

No Child Left Behind: Misguided Approach to School Accountability for English Language Learners¹

By James Crawford

“It is crucial ... to include all students in testing designed to hold teachers and administrators accountable for the education they are providing these students. However, testing students whose language skills are likely to significantly affect their test performance will yield inaccurate results. ... The aggregate performance of language subgroups that are inappropriately tested can be seriously misunderstood, and decisions influenced by invalid test results can have a significant impact on their lives.”
– National Research Council (2000)

“There is always an easy solution to every human problem – neat, plausible, and wrong.”
– H. L. Mencken (1917)

Holding schools accountable for results is a goal with broad support among the American public, policymakers, and educators themselves. There is a growing recognition that our children deserve no less – especially children whose academic needs have often been ignored, leading to achievement gaps that no just society should tolerate. The consensus falls apart, however, when it comes to means: how to design accountability systems that yield fair, accurate, and useful information on which to base decisions about school improvement. What kinds of oversight will ensure that students are achieving to their full potential, yet avoid arbitrary, one-size-fits-all mandates that disrupt the educational process? In short, how can we ensure that the “solution” does not exacerbate the problem?

The No Child Left Behind Act (NCLB) is the latest attempt to resolve this question. The law’s stated aims are worthy, to be sure. Unfortunately, its approach to school accountability is overly rigid, punitive, unscientific, and likely to do more harm than good for the students who are now being left behind. Nowhere is this more true than in the case of English language learners (ELLs).

¹ This paper was presented at a Forum on Ideas to Improve the NCLB Accountability Provisions for Students with Disabilities and English Language Learners, sponsored by the Center on Education Policy, Washington, DC, September 14, 2004.

Many of those supported passage of this legislation in hopes that its stress on high standards for all students, combined with enforceable requirements for meeting those standards, would lead schools to pay increased attention to the academic progress of ELLs. That has indeed occurred. But experience has also shown that NCLB is not bringing the *kind of attention* that would benefit these children.

To the contrary, the law does little to address the most formidable obstacles to their achievement: resource inequities, critical shortages of teachers trained to serve ELLs, inadequate instructional materials, substandard school facilities, and poorly designed instructional programs. Meanwhile, its emphasis on short-term test results – backed up by punitive sanctions for schools – is narrowing the curriculum, encouraging excessive amounts of test preparation, undercutting best practices based on scientific research, demoralizing dedicated educators, and pressuring schools to abandon programs that have proven successful for ELLs over the long term.

After just two years, NCLB is clearly failing to meet its goals. By setting arbitrary and unrealistic targets for student achievement, this accountability system cannot distinguish between schools that are neglecting ELLs and those that are making improvements. As achievement targets become increasingly stringent, virtually all schools serving ELLs are destined to be branded failures. The inevitable result will be to derail efforts toward genuine reform. Ultimately, a misguided accountability system means no accountability at all.

No Accountability Without Valid Assessments

Many of NCLB's shortcomings for ELLs can be traced to its failure to consider what is unique about these children. Setting benchmarks for student achievement, testing the progress of students against these benchmarks, then punishing schools where students fail is a simple, straightforward approach to accountability. It is also plausible and easily understood by the public. Yet, for ELL students in particular, it is an inappropriate, unworkable, and inequitable approach.

To succeed in school, ELLs must master academic knowledge and skills at the same time they are acquiring a second language. This is not an easy task. Nor is it a simple matter to monitor their progress, because existing assessment tools are generally unable to separate language errors from academic errors (Hakuta, 2001). When measuring the progress of ELLs, little confidence can be placed in tests that assume a mastery of English skills and that were never designed with ELLs in mind. This principle holds true not only in reading/language arts assessments but in mathematics assessments as well (Hakuta and Beatty, 2000). English-language achievement tests may be valid and reliable² for some ELLs, but not for others; the point is that no one can say with certainty. Research remains extremely limited on the level of English proficiency that students need to participate in the same assessments administered to native English speakers (August and Hakuta, 1997). Nevertheless, under Title I regulations proposed by the U.S. Department of

² Validity refers to whether assessments actually test what they are designed to test (e.g., whether results are distorted by language barriers). Reliability concerns the accuracy and consistency of assessment outcomes (e.g., whether results vary because of unrepresentative sampling of the populations being tested).

