

The Use of High School Exit Examinations in Four Southwestern States

Paula García
Northern Arizona University

Abstract

Some 25 U.S. states have implemented or plan to implement a policy requiring that students pass a basic-skills test to graduate from high school. These tests, designed for native English speakers, have doubtful validity and reliability when administered to the English language learner (ELL) population, which is growing every year. Additionally, requiring students to pass an examination in English for high school graduation has two major consequences. To the individual, failure of the test could result in denial of gainful employment. To the larger society, an incidental English-only policy is implemented. This paper reviews some of the validity issues associated with high school exit exams used with ELL students and some of the solutions proposed by researchers and educators. This paper highlights how high-stakes testing has been implemented in four southwestern states with large ELL populations: Arizona, California, New Mexico, and Texas.

Introduction

In the United States, the use of standardized tests as a high school graduation requirement emerged from the push to hold schools accountable for adequately preparing students for the workplace (Rivera & Vincent, 1997). Since a 1983 report by the National Commission on Excellence in Education alerting lawmakers and the public to the increasing failure of public high schools to prepare young people for the workplace, 25 states have implemented or plan to implement high school exit exams. These exams, developed for the native English-speaking student population without addressing the unique linguistic and cultural issues that affect English language learners (ELLs), have the potential to deny ELLs a high school diploma. In this paper, the use

of high school exit exams in four southwestern states, which historically have had large multilingual populations, is examined as a test validity issue that has important consequences for ELLs.

Literature Review

Standardized Testing: Pros and Cons

The term “standardized test” refers to a test that has been screened for reliability and validity on a large population, and calibrated on the group of test takers for whom it is intended (National Research Council [NRC], 1999). High-stakes standardized tests have been associated with the “washback” phenomenon (e.g., Shohamy, Donitsa-Schmidt, & Ferman, 1996; Wall & Alderson, 1993), educational equality (e.g., Gipps, 1999), school improvement (e.g., Hoff & Manzo, 1999), and identification of student needs (Valdés & Figueroa, 1994).

Standardized tests, like all tests, have an inevitable impact on teaching known as “washback.” Washback can be positive when it improves the instruction that students receive or negative when it limits teaching to the content of the test (Hamp-Lyons, 1997; Messick, 1996). Standardized testing has been blamed for inducing negative washback because many schools and teachers, overly concerned with raising test scores, often reduce instruction to drill and practice of discrete facts and abandon the development of critical-thinking skills (Haladyna, 1992; Messick, 1996; NRC, 1999; Shepard, 1997).

Despite possible harmful effects from negative washback, standardized testing has been associated with effective educational reform. First, some educators believe that the use of standardized tests has the effect of promoting educational equality by standardizing curricula across school districts (Gipps, 1999). The United States does not have a centralized national curriculum; instead, high-stakes testing practices are used to make states accountable and control what is taught. Second, standardized testing has the effect of alerting the public to test-score disparities among districts that have varying socioeconomic situations, thus signaling unfair differences among schools and school districts. When such disparities are reported, the public often calls for educational reform. Therefore, many advocates of equality in education support standardized testing practices (e.g., American Educational Research Association, 2000). Third, there has been some evidence of school improvement related to standards-based reforms. For example, Colorado, Connecticut, and Kentucky, each of which has built reforms around standards and assessment, reported significantly improved scores on the National Assessment of Educational Progress fourth- and eighth-grade reading exams in 1992 and 1994 (Hoff & Manzo, 1999).

Standardized testing can benefit individual students by identifying those in need of academic intervention (Valdés & Figueroa, 1994). Many high school exit exams are administered in Grade 10 or earlier so that students who need to improve their skills can be identified for additional remedial work while there is still time to raise their achievement level before graduation in Grade 12. Virginia and Maryland begin administering the high school test in sixth and seventh grades respectively, with the intention that the standards will influence the elementary schools to improve instruction (Rivera & Vincent, 1997). Of course, such a system only works if students actually receive the necessary additional instruction to help them pass the test. If a school system lacks a clearly defined policy on how failing students will receive the necessary services to bring their skills to the level needed to pass the test, the test remains purely punitive and disciplinary (NRC, 1999).

Although standardized testing practices can benefit school systems and individual students, there is one group of students who are regularly overlooked in attempts at educational reform: ELLs. With the proliferation of standardized tests as graduation requirements, this ever-growing group of students faces special problems.

Problems of Validity in Assessing ELLs

When interpreting the scores of ELLs, educators need to be aware of specific problems with construct and instructional validity, reliability, and other influences that affect performance. Construct validity centers on the use of test scores to make inferences regarding the abilities of the intended test takers. Instructional validity deals with the students' opportunity to learn the content of the test. Without evidence of validity, it is inadvisable to trust inferences based on the scores.

