking, listening, reading and writing skills in either English or Spanish for K-12 students; most often it is administered in English in order to determine proficiency level for placement in, and exit from, programs for ELLs. Given its emphasis on reading and writing as well as oral proficiency, the LAS itself represented a vast improvement on the Bilingual Syntax Measure (Burt, Dulay, & Hernandez Chavez, 1976), the instrument most widely used previously, which focused exclusively on oral proficiency, and primarily in conversational contexts. In spite of this, in the context of NCLB requirements, the LAS is not optimal as a tool for classification, reclassification, or progress monitoring.

Thus, in the wake of NCLB's requirement that all ELLs be tested annually on English speaking, listening, reading, and writing skills, several proficiency tests have been developed (some states developed their own tests, while other states formed themselves into consortia to reduce development costs). Presumably these new tests will be used as part of the system for classifying students as ELL or formerly ELL as well as to meet the accountability requirements of NCLB's Title III; since their psychometric properties are almost certainly better than those of tests used previously, they will probably represent an improvement.

Furthermore, whereas the previously used tests focus primarily on social language as opposed to academic content, the state-developed tests are typically aligned with state standards for second language development. In addition, the NCLB guidelines for these proficiency assessments specify reading and writing as well as listening and speaking, which may help educators to conceptualize language proficiency as more than basic oral conversational skills. In theory, this feature should promote greater instructional emphasis on academic language and content. However, the guidelines do not specify anything in particular about academic language or predicting performance in content area classrooms; after about 3rd grade, it is performance in content areas, not just proficiency in English, which constitutes academic success. Thus, the likelihood of being able to do the work in math, science, social studies, and literature study ought to be the criterion against which LM learners' progress is monitored. To the extent that 'academic performance' is included in current tests, it is typically operationalized as alignment with English Language Arts standards, rather than as the full range of skills needed for achievement across all content areas. Some districts and states include grades or performance on achievement tests as additional criteria for exiting ELL programs (Ragan & Lesaux, 2006), but again ELA performance is the most likely to be the focus.

Figure 1 provides two examples drawn from Grade 6-8 Texas' Reading Proficiency Test in English (TRPT), one of the older state tests of English proficiency. Although it is a reading test with some content-based items, the TRPT shows a strong emphasis on basic vocabulary and everyday uses of language, such as telling time or reading a calendar, even in this version of the test for middle school students. These language characteristics contrast markedly with those of math or English Language Arts achievement tests used for accountability purposes.

The absence of sufficient attention to the full range of meaningful indicators of academic performance on the proficiency tests may be biasing instruction in the bilingual and Structured English Immersion (SEI) classrooms that serve ELLs toward conversational English or reading/writing the simple narrative texts that predominate in literacy instruction in the primary grades. Of course, the different language proficiency tests vary enormously in the degree to which they attend to higher-level academic performance. Even the format of the listening and speaking components of the assessments vary widely. California and Arizona use individually administered tests with standardized listening and speaking prompts whereas Massachusetts and Texas have teachers, trained and qualified on the measure, administer a tool to assess language proficiency by observing students performing academic and social tasks in the classroom, over a period

of time. The (structured) observations are taken in conjunction with a rubric that focuses on students' ability to communicate, and result in a set of scores that reflect language proficiency. It is an open question which of these two approaches will have the greatest benefits for students. Although there is very little empirical evidence comparing these two approaches, we might suspect that the standardized test approach will likely yield more reliable and comparable scores whereas the rubric approach, if well implemented, may provide a more valid measure of students' performance in class and has the potential to raise teachers' awareness of the importance of academic language skills, given the structured observations of language in which they must engage.

The range of skills associated with reading and writing proficiency, successful use of academic language, and adequate oral English are difficult to characterize for either native speakers or for second language learners (e.g., Scarcella, 2003; Bailey & Butler, 2003). However, it is very clear that, in most states, the expectations for language minority students are derived from English Language Arts standards developed for native speakers, and thus may well be quite irrelevant to success in reading and writing for purposes of learning math, science, or social studies. As previously noted, it is important to distinguish academic from conversational language skills, and noteworthy that many of the LM Learners who struggle academically have well-developed conversational English skills. Thus, there is a need for academic language to be a prominent feature of language proficiency assessments, especially as it relates to content area material.