Education, ELLs must be tested in mathematics from day one and in reading/language arts after just 10 months in American schools. This is an arbitrary determination, without scientific support. Until appropriate assessments are widely available, it will inevitably yield inaccurate data about the quality of ELL programs.

Sometimes accommodations are provided, such as translations of test questions into the native language or the use of simplified English, which tend to raise the scores of ELLs (Abedi, 2004). Nevertheless, the extent to which English-language tests with accommodations fully measure student learning remains to be determined, especially for students just beginning to acquire English.

A provision of NCLB allowing states to test ELLs in their native language for up to three years (or five years on a case-by-case basis) appears to add a measure of flexibility to the system. In reality, it does little to mitigate the validity and reliability problem. Native-language assessments are often unavailable and are rarely aligned with state standards. Some merely translate English-language tests into Spanish, a procedure that psychometricians consider invalid because, among other things, the difficulty of vocabulary tends to differ across languages (August and Hakuta, 1997). Native-language tests are also inappropriate for students who are taught primarily in English and have limited literacy development in their first language.

Thus it is fair to say that existing instruments for assessing the academic achievement of ELLs, whose validity and reliability are questionable at best, cannot be counted on to generate meaningful information for accountability purposes. Yet state plans approved under NCLB rely heavily on such achievement tests – in many cases, a single test that is largely incomprehensible to many ELLs and unable to measure what they know.

No Accountability Without Reasonable Expectations

In NCLB's accountability system, the ELL subgroup itself is a problematic construct. This is a highly diverse population in terms of socioeconomic status, linguistic and cultural background, level of English proficiency, amount of prior education, and instructional program experience. It is also a highly fluid population, as newcomers enter – often speaking little English – and others leave after being reclassified as fully proficient in English; a common exit criterion is the 36th percentile in English reading/language arts. In other words, ELLs are *defined* by their low achievement level. When they have learned English, they exit the subgroup and their scores are no longer counted in the computation of adequate yearly progress (AYP).³ So it not merely unrealistic – it is a mathematical impossibility – for the ELL subgroup to reach full proficiency, as required by NCLB. It hardly makes sense to “hold schools accountable” for failing to achieve the impossible. Indeed, lumping virtually all schools with significant ELL enrollments in the same “needs improvement” category would defeat the purpose of accountability.

³ Citing a need for “flexibility,” the Department of Education has proposed to allow schools to count former ELLs in the subgroup for two years after they are reclassified as fluent in English. No doubt this will raise average scores for AYP purposes, but only to a limited extent. While it may postpone the date at which schools are defined as failing, it does not solve the problem.

Another difficulty in setting reasonable AYP targets is variability in the time it takes children to acquire a second language, especially the kinds of language needed for success in school. Research has shown that students in bilingual and English-as-a-second-language (ESL) programs require four to seven years to achieve grade-level academic performance in English (Hakuta, Butler, and Witt, 2000; Collier and Thomas, 1989). Another study found that a group of bilingual students in Arizona needed, on average, 3.3 years to acquire “native-like” oral proficiency in English. But the pace of acquisition varied widely, from one year to 6.5 years (Pray and MacSwan, 2002). In other words, scientifically based research shows there is no “standard” learning curve when it comes to second-language acquisition. As in other developmental processes, numerous hereditary and environmental factors are involved in learning a language. Among the most important is socioeconomic status; children from high-poverty, less educated backgrounds tend to need more time to acquire English (Hakuta, Butler, and Witt, 2000). ELLs are over-represented in this category. Thus they are likely to be disproportionately affected when AYP targets are set on an arbitrary basis.

Diversity and mobility in the ELL subgroup also complicate the task of determining reasonable rates of progress. A school that experiences a sudden influx of poor, immigrant children who speak little or no English – which is not uncommon – is likely to see a decline in its average ELL scores. If significant numbers of these students move elsewhere the following year – also common – scores are likely to increase. Neither result reveals anything meaningful about the quality of education offered to these students. Yet NCLB “holds schools accountable” for such yearly fluctuations. (As a statistical matter, the smaller the ELL subgroup, the wilder these random variations become; Abedi, 2004). Schools with ELLs from higher socioeconomic backgrounds and higher levels of English proficiency will find it easier to make AYP than those at the other end of the scale. But again, these patterns tell us little or nothing about how well ELLs are being served. In effect, schools are being “held accountable” for the demographic profile of their students.