Construct validity is problematic for tests administered to ELLs because most standardized tests used in the public schools were not developed with this population in mind (Valdés & Figueroa, 1994). Construct definitions of academic standards are usually not designed to include a second language as part of the construct and therefore cannot account for ELLs' performance on the test. According to guidelines of psychometric practices recommended by groups such as the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council of Measurement in Education (NCME), test items are assumed to be unidimensional (AERA, APA, & NCME, 1999; NRC, 1999); that is, each item should measure one attribute or trait, in order to support construct-related validity. When an ELL takes an academic achievement test in English, the test items measure not only the test taker's content knowledge, but also his or her English-language ability. This means that test items measure more than one skill, and therefore, it is difficult to interpret the scores. High school exit

exams, which usually focus on reading, writing, and mathematics, measure both content knowledge and English-language ability when administered to ELLs. Thus, this raises serious ethical questions when test scores are used for high-stakes decisions, such as the award of a high school diploma.

Instructional validity must also be considered because ELL students are often taught in different classes from those attended by native English speakers. Many schools have sheltered instruction or pull-out programs in which ELLs are taught by English as a Second Language teachers all the time or for a certain amount of time every day. Therefore, ELLs are often not exposed to the same instruction as their non-ELL counterparts (Geisinger, 1992; Valdés, 2001), yet the same instruments used for the non-ELL students are used to assess ELLs' abilities. Wang and Goldschmidt (1999) found evidence of the effect of differential educational opportunities on achievement. In their study of over 2,000 California sixth, seventh, and eighth graders of various ethnicities, they found that lower ELL test scores could be attributed to lack of opportunity to learn. Non-ELL students were enrolled in more advanced classes and had higher levels of achievement. This study provides support for the importance of instructional validity. If ELL students are enrolled in classes for which achievement tests are not designed, they will perform poorly on the tests because they do not receive the instruction that the test evaluates.

Reliability, another indicator of test validity, is also problematic when ELLs are involved because statistical reliability estimates are commonly calibrated on native English speakers. For example, Texas describes its field testing and data review of newly developed tests as including African American and Hispanic students in samples that are proportional to their total student populations in Texas (Texas Education Agency, 2001b). However, ELL students, who make up 14% of the total student population in Texas, are not included in the field-testing sample, and therefore are not included in estimates of reliability. This is one example of the ways in which many tests do not demonstrate reliability for the ELL population.

Reliability is an important indicator of validity because it indicates a test's consistency in measuring test takers' abilities. A study released by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) found that ELL scores in two large-scale multi-site standardized test administrations were lower than non-ELL scores, and that the gaps between the scores increased as the grade level advanced (Abedi, Leon, & Mirocha, 2000). These differences appeared in all subject areas: math, language, social studies, science, and especially reading. The smallest gap appeared in math computation, which was the content area with the lowest level of language load. Standardized tests administered to Grades 2, 7, and 9 proved much less reliable in the case of the ELL students. The researchers attributed the lower reliability for ELLs to linguistic complexity, which adds another source of measurement error to the test. The study showed that standardized assessments

disadvantage language-minority students because they do not assess their academic abilities reliably, which impacts test validity. In other words, when used to interpret the academic abilities of ELLs, the tests do not demonstrate evidence for validity. Using standardized testing on ELLs is problematic for two reasons: (a) it adds another dimension of cognitive ability (i.e., the second-language dimension) and (b) this added dimension was not intended to be measured by the test developers because the tests were created with non-ELLs in mind.

Another problem in interpreting ELLs' test scores is test score pollution. Haladyna (1992) points out three sources of test score pollution: test-taking strategy training, situational factors, and external conditions. Test score pollution attributed to test-taking strategy training and situational factors such as test anxiety and motivation are applicable to all types of students. The ELL population is especially susceptible to external factors, such as language deficits, socioeconomic context, and family mobility, all of which can affect test performance. These factors raise further concern about the interpretation of test scores of ELLs. One study by García (1991) showed the differences between the test scores of 51 Spanish-speaking Hispanic children and 53 native English-speaking children. The study found that the Hispanic fifth and sixth graders achieved significantly lower scores than other children at the same grade level on a reading comprehension test because of differences in background knowledge and test-taking strategies, not due to lack of knowledge or ability.

The Consequences of High School Exit Exams

Testing experts (e.g., Cronbach, 1989; Linn, 1998; Messick, 1989; Shepard, 1997), and more recently language testers (e.g., Hamp-Lyons, 1997; McNamara, *in press*), have been paying more attention to the aspect of validity related to the social consequences of test use. The consequential aspect of validity focuses on test use as one of several indicators of test validity. If the use of a test has negative social consequences, the test's validity is called into question. In the case of ELLs and high school exit exams, there are two main social consequences: the direct impact on the lives of individual students, and the incidental implementation of English-only language policy.