Finally in addition to considerations of the content and validity of Language Proficiency assessments for LM students, a significant danger with the shift to an assessment-based system relates to the way in which the assessment results will be used. Although the proficiency tests have been, and will be, designed for the purposes of evaluating schools' success in moving ELLs toward English proficiency, states and districts undoubtedly use these assessments for other purposes for which they are not designed. For instance, California districts routinely use the California English Language Development Test for initial placement, annual monitoring, and reclassification of students, as well as to inform decisions about interventions for struggling learners. Given the psychometric properties of the test and the complexity of the language proficiency construct, this single measure cannot possibly serve all of these four purposes well. A test designed strictly to identify whether a learner is above or below a particular threshold is likely to be insensitive to fine distinctions such as those between beginning and early intermediate students. This same test is likely to provide little, or no, information on which to base interventions for individual children who are struggling. Such practices are examples of inappropriate and unethical test use¹; in these instances the benefits of tracking the academic achievement of the population are in fact outweighed by the costs of using the tests in inappropriate ways.

Mainstream Accountability Assessments

Under NCLB guidelines, students classified as ELLs must be included in the state accountability assessments after one year in U.S. schools, and must be tested entirely in

¹ For a discussion of appropriate and ethical test use, see Sattler, 2001

English after three years². Here the potential benefits include tracking the academic achievement of these learners and ensuring that districts, schools, and teachers incorporate this population in their instructional plans and efforts. In contrast, the very significant potential danger derives from the fact that any test is to some degree a language test and thus, that LM students' scores are not necessarily a reflection of their academic ability, but instead a reflection of degree of understanding of the test items and the academic language needed for content area success.

It might seem that ELL students are not particularly disadvantaged in performance on some content areas of standards-based tests. For example, math is thought of as a rather language-free zone, especially in the elementary school years. A very common misconception about mathematics is that it is a "universal language," one that is synonymous with numbers and symbols, and a "culture-free" static body of knowledge. Clearly though, especially in the elementary grades, learning of mathematics is verbally mediated; the association of verbal labels to mathematical forms and expressions is common (e.g., Lager, 2006). Mathematics language is often a specialized form of natural, conventional language and requires a re-interpretation of the way language is used in everyday settings (e.g., Cuevas, 1984). Much instruction and assessment in mathematics curriculum occurs via discourse and text that is characterized by academic language (e.g., Cazden, 1986).

A careful look at the math items on many state tests, in particular the more challenging state tests, reveals that they make enormous language and reading comprehension demands on students. Some of these demands derive from unnecessary linguistic complexity that has nothing to do with the central domain being assessed and thus calls into question the validity of inferences based on the scores (e.g., Abedi, Lord, & Hofstetter, 1998). However some of these demands are intricately related to the conceptual understanding and application involved in the domain. Figure 2 shows two released items from the California Standards Test for grade 6. In each, notice that there is unnecessarily rare vocabulary that may be unknown to some ELLs, such as *orchard*, *harvested*, *acre*, and *band of a hat*, as well as unnecessarily complex sentence structure, such as the verb phrases *could be solved* and *is shaped*. However, each item also presents linguistic complexity that is central to the mathematical concepts being assessed, including the math vocabulary *proportion*, *cylinder*, *measure*, and *diameter*.

There is substantial evidence that the size of the math achievement gap between ELLs and native English speakers differs as a function of the language demands of the items (e.g., Abedi, Lord, & Hofstetter, 1998; Abedi, 2003). As noted above, some of these language demands are irrelevant to the measurement of math achievement, but derive from the academic language central to understanding and solving mathematical problems. This is evident through studies showing that even when the unnecessary linguistic complexity (e.g., syntax and the sentence structures) of math test items is removed, ELLs often perform no better than they did on the original items and continue to perform

² Originally, NCLB required assessment of all students but a 2004 decision allowed states the option to give recent immigrants a single year until they are required to be tested (U.S. Department of Education, 2004). There have been proposals to make this option permanent as well as to extend this time period.

substantially worse than native English speakers (e.g., Abedi, Courtney, & Leon, 2003; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Abedi, Hofstetter, Baker, & Lord, 2001; Abedi, Lord, & Hofstetter, 1998). This suggests that ELLs in the U. S. are not being taught the *necessary* academic language involved in doing sophisticated mathematical problem solving.