The Proper Role of Assessments for ELLs

It is essential to develop high-quality assessments for ELLs – valid and reliable instruments to measure their academic achievement and their progress in acquiring English. Such assessments are needed to serve numerous purposes. These include:

- identifying students with limited English proficiency, placing them in appropriate instructional programs, and determining when they are ready to be reassigned to mainstream classrooms;
- evaluating alternative program models to gauge their effectiveness in serving ELLs;
- diagnosing student strengths and weaknesses to assist educators in improving instruction;
- tracking long-term trends of achievement in various groups and contexts; and
- holding schools accountable for student performance (Hakuta and Beatty, 2000).

Some progress is being made, especially in developing assessments of English-language proficiency.⁴ Many of these tests still suffer from “contamination” by extraneous factors – that is, most tend to test academic development as well as language development to varying degrees. Nevertheless, existing ELL assessments can be useful, if used appropriately, for certain purposes. At the present time, school accountability is not among these appropriate uses – at least, not the kind of accountability enforced by NCLB, with punitive consequences for schools based on ELLs’ test performance.

Consequences of a Misguided Accountability System

A broad consensus has emerged among testing experts: *achievement tests of questionable validity and reliability – or, indeed, a single test of any kind – should not be used for high-stakes decision-making* (Gottlieb, 2003). The focus of concern has been primarily on decisions involving individual students, such as grade promotion and graduation. ELLs are at a huge disadvantage where test results are employed in this fashion. Given the widespread inequities in resources available to schools where minority students are concentrated, high-stakes testing has spawned civil-rights litigation in several states.

By contrast, the policy of punishing schools on the basis of unreliable scores on a single test has received limited attention. What is often overlooked is that high stakes for schools are in many ways high stakes for children as well. As noted above, NCLB’s misguided approach to accountability is likely to be indiscriminate in identifying “failing schools.” It cannot benefit students to stigmatize and dismantle good programs along with the bad.

Equally pernicious are schools’ responses to the *threat* of labels and sanctions. If educators know that their careers can be jeopardized by results on a single round of achievement tests covering just two subjects, they will tailor instruction accordingly. Education will be reduced to language arts, mathematics, and of course, large doses of test preparation. Indeed, as reports from classroom teachers make clear, these consequences are already beginning to be felt. Well before the most punitive features of NCLB are scheduled to take effect, music, art, physical education, even social studies are being eliminated in many schools.⁵ Few of these are well resourced schools that teach socially advantaged students. The curriculum is being impoverished primarily in schools that enroll large numbers of poor, minority, disabled, and ELL students: the “problem subgroups.” Ironically, in the name of high standards, these children are being fed a steady diet of basic skills.

For ELLs in particular, this marks a giant step backward. The Improving America’s Schools Act (IASA) of 1994 had broken with the compensatory, remedial mindset in serving these children. For the first time it gave priority in federal funding to ELL programs whose goals included proficient bilingualism and biliteracy along with academic achievement in English. The law reflected research findings demonstrating that the most effective approaches seek to develop rather than replace the native-language

⁴ Some efforts in this direction have been funded under Title III of NCLB.

⁵ It appears that science is more often spared because it will become a high-stakes subject in 2007.

skills that ELLs bring to school, while teaching English through academic content rather than through instruction in discrete language skills (Ramírez et al., 1991). Over the last decade IASA helped numerous school districts build their capacity to offer *developmental* bilingual education programs, many of which have proven highly successful. NCLB, by contrast, eliminated the goal of bilingualism – in fact, any mention of the concept – instead narrowing the federal role to “curing” students of their limited English proficiency.