In standardized testing situations in the lower grades, there are few direct consequences to individual students. However, at the high school level, the consequences of failing a high school exit exam greatly impact the student. Without a high school diploma, students may be denied jobs or entrance to higher education or vocational programs. Such consequences leave those who fail high school exit exams, including ELLs, with few options to make a living, thus creating a permanent lower class of uneducated, and possibly unemployed, citizens and residents.

A larger picture of the consequences of English-language high school exit exams is the creation of an incidental English-only policy. Dery (1999) labeled the incidental creation of a policy caused by the implementation of another seemingly unrelated policy as “policy by the way.” Kaplan and Baldauf (1997) point out that linguistic unity created through the unintentional suppression of minority languages may be due to benign ignorance on the part of language policymakers. Incidental or intentional, English-only policy has been argued to affect minority-language speakers in a variety of ways: by decreasing their income potential (García, 1995), by denying them freedom to use their own languages in our democratic society (De Villar & Sugino, 1999), and by infringing on their linguistic human rights granted by the United Nations (Skutnabb-Kangas, 2000). It is important to keep in mind that the effects of an incidental English-only policy will not be known until the exit exams have been in place for some time; therefore, we can only speculate about the possible long-term effects. We can look at the case of German language loss and suppression through systematic anti-German policy imposed early in the 20th century as an example of how a minority language diminishes to near oblivion when that language is barred from use in schools (Wiley, 1998). Educators need to be aware of the larger social consequences of high school exit exam policies, and include them in the evaluation of validity.

Recommendations for Policy Decisions

In response to the problem of assessing ELL students with standardized tests, research organizations and scholars have presented a variety of recommendations such as exemption from taking the test, test accommodations (e.g., extra time, use of bilingual dictionaries), test translation, and administration of a language readiness test.

Test exemption may mean excluding ELLs from taking the test or disaggregating their scores from the native English-speaking population (LaCelle-Peterson & Rivera, 1994; Valdés & Figueroa, 1994). However, if ELLs are excluded from testing, they will not be identified for intervention programs. Many high school graduation requirement tests are given 2 to 3 years prior to graduation. ELLs who cannot pass the test at these earlier stages can benefit from more intensive study to help them pass at a later stage. Also, if assessment is used to make schools accountable, then schools that exclude ELLs from testing are not being held accountable for the achievement of ELLs. Schools need to demonstrate that they are providing ELLs and non-ELLs with educational equality to teachers, administrators, parents, and the public. This would not be accomplished if ELLs are exempted from testing.

Another option adopted by several state education agencies (SEAs) is to accommodate ELLs through the provision of bilingual dictionaries, simplified test instructions, or the provision of extra time (Butler & Stevens, 1997; NRC, 1999). These accommodations are intended to reduce the language load for

ELL test takers. In the case of bilingual dictionaries, ELLs may be aided by looking up specific English vocabulary, but they also could be disadvantaged if they have not mastered dictionary skills or apply the wrong meaning to a word. The use of various types of accommodations still affects the testing situation for ELLs and contributes to test score pollution (Haladyna, 1992). When some test takers are granted accommodations and others are not, test scores are no longer comparable (Rivera & Vincent, 1997). Some experts caution that such modifications are not the best answer to making standardized tests more usable for this complex population.

A third option is to provide a translation of the test in the native language of the ELL. This also is a questionable solution because two languages almost never conform in both form and meaning. Therefore, a translation of a test will likely have a different difficulty level from the original (Valdés, Barrera, & Cardenas, 1984) and will have to be validated independently of the original test version. Another problem with assessing ELLs in their native language is that the students may not have received instruction or performed academic work in their native language. Since many ELLs have had limited or interrupted schooling due to family mobility (Haladyna, 1992), they may have had limited exposure to literacy in their native language. A further complication is that many school districts provide bilingual education only at the elementary level; thus, few high school ELL students receive instruction in their native language. Due to these factors, test translation provides a problematic solution because it presupposes that ELL students can use their native language to process school-related content.

A fourth option, recommended by a report from CRESST, is the development and use of an academic language measure that can determine ELLs' English-language readiness for standardized tests (Stevens, Butler, & Castellon-Wellington, 2000). CRESST researchers claim that such a test would help to assure the validity of standardized tests administered to ELLs because it would indicate that the ELLs are at a linguistic level similar to the native English-speaking test takers, thus reducing the effect of language ability on test scores.

Given the complexity of these recommendations, it is of little surprise that there is no uniform way of addressing the assessment of ELLs (Valdés & Figueroa, 1994). This study examines how four states are addressing ELL assessment in regard to high school exit examinations as graduation requirements.

The Present Study

This investigation compares the policies regarding ELLs in four southwestern states: Arizona, California, New Mexico, and Texas. These states require or plan to require passing standardized assessments for high school

graduation. These states were selected for this study because they are geographically similar and contain large populations of ELL students. In fact, these four states rank among the top seven in ELL student enrollment, with California and Texas ranking first and second respectively (Rivera & Vincent, 1997). Also, these states have rich multicultural heritages with prominent and politically active multilingual communities, mainly Spanish speaking.