As a way to address some of concerns about the language demands of content area assessments, NCLB requires that states provide testing accommodations for ELLs, such as English dictionaries, bilingual dictionaries, extra time, and native language versions of the test. However, states have not received guidance about which accommodations to provide or how to provide them, which has led some states to adopt accommodations for ELLs that have no theoretical justification, such as preferred seating or testing in small groups (Rivera & Collum, 2006). To date, there is little supporting evidence for the efficacy of even those accommodations considered sensible and appropriate in addressing ELLs' language difficulties without changing the construct being measured; a recent review and meta-analysis of eleven experimental studies comprising 37 samples found the most commonly used accommodations to be largely ineffective (Francis, Lesaux, Rivera, Kieffer, & Rivera, 2006). Providing bilingual dictionaries or native language versions of the test, although often touted as the most fair treatment, does not necessarily yield higher or more accurate test scores for ELLs; there is some evidence that ELLs perform no differently when provided bilingual dictionaries and even perform worse when tested in their native language. Of course, the efficacy of native language tests or dictionaries depends greatly on whether the students have ever learned the material being tested in their native language. Francis et al. did find that providing English dictionaries in some cases was somewhat effective, but had only a very small, if significant, effect on narrowing the substantial gap in content area performance between ELLs and native English speakers.

Although test-makers must ensure that they do not introduce unnecessary linguistic complexity, educators must also realize that better assessments will not eliminate the real differences in content area achievement between ELLs and their native English speaking peers. Academic language is indispensable in presenting higher grade-level material in every content area and in providing all students the skills they need to function at higher levels in those subjects. Having tests that reduce the use of complex, academic language is in fact a disservice to these learners if those tests are also omitting crucial content needed for academic success. Ultimately, the LM population must receive high quality instruction—which includes an emphasis on the language of the domain—in content areas, and in doing so be held to the same academic standards as their English-only peers.

Summary

In contrast to a historic lack of emphasis on tracking the achievement patterns of all LM learners across the nation, and ensuring their academic growth, one of the significant benefits of NCLB has been an increase in awareness of the academic needs and achievement of students from non-English speaking homes; schools are now accountable for teaching English and content knowledge to these learners. There is little disagreement

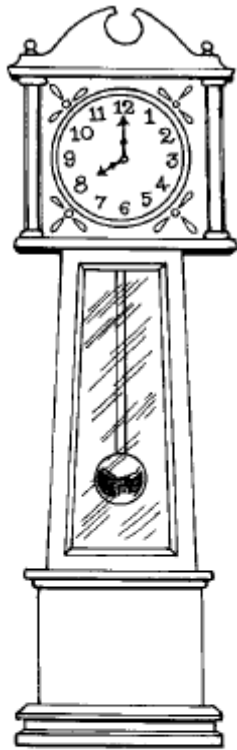
about the spirit of the law as it relates to LM learners, that is, to ensure to that states and districts meet their academic needs. However, if all learners are to participate in large-scale assessments, then a prerequisite is that the tests being used are valid and ensure equitable outcomes. This is particularly the case for LM learners given the complexity of second language acquisition, the differences in language proficiency within the population, and the difficulties in designing tests in which language proficiency is *not* one of the primary skills measured.

Although we are not opposed to the use of a test-based system such as NCLB to hold schools and districts accountable for serving LM students, to be successful such a system requires a rational approach to defining the population and careful attention to the valid assessment of their skills. This growing population indeed deserves to be part of the accountability system if in fact, as planned, accountability results in a more systematic and thoughtful delivery of educational services to meet their needs. This is particularly important because this population has historically been underserved and subject to uneven standards of instruction. However, the current design of the system under NCLB, in particular the procedures for defining the ELL population and devising assessments for language minority children, fails to serve the purposes of the law in a number of ways. To lessen the negative and increase the positive impact, and to be consistent with the spirit of the law, subsequent attention must focus on:

- A national definition and operationalization of the constructs of LM learner and ELL:
 1. A LM learner comes from a home where the primary language of use, as reported by parents, is *not* English. LM learners would then constitute a fixed category for the purposes of data reporting under NCLB. This would eliminate the problem of an inauthentic and inaccurate picture of achievement among LM learners that is created when the focus is only on a small subgroup—which happens to represent the tail end of the distribution of language and literacy achievement—of the population. This would not preclude the need for a distinction between ELL and LM in order to determine which students receive more intensive support services.
 2. ELL: the subset of this population who need intensive language support services in order to participate meaningfully in mainstream classrooms, based on English language proficiency measures that include academic language and provide valid inferences for students' future success. The use of multiple measures and procedures to identify this subset of the LM population should be made uniform across states and districts. An ELL student redesignated as fluent English proficient (R-FEP) no longer receives language support because he or she has attained proficiency in English; such students, however, remain in the LM group for the purposes of accountability and progress monitoring.
- The need to ensure that academic language plays a well-defined role in the assessment of language proficiency. Different language proficiency assessments

place very different degrees of emphasis on academic language as an indicator of progress toward full proficiency. A systematic study of the progress of ELLs (both progress in getting reclassified, and concomitant performance on state accountability and NAEP assessments) could exploit this natural experiment to explore the impact of different test designs on long-term student performance, and establish the extent to which reclassification is a valid indicator of ability to thrive in mainstream content area classrooms. Results of this research would address basic policy questions: On what grounds should ELLs be reclassified? Do lax or stringent criteria for reclassification make a difference in long-term outcomes? How much support do reclassified ELLs need to access classroom learning, and for how long?

- The need for content area assessments to reflect the academic language demands of the content and the concepts reflected by that language. While we are not arguing for unnecessary complexity in the tests, the tests should motivate instruction for LM students that reflects high academic standards and provides the opportunity for them to develop the academic language needed for content area learning.
- The need for tests and measures to be used in an appropriate and ethical manner. Currently, many tests are being used for multiple, competing purposes despite designs and psychometric properties that do not support such use. The implementation of the Reading First provisions of NCLB included recommendations concerning assessments to be used, as well as guidelines for preparing teachers to use the information derived from assessment. A similar set of policies should be introduced for those teaching the LM population: a set of assessments together with guidelines for their use, and guidance to teachers about interpreting test results as a basis for planning instruction. Furthermore, incentives for the development of multiple measures of language and content for LM students would help ensure the availability of tests to serve different functions (e.g., diagnosis, placement, progress monitoring) appropriately. A test that reliably monitors the achievement of students at the population level cannot simultaneously provide student-level information that is useful for purposes of placement and/or to select interventions (e.g., to identify if students need code-focused, fluency-focused, or oral language-focused interventions); thus a wide array of psychometrically sound measures are needed. The development of multiple measures would also increase opportunities to establish validity of the measures for LM learners



- 5 What time is it?
- A It is eight o'clock.
 - B The clock is old.
 - C It is early in the afternoon.
 - D It is time for lunch.

Figure 1. Items from Texas' Reading Proficiency Test in English for grades 6 - 8.

School Vacation

- 1 Jack gets a letter from his cousin Ted. Ted lives in the mountains. It is cold there in winter. There is a lot of snow.
- 2 Jack wants to visit Ted. Jack checks his school calendar to see when he can go to Ted's house.

February School Calendar						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1 Game Day	2	3	4	5 Bake Sale	6
7	8	9	10	11	12 Report Cards	13
14	← WINTER BREAK—NO SCHOOL →					20
21	22	23	24	25	26 Parents' Day	27
28						



- | | |
|--|---|
| <p>22 What event will take place on February 26?</p> <p>F School Play</p> <p>G Game Day</p> <p>H Parents' Day</p> <p>J Bake Sale</p> | <p>24 This story is mostly about a boy who —</p> <p>F wants to visit his cousin</p> <p>G likes to read</p> <p>H plays in the snow</p> <p>J likes cold weather</p> |
|--|---|

Figure 1. Items from Texas' Reading Proficiency Test in English for grades 6 - 8.

- 6** A farmer harvested 14,000 pounds of almonds from an 8-acre orchard. Which proportion could be solved to find x , the expected harvest from a 30-acre orchard?