One example of the policy impact was reported earlier this year in Montgomery County, Maryland. The casualty was a two-way bilingual education program at the Highland Elementary School, which was designed to cultivate bilingualism in both native English speakers and native Spanish speakers. Children took classes in both languages, learned a full range of challenging subjects, and served as language models for each other. This approach is both popular with parents and academically promising, according to numerous studies. Yet the school district, concerned about ELL reading scores, mandated a 2½-hour block of English phonics each day, thoroughly disrupting the program. This decision was not mandated by NCLB,⁶ which does not explicitly require English-only instruction. Yet it was a direct result of district administrators’ worries about making AYP on English-language achievement tests. To Highland parents who objected that their children’s program was being dismantled to focus on basic skills, the superintendent responded: “Once they learn the fundamentals of reading, writing and math, they can pick up science and social studies on the double-quick” (Perlstein, 2004). This approach does not reflect what is known about best practices for ELLs. Moreover, it was condemned by an important civil-rights decision (*Castañeda v. Pickard*, 1981).

As a result of NCLB, similar decisions are being made in districts across the country. Pressures for the ELL subgroup to reach arbitrary targets – as measured by questionable achievement tests – are guiding local policymakers, rather than research-based practices. This is not where “accountability” should lead.

Authentic Accountability for ELLs

What matters most in the final analysis is not the progress of the ELL subgroup, but the progress of individual ELLs. A fair, reasonable, and useful accountability system would track cohorts of students to gauge their long-term academic achievement. It would use multiple measures, including grades; graduation, promotion, and dropout rates; and alternate forms of assessment.⁷ It would be accountable to local parents and communities, not just to top-down directives. Finally, it would consider a school’s “inputs” in serving ELLs, such as program design and teacher qualifications, rather than merely “outputs” – test scores alone.

⁶ It was, however, supported by a Reading First grant.

⁷ The state of Wisconsin, for example, uses assessments based on alternate performance indicators for ELLs to measure progress toward meeting standards in English language arts/reading, mathematics, science, and social studies. “Teachers collect classroom evidence from students in each content area that is scored by universal, content-based rubrics designed specifically for English language learners” (Gottlieb, 2003). So far, Wisconsin’s approach is the exception.

There is no question that schools' performance in educating ELLs requires close scrutiny. Services for these students remain inadequate in many districts, especially in parts of the country only recently impacted by immigration. School officials have often been slow to respond to cultural and linguistic diversity, to recognize the unique needs of ELLs, and to adapt instructional practices accordingly. They should be held accountable for providing equal opportunities for these students. But judgments about school performance should be broad-based and well informed. Indicators of progress, or lack thereof, should be not only accurate but also sensitive enough to assist in the process of school improvement. NCLB's simplistic approach fails ELLs on all of these counts.

Fortunately, a more promising framework for accountability already exists. Known as the *Castañeda* standard, it provides a proven set of tools for determining whether schools are meeting their obligations toward limited-English-proficient students. For two decades it has guided enforcement activities by the Office for Civil Rights of the U.S. Department of Education. The framework, first outlined by a federal appeals court, established a three-prong test to gauge whether school districts are taking "affirmative steps to overcome language barriers," as required by federal law.⁸ The court ruled that "good faith" efforts are insufficient. In serving ELLs, schools are obligated to meet three standards:

- Programs must be based on an educational theory recognized as sound by experts.
- Resources, personnel, and practices must be reasonably calculated to implement the program effectively.
- Programs must be evaluated and restructured, if necessary, to ensure that language barriers are being overcome.

Castañeda thus offers a comprehensive approach to school accountability, encompassing both inputs and outputs. Its broad focus includes instructional quality, teacher qualifications, language assessment and placement, classroom materials, and student outcomes. It emphasizes capacity-building, requiring districts to address the specific needs of ELLs, while allowing them the flexibility to choose programs suited to local conditions and preferences. It stresses not merely the development of English language skills but also students' progress in reaching academic standards (Hakuta, 2001). And it emphasizes instructional reform – getting to the roots of underperformance – rather than imposing punitive sanctions for failing to reach arbitrary AYP targets.

Where the *Castañeda* standard has been applied by federal courts and the Office for Civil Rights, results have often been promising. Districts have been required to initiate serious capacity-building efforts for serving ELLs, sometimes with federal funding under Title VII of IASA.⁹ The problem is that *Castañeda* has been applied on a very small scale, owing to political resistance and limited resources for enforcement. Moreover, the program evaluation component of Title VII was never adequately funded; nor were its

⁸ The law was the Equal Educational Opportunity Act of 1974, a codified version of the U.S. Supreme Court's decision in *Lau v. Nichols* (1974); the decision was *Castañeda v. Pickard* (1981).