Requiring students to pass a test, which in most cases is in English, for high school graduation has the incidental effect of creating English-only language policy. Dery (1999) warns of policies that are byproducts of other policies. The United States has never had an officially mandated national language; English has been considered the national language not by vote or law, but by social pressure (Ricento, 1998). In the American Southwest, Spanish has dominated at times, and a multiplicity of Native American languages (e.g., Apache, Hopi, Navajo) have had a strong, but diminishing, presence. English is the newcomer language, yet it presently dominates; policies such as English-language high school graduation exams reinforce its place in the southwestern social milieu. Two research questions guided this investigation:

1. What is the policy regarding the high school examination graduation requirement for ELLs in Arizona, California, New Mexico, and Texas? What testing accommodations are allowed, if any?
2. What pass-rate results for ELLs have been reported from the SEAs of these four states?

Methodology

Information regarding high school examination requirements for ELLs was collected from SEAs through their official Web sites. Each of the SEAs of these four states has a Web site that is accessible to the public. SEAs utilize Web sites as their main avenue for dissemination so that thousands of school principals, teachers, and local school boards across each state can access current information. The use of Web sites facilitates the dissemination of information, which is constantly changing and being updated, in a way that formal printing and publishing of information cannot. While questions may reasonably be raised about relying on information on the Internet, this approach did permit the inclusion of current information on state educational policies, which change regularly. Information was also obtained through the newspaper and through telephone conversations with staff members from their respective education agencies.

Results

Each of the four states is at a different stage of implementing high school exit exams. New Mexico has required a high school exit exam for the longest period of time but is currently in a state of transition. Texas has also had a requirement for some time, while Arizona and California are working toward full implementation. This section reviews background information, accommodation policies, and pass rates of each state.

New Mexico

The New Mexico State Department of Education (NMSDE) began requiring passage of the High School Competency Exam (HSCE) with the senior class of 1990 as part of a standards-based reform project. The goal of this reform was to redesign high school education to meet the needs of all stakeholders: students, community members, business, and industry. The policy expresses a desire to graduate high school seniors who have the skills that employers want. In order to meet these needs, NMSDE has implemented a variety of programs, such as alternative credit and school-to-work, meant to ensure that a high school diploma is a valued credential. The HSCE, which is administered in the 10th grade, is part of this high-standards goal (NMSDE, 1999). It consists of six subtests in reading, language arts, math, science, social studies, and writing.

ELLs make up about a third of the total student body in New Mexico. To serve this population of students, NMSDE allows accommodations that fall into three categories: presentation, setting, and timing and scheduling. The category dealing with test presentation includes oral reading of test directions and questions in English, clarification of English words, and use of bilingual dictionaries. The setting category includes small-group or individual test administration, seating in the front of the class, or administration in a bilingual classroom familiar to the student. Timing and scheduling accommodations allow students to take additional time to complete the test, schedule the test over several days, and take more frequent breaks while testing. Multiple accommodations are allowed in any combination. The purpose of these accommodations is to minimize ELLs' linguistic differences and help them demonstrate their true knowledge without providing an unfair advantage or changing the focus of what is being assessed (NMSDE, 2003). NMSDE makes clear that only those accommodations that have been approved by NMSDE may be used and only with students who have been identified as ELLs.

A final way that NMSDE accommodates ELLs is through the provision of a Spanish-language version of the HSCE. Since nearly half of New Mexico's student body is Hispanic, a Spanish-language translation seems like an obvious way to address student needs (NMSDE, 2001). However, in a typical year, only about 300 to 500 high schoolers statewide request the Spanish version.

Throughout the late 1990s, pass rates for the HSCE were fairly high, at 82% or above, for all students combined. This changed in 2001, when NMSDE raised the passing score from 150 to 175. Pass rates for non-ELLs sank to 64.4% in that year, and for ELLs the pass rate was 36.8%, compared with an 80% pass rate for ELLs in 2000 (NMSDE, 2000). Of course, this decrease in pass rates was to be expected with the increase of the pass cutoff score. In 2002, 66% of the non-ELL 10th graders passed, and 40.9% of the ELLs passed (NMSDE, 2002), reflecting a small increase from the 2001 pass rates. For those taking the Spanish version in 2002, only 50% of the 10th graders passed the math subtest, and 40% or below passed in all other subtests. Beginning in 2003, graduating seniors had to meet the new, higher score requirement; however, pass rates for graduates were not available to the public at the time of this writing.

Also in 2003, NMSDE field tested a new test aimed at assessing the skills and knowledge of graduates. The new test, called the High School Standards Assessment, was administered to 10th graders in May 2003 (the results are being analyzed by the test developer and are unavailable at this time). The purpose of this new test is to meet even higher standards than the HSCE. For the time being, all accommodations for ELLs will continue to be allowed, but there is no Spanish version of the new test planned (G. Pitzl, personal communication, August 7, 2003).