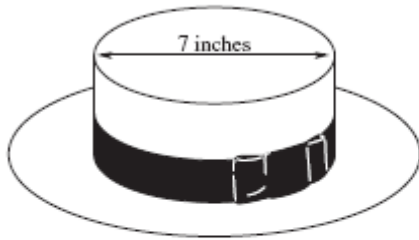
A $\frac{8}{14,000} = \frac{x}{30}$

B $\frac{8}{14,000} = \frac{30}{x}$

C $\frac{30}{14,000} = \frac{x}{8}$

D $\frac{30}{14,000} = \frac{8}{x}$

-
- 37** The top part of this hat is shaped like a cylinder with a diameter of 7 inches.



Which measure is *closest* to the length of the band that goes around the outside of the hat?

- A 10.1 inches
B 11.0 inches
C 22.0 inches
D 38.5 inches

Figure 2. Items from the California Math Standards Test for grade 6

References

- Abedi, J., Courtney, M., & Leon, S. (2003). *Effectiveness and validity of accommodations for English language learners in large-scale assessments*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance test accommodations: Interaction with student language background*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J. (2003). *Impact of student language background on content-based performance: Analysis of extant data*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Lord, C. & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- August, D., & Hakuta, K. E. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- August, D. & Shanahan, T. (2006). *Understanding literacy development in a second language: The report of the national literacy panel*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bailey, A. L., & Butler, F. A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Burt, M. K., Dulay, H. C., & Hernández-Chávez, E. (1976). *Bilingual syntax measure*. San Antonio, TX: Harcourt, Brace, Jovanovich, Inc.
- Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39, 188-215.
- Cazden, C. (1986). Classroom discourse. IN M.C. Wittrock (ed.), *Handbook of research on teaching* (3rd ed.) (pp. 432-463). New York: MacMillan.

- Cuevas, G. J. (1984). Mathematics learning in English as a second language. *Journal for Research in Mathematics Education*, 15(2), 134-144.
- Del Vecchio, A. & Guerrero, M. (1995). *Handbook of English language proficiency tests*. Albuquerque, New Mexico: Evaluation Assistance Center.
- Development Associates (2003). *Descriptive Study of Services to LEP Students and LEP Students with Disabilities. Volume I: Research Report*. Report submitted to U.S. Department of Education, OELA. Arlington VA: Author.
- de Avila, E. A., & Duncan, S. E. (2005). *Language assessment scales, English*. Monterey, CA: CTB MacMillan McGraw-Hill.
- de Jong, E. J. (2004). After exit: Academic achievement patterns of former English language learners. *Education Policy Analysis Archives*, 12 (50), 1-18. Retrieved November 3, 2006 from <http://epaa.asu.edu/epaa/v12n50/>.
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the Diagnostic Assessment of Reading Comprehension. *Scientific Studies of Reading*, 10(3), 301-322.
- Francis, D. J., Lesaux, N., Rivera, M., Kieffer, M. , & Rivera, H. (2006). *Research-based recommendations for the use of accommodations in large-scale assessments*. Portsmouth, NH: Center on Instruction.
- Gandara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, 11 (36), 1-52. Retrieved November 3, 2006 from <http://epaa.asu.edu/epaa/v11n36/>.
- Kieffer, M. J., & Lesaux, N. K. (in press). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher*.
- Lager, C. A. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology*, 27, 165-204.
- Linquanti, R. (2001). The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners. Policy report to University of California Linguistic Minority Research Institute.
- Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal*, 29 (2), 387-409.

- Proctor, C. P., Carlo, M., August, D., & Snow, C. E. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, 97(2), 246-256.
- Ragan, A., & Lesaux, N. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives*, 14(20). Retrieved November 3, 2006 from <http://epaa.asu.edu/epaa/v14n20/>.
- Rivera, C. & Collum, E. (2006). *State assessment policy and practice for English language learners: A national perspective* (pp. 1-173). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sattler, J. (2001). *Assessment of Children: Cognitive Applications (4th Ed.)* La Mesa, CA: Sattler Publishing Inc.
- Scarcella, R. (2003). *Academic English: A conceptual framework*. Los Angeles: Linguistic Minority Research Institute.
- Tabors, P., Paez, M., & Lopez, L. (2003). Dual language abilities of bilingual four-year-olds: Initial findings from the Early Childhood Study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice*, 1(1), 70-91.
- U.S. Department of Education. (2004). Fact sheet: NCLB provisions ensure flexibility and accountability for limited English proficient students. Retrieved on November 3, 2006 from <http://ed.gov>.