⁹ Also known as the Bilingual Education Act of 1968, which was replaced by Title III of NCLB.

provisions for professional development. NCLB exacerbated the situation by eliminating requirements for evaluating ELL programs altogether and capping funds for professional development at less than half the FY 2001 level.

As a result, *Castañeda* has thus far played a relatively limited role in improving the education of ELLs. Yet there is no reason why this framework could not be successfully used in a comprehensive school accountability system. Under federal court orders, states including Illinois and Florida are already providing this type of oversight to ensure that districts are adequately serving ELLs (*Gómez v. Illinois State Board of Education*, 1987; *LULAC v. Florida Board of Education*, 1990). The principles of *Castañeda* should be developed, refined, and extended for use in all state accountability plans.

Recommendations for Reforming NCLB

School accountability for ELLs should be authentic, comprehensive, and oriented toward reforming instruction to reflect what is known about best practices in the classroom.

Toward that end, the following recommendations should be considered:

- Until assessments for ELLs have been proven valid and reliable, they should never be used to make high-stakes decisions for students, educators, or schools. Meanwhile, the federal government should substantially increase funding for scientific research in ELL assessment.
- AYP should not be calculated for an ELL subgroup. Instead, the progress of ELLs toward English proficiency and high academic standards should be tracked on a longitudinal, cohort basis. Arbitrary achievement targets – not based on scientific research – should never be used.
- ELLs’ achievement should be measured using multiple indicators, including grades, graduation and dropout rates, and alternate forms of assessment. Local authorities should be responsible for deciding, on a case-by-case basis, when ELLs are ready to be assessed in English and what test accommodations may be used. ELLs should never be required to take standardized tests that have not been normed for children whose English is limited. The most important goal of assessment should be to help educators improve instruction and students achieve long-term academic success.
- Accountability should concentrate on building schools’ capacity to serve ELLs, not on stigmatizing labels or punitive sanctions. Sanctions should only be used as a last resort, as a response to clear resistance to school improvement.
- Schools should be accountable to all stakeholders – in particular, local parents and communities, who should play an active role in accountability systems. Efforts should be required to facilitate the participation of limited English speakers.
- Accountability for serving ELLs should consider both *inputs* and *outputs*, using the *Castañeda* framework to determine (1) whether schools are providing well

designed instructional programs based on sound theory; (2) whether programs are supported with sufficient funding, qualified teachers, appropriate assessment and placement, and adequate materials; (3) whether programs are evaluated comprehensively for effectiveness; and (4) whether programs are being restructured, when necessary, to ensure that students are acquiring high levels of English proficiency and academic achievement.

References

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33, no 1: 1–14.
- August, D. and Hakuta, K., eds. 1997. *Improving schooling for language-minority students: A research agenda*. National Research Council. Washington, DC: National Academy Press.
- Castañeda v. Pickard*. 648 F.2d 989 (5th Cir. 1981).
- Collier, V.P. and Thomas, W.P. 1989. How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students*, 5 (Fall), 26–39.
- Gómez v. Illinois State Board of Education*. 811 F.2d 1030 (7th Cir. 1987).
- Gottlieb, M. (2003). *Large-scale assessment of English language learners*. Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Hakuta, K. (2001). The education of language-minority students. Testimony to the U.S. Commission on Civil Rights. Washington, DC, April 13.
- Hakuta, K., and Beatty, A., eds. (2000). *Testing English-language learners in U.S. Schools: Report and workshop summary*. National Research Council. Washington, DC: National Academy Press.
- Hakuta, K., Butler, Y.G., and Witt, D. (2000). *How long does it take for English learners to attain proficiency?* Santa Barbara, CA: Linguistic Minority Research Institute.
- Lau v. Nichols*. 414 U.S. 563 (1974).
- LULAC et al. v. Florida Board of Education*. Case No. 90-1913 (S.D.Fla. 1990)
- Perlstein, L. (2004). School pushes reading, writing, reform: Sciences shelved in efforts to boost students to “No Child” standards. *Washington Post*, May 31, p. A1.

Pray, L.C., and MacSwan, J. (2002). Different question, same answer: How long does it take for English learners to acquire proficiency? Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 4.

Ramírez, J.D., Yuen, S.D. and Ramey, D.R. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children*. San Mateo, CA: Aguirre International.