These upcoming changes reflect a transition period for NMSDE's standards-based reform, with higher passing scores and a new test in the works. Exactly how this new test will impact the lives of ELLs will not be known for at least a couple of years. New Mexico is not alone in transitioning to higher standards. Texas has also been in the process of implementing newly developed curriculum.

Texas

Texas implemented a high school exit exam with the graduating class of 1992. That test, called the Texas Assessment of Academic Skills (TAAS), was part of the Texas Education Agency's (TEA) criterion-referenced testing program that reflected a desire to show that students had more than just minimum basic skills. It was administered to students in the 10th grade and did not include accommodations or a Spanish version (TEA, 2001a). In 1999, legislation was passed that called for an expanded testing program to match a more developed educational standards initiative. This new program, still in the phases of implementation, is called the Texas Assessment of Knowledge and Skills (TAKS). Beginning with the graduating class of 2005, the TAKS will be administered in the 11th grade, and all Texas students will be required to pass it in order to receive a high school diploma. The TAKS has four parts: English language arts, mathematics, social studies, and science.

As with the TAAS, there is no Spanish version of the TAKS exit-level exam, nor are other accommodations allowed, such as English-language or bilingual dictionaries, translation, or simplification and rephrasing of test questions. TEA does have a test that is used to assess ELLs' English-language readiness. The Reading Proficiency Tests in English (RPTE) are designed specifically for ELLs and rate their academic reading ability as beginning, intermediate, or advanced. ELLs who rate as beginning or intermediate, and who have been enrolled in a school in the United States for less than 12 months, may postpone taking the TAKS. In fact, this one-time postponement applies to all ELL students who have attended high school in the United States for less than 12 consecutive months, but they must take it once that time period has expired. A student who fails the test can retake it as often as the test is administered by the school district (TEA, 2003a).

According to TEA, failure for ELLs is likely. In the May 2003 administration of the TAKS to 11th graders, only 15% of ELL students passed (TEA, 2003b), compared to a 39% pass rate for 2002's 10th-grade ELL students on the TAAS (TEA, 2002). These low pass rates, however, were reported across all Texas student populations, with only 49% of the total student body passing the TAKS. Perhaps the new learning standards are too high for Texas high school students, which has been a similar situation in the state of Arizona.

Arizona

The Arizona Instrument to Measure Standards (AIMS) was implemented as a catalyst to improve the state's public schools by making them accountable for high academic standards. The Arizona Department of Education (ADE) intended to require passage of the high school AIMS test, administered in the 10th grade, for high school graduation beginning with the class of 2001. However, miserable pass rates for students of all ethnicities and socioeconomic levels, along with protests from parents and students, urged ADE to push back the high school requirement in order to allow sufficient time to align the curriculum with the test (Kossan, 2003). The class of 2006 will be the first class required to pass all three sections—reading, writing, and math—in order to earn a high school diploma (ADE, 2002a).

ELL students are also required to pass the test in order to graduate, but they are allowed some accommodations. For example in the language arts section, ELLs are allowed to use bilingual dictionaries although their answers must be in English. Also, they can have the test administered to them individually or in small groups, with the instructions simplified or read repeatedly. For the math section, ELLs can have an interpreter, translated test items, individual and small-group administration, and translation dictionaries (ADE, 2002b). The high school AIMS test is available in English only, and no English-language readiness test is offered or planned for the future (A. McGee, personal communication, September 15, 2003).

Despite the range of allowable accommodations, ELLs, along with other Arizona high school students, had tremendous difficulties with the test. Scores from spring 2003 showed pass rates for 10th-grade ELLs of 7% on the math section, 13% on the reading section, and 22% on the writing section. Pass rates for 10th-grade non-ELLs were a bit higher: 36% in math, 59% in reading, and 67% in writing. However, ADE hopes that ELLs' scores will continue to go up with each administration, and is therefore committed to maintaining the requirement for the class of 2006. This commitment, despite low pass rates, resembles the situation in California.

California

California wanted all graduates to demonstrate knowledge of minimum standards for verbal and quantitative skills by taking a high school exit exam. Taken in the 10th grade, the California High School Exit Exam (CAHSEE) was meant to be required for graduation starting with the class of 2004. However, after receiving news of pass rates below 50% for all students combined and an independent study showing non-alignment of the exam with curricula (Wise, Harris, Sipes, Hoffman, & Ford, 2000), the California Department of Education (CDE) decided to postpone the CAHSEE requirement until 2006, meaning that 10th graders taking the test in spring 2004 will be required to pass it for graduation. Until then, the test will be revised and shortened in length by its developers (CDE, 2003).

In California, ELLs represent approximately 25% of all public school students. The only accommodation allowed to ELLs taking the CAHSEE is extra time (CDE, 2002). However, there is a relatively new test developed by the CDE for the purpose of assessing ELLs' English-language readiness. The California English Language Development Test (CELDT) assesses ELLs' reading, writing, listening, and speaking abilities in order to determine their readiness for mainstream classes. Although it is available for all grade levels including high school, there are no plans for the CELDT to be used as an English-language readiness test for ELLs who will need to pass the CAHSEE for graduation (M. Center, personal communication, September 5, 2003).

The CAHSEE is made up of two main parts: English language arts, which is comprised of reading and writing, and mathematics. Similar to high school exit exam programs in New Mexico, Texas, and Arizona, California also reported low pass rates for all students in 2003, but especially for ELLs. On the English language arts section, 54% of all students passed, compared with 28% of ELLs. On the math section, 32% of all students passed, but only 18% of ELLs passed.

Table 1

Summary of High School Exit Exams by State

Name of high school exit exam	Required for ELLs	Accommodations or other support for ELLs	Current status	Most recently reported pass rates for ELLs
New Mexico High School Competency Exam (NMHSCE)	Yes	A variety of accommodations dealing with presentation, setting, timing, and scheduling; phasing out Spanish version	In use, new exam field tested in 2003	40.9% on all test parts combined
Texas assessment of Knowledge and Skills (TAKS)	Yes	No accommodations; test of language readiness available	In use, phasing in new exam	15% on all test parts combined
Arizona's Instrument to Measure Standards (AIMS)	Yes	Bilingual dictionary, individual and small-group administration, simplified directions	Postponed as a requirement until 2006	7% on math 13% on reading 22% on writing
California High School Exit Exam (CAHSEE)	Yes	Extra time	Postponed as a requirement until 2006	18% on math 28% on English language arts

Discussion

In answer to the first research question regarding the high school exam graduation requirement for ELLs in the four southwestern states of interest, the results show that Texas and New Mexico currently require ELLs to pass the exam in order to receive a high school diploma. Also, both Texas and New Mexico are in the process of increasing the difficulty level of their respective tests. Arizona and California intend to require the exam after aligning their tests with curricula. Regarding accommodations, results show that currently New Mexico and Arizona offer flexibility for ELLs in terms of time, simplification, and setting of test administration. California allows extra time but this state, like Arizona, has not fully implemented the requirement of passing the exam to graduate, and this accommodation may be eliminated in

the future. Texas offers no accommodations. New Mexico, the only state to offer an exit exam in a language other than English, will discontinue this option once a new test is in place (see Table 1).

One interesting finding was that two states have an English-language readiness test available, as recommended by CRESST (Stevens et al., 2000) and discussed earlier in this paper. Texas has already integrated such a test, the RPTE, into its high school exit exam program. California has chosen not to use its CELDT as an indicator of ELLs' readiness to take the exit exam. This may be because the test is still in the development phase, or because CDE feels that the purpose of the CELDT is to assess ELLs' readiness for the mainstream classroom and should not be applied to other uses. Regardless, the trend of developing English-language readiness tests show that SEAs are becoming more sensitive to the special needs of their ELL student population.

In answer to the second research question about what pass rates have been reported, it was discovered that Texas reported the lowest pass rates for ELLs and that it has the strictest requirements. New Mexico had the highest pass rates and is currently the most lenient, yet it is in the process of changing its test. California and Arizona had low pass rates for both ELL and non-ELL students on their pilot tests, but still plan to implement the tests with some revision and curricular alignment.

The implementation of high school exit exams invokes three major problems for ELLs in these southwestern states. The first problem, as discussed earlier, is that of invalid testing practices. Standardized tests assess more than content when administered to ELLs; they also assess language. This may be one reason why there are such low pass rates for ELLs.

The second problem is the social consequence of not having a diploma. The denial of a high school diploma is a very serious matter for U.S. citizens and residents. There is no alternative vocational education system such as those that exist in other developed and developing countries. For many Americans, a high school diploma is the only qualification to join the work force; this is why high school exit exams represent high-stakes testing practices. Professional organizations such as AERA (2000) strongly advise against using test scores as the sole basis for a high-stakes decision such as high school graduation. Course grades, project work, and other measures of achievement can be used in addition to a standardized test to help teachers decide whether a student has fulfilled state educational objectives.

A third problem is the possible simultaneous adoption of English-only language policy as a byproduct of education policy. This problem is most evident in Arizona and California, where "English for the Children" legislation has promoted an English-only stance. The implementation of a high school requirement demanding the use of English, as most exit exams do, is tantamount to making English the official language. The Southwest has historically been

multilingual, with Spanish and Native American languages present from the beginning. The loss of these languages in the public schools will continue to discourage multilingualism in this multilingual region.

In addition to the possible English-only implications, the implementation of high school exit exams embodies two main contradictions. First, when they were originally conceived, high school exit exam scores were supposed to indicate that graduates had the academic skills needed for the workplace. However, real-world job options at the high school diploma level mainly include vocational and service-oriented positions. Do high school graduates really need academic skills to perform non-academic jobs? Second, many SEAs implement exit exams in the 10th grade so that there is time for remediation if necessary; yet the test scores are interpreted to mean that students possess 12th-grade, or graduation-level, skills. Is there no advancement of skill level between the 10th grade and 12th grade? These two questions stated here relate to all high school students, not just ELLs, and therefore educators and administrators need to address them as the high school exit-exam trend continues.

Conclusion

This study highlighted the situation of ELL high school students in four southwestern states that require or will require passage of an exit exam for graduation. Results showed that ELLs have tremendous difficulty passing such tests. But results also show, in light of allowed accommodations and new language readiness tests, that these states are aware of the challenges that their ELL students face. In its document for implementing appropriate accommodations, NMSDE (2003) warns that little is known about the effects of accommodations on test scores because few studies have been done to investigate this aspect. This sheds light on the fact that SEAs are implementing testing policies when the effects of support systems, such as accommodations, are still being investigated.

The need for continued research on the effects of exit exams, accommodations, and language readiness tests cannot be more urgent. The consequences of not having a high school diploma are too great for SEAs to implement exit exams, and other policies that go along with them, as graduation requirements without knowing their full impact. One study (Jacob, 2001) on the general student population across 31 states showed that high school exit exams have no significant impact on reading and math achievement but are associated with increased dropout rates among low-performing students. Therefore, if high school exit exams are not serving the educational objective of improving basic skills for all students, then what purpose are they serving? Some may argue that they are only serving to further the English-only agenda as evidenced by the “English for the Children” movements in California and Arizona.

The issues presented here are just a starting point for investigating the effects of high school exit exams on ELLs. Perhaps after taking into account the impact on ELLs and the possible ineffectiveness of this high-stakes testing practice, educators and policymakers will rethink the continuation and expansion of high school exit exams.

References

- Abedi, J., Leon, S., & Mirocha, J. (2000). Examining ELL and non-ELL student performance differences and their relationship to background factors: Continued analysis of extant data. In Eva Baker (Principal Investigator), *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives* (pp. 3–49). Los Angeles: National Center for Research and Evaluation, Standards, and Student Testing, University of California.
- American Educational Research Association. (2000). Position statement of the American Educational Research Association concerning high-stakes testing in preK–12 education. *Educational Researcher*, 29(8), 24–25.
- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Arizona Department of Education. (2002a). *Summary report for Spring 2002 test administration*. Retrieved September 29, 2003, from <http://www.ade.state.az.us/standards/aims/Results/2002AIMSExecSum.pdf>
- Arizona Department of Education. (2002b, Spring). *LEP Guidelines: Arizona student achievement program accountability for all students*. Phoenix: Author.
- Butler, F., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (Technical Report No. 448). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- California Department of Education. (2002). *Questions and answers for administrators about the postponement of the CAHSEE requirement*. Retrieved September 30, 2003, from <http://www.cde.ca.gov/statetests/cahsee/updates/sbedecqa703.pdf>
- California Department of Education. (2003). *Cancellation of July 2003 test*. [Letter from Jack O'Connell, state superintendent of public instruction.] Retrieved September 30, 2003, from <http://www.cde.ca.gov/statetests/cahsee/updates/julycahsee.pdf>

- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147–171). Urbana: University of Illinois Press.
- Dery, D. (1999). Policy by the way: When policy is incidental to making other policies. *Journal of Public Policy*, 18(2), 163–176.
- De Villar, R. A., & Sugino, T. (Eds.). (1999). One world, many tongues: Special issue on language policies and the rights of learners. *TESOL Journal*, 8(3).
- García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, 26, 371–392.
- García, O. (1995). Spanish language loss as a determinant of income among Latinos in the United States: Implications for language policy in schools. In J. W. Tollefson (Ed.), *Power and inequality in language education* (pp. 142–160). Cambridge, England: Cambridge University Press.
- Geisinger, K. (1992). Testing limited English proficient students for minimum competency and high school graduation. In *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement* (Vol. 2, pp. 33–65). Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs.
- Gipps, C. (1999). Socio-cultural aspects of assessment. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education* (Vol. 24, pp. 355–392). Washington, DC: American Educational Research Association.
- Haladyna, T. (1992). Test score pollution: Implications for limited English proficient students. In *Proceedings of the Second National Research Symposium on limited English proficient student issues: Focus on evaluation and measurement* (Vol. 2, pp. 135–163). Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14, 295–303.
- Hoff, D. J., & Manzo, K. (1999, March). States committed to standards reforms reap NAEP gains. *Education Week on the Web*, 18(26). Retrieved September 29, 2003, from <http://www.edweek.org/ew/vol-18/26naep.h18>
- Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23, 99–121.
- Kaplan, R. B., & Baldauf, R. B. (1997). *Language planning from practice to theory*. Clevedon, England: Multilingual Matters.

- Kossan, P. (2003, August 31). AIMS causing worry, dread: Impact of test hitting Class of '06. *The Arizona Republic*, pp. A1, A8.
- LaCelle-Peterson, M. W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64(1), 55–75.
- Linn, R. L. (1998). Partitioning responsibility for the evaluation of the consequences of assessment programs. *Educational Measurement: Issues and Practice*, 17(2), 28–37.
- McNamara, T. F. (in press). Tearing us apart again: The paradigm wars and the search for validity. In S. Foster-Cohen & S. Pekarek Doehler (Eds.), *EUROSLA Yearbook* (Vol. 3). Amsterdam: John Benjamins.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241–256.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Research Council. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- New Mexico State Department of Education. (1999, August). *Strategic plan*. Retrieved September 30, 2003, from <http://www.sde.state.nm.us/sbe.desktop/dl/stratplan.pdf>
- New Mexico State Department of Education. (2000). *New Mexico statewide articulated assessment program: Phase II*. Retrieved September 30, 2003, from <http://www.sde.state.nm.us/div/ais/assess/dl/SumRepPhaseII.pdf>
- New Mexico State Department of Education. (2001, January). *Statewide—Data collection factsheets*. Retrieved September 30, 2003, from <http://www.sde.state.nm.us/div/ais/data/FS/state.st.factsheets.html>
- New Mexico State Department of Education. (2002, Spring). *Executive summary report for the New Mexico articulated assessment program*. Retrieved September 30, 2003, from <http://www.sde.state.nm.us/div/ais/assess/dl/exec.summary.hsce.10.23.02.pdf>
- New Mexico State Department of Education. (2003). *Appropriate accommodations for English language learners*. Retrieved September 30, 2003, from <http://www.sde.state.nm.us/div/ais/assess/ell.accomm.assess.10.31.02.htm>

- Ricento, T. (1998). National language policy in the United States. In T. Ricento & B. Burnaby (Eds.), *Language and politics in the United States and Canada: Myths and realities* (pp. 85–112). Mahwah, NJ: Lawrence Erlbaum.
- Rivera, C., & Vincent, C. (1997). High school graduation testing: Policies and practices in the assessment of English language learners. *Educational Assessment, 4*, 335–355.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practices, 16*(2), 5–8, 13, 24.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*, 298–317.
- Skutnabb-Kangas, T. (2000). Linguistic human rights and teachers of English. In J. K. Hall & W. E. Eggington (Eds.), *The sociopolitics of English language teaching* (pp. 22–44). Clevedon, England: Multilingual Matters.
- Stevens, R., Butler, F., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs*. Los Angeles: National Center for Research and Evaluation, Standards, and Student Testing, University of California.
- Texas Education Agency. (2001a). *Student assessment program technical digest chapter 1: Background*. Retrieved September 2, 2003, from <http://www.tea.state.tx.us/student.assessment/resources/techdig/chap1.pdf>
- Texas Education Agency. (2001b). *Student assessment program technical chapter 2: TAAS test development*. Retrieved September 2, 2003, <http://www.tea.state.tx.us/student.assessment/resources/techdig/chap2.pdf>
- Texas Education Agency. (2002). *TAAS summary report: Test performance*. Austin: Author.
- Texas Education Agency. (2003a). *LPAC decision-making process for the Texas Assessment Program (Grades 3–12)*. Austin: Student Assessment Division and Bilingual Education Program Unit, Texas Education Agency.
- Texas Education Agency. (2003b). *TAKS summary report: Test performance*. Austin: Author.
- Valdés, G. (2001). *Learning and not learning English: Latino students in American schools*. New York: Teachers College Press.
- Valdés, G., Barrera, R., & Cardenas, M. (1984). Constructing matching texts in two languages: The application of propositional analysis. *Journal of the National Association for Bilingual Education, 9*(1), 3–19.
- Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.

- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41–69.
- Wang, J., & Goldschmidt, P. (1999). Opportunity to learn, language proficiency, and immigrant status effects on mathematics achievement. *Journal of Educational Research*, 93(2), 101–111.
- Wiley, T. G. (1998). The imposition of World War I era English-only policies and the fate of German in North America. In T. Ricento & B. Burnaby (Eds.), *Language and politics in the United States and Canada: Myths and realities* (pp. 211–241). Mahwah, NJ: Lawrence Erlbaum.
- Wise, L., Harris, C., Sipes, S., Hoffman, R. G., & Ford, J. P. (2000). *High school exit examination (HSEE): Year 1 evaluation report*. Retrieved September 30, 2003, from <http://www.cde.ca.gov/statetests/cahsee/eval/2000/humrro1.pdf>

Acknowledgments

I would like to thank Mary McGroarty for valuable comments on earlier drafts of this paper. I would also like to thank the anonymous reviewers and editors of *Bilingual Research Journal* for contributing to its quality.