Academic Achievement of English Language Learners in Post Proposition 203 Arizona

by

Wayne E. Wright
Assistant Professor
University of Texas, San Antonio

Chang Pu
Doctoral Student
University of Texas, San Antonio

Language Policy Research Unit (LPRU)
Education Policy Studies Laboratory
College of Education
Division of Educational Leadership and Policy Studies
Box 872411
Arizona State University
Tempe, AZ 85287-2411

September 2005

EPSL | EDUCATION POLICY STUDIES LABORATORY
Language Policy Research Unit

EPSL-0509-103-LPRU
http://edpolicylab.org

Education Policy Studies Laboratory
Division of Educational Leadership and Policy Studies
College of Education, Arizona State University
P.O. Box 872411, Tempe, AZ 85287-2411
Telephone: (480) 965-1886
Fax: (480) 965-0303
E-mail: epsl@asu.edu
http://edpolicylab.org

Academic Achievement of English Language Learners in Post Proposition 203 Arizona

Wayne E. Wright
and
Chang Pu

University of Texas, San Antonio

Executive Summary

This report reveals the problems with claims made by Arizona state public education officials that English Language Learners (ELLs) are thriving under English-only instruction.

The No Child Left Behind Act of 2001 (NCLB) and the state's accountability system, Arizona LEARNS, require all students, including ELLs, to participate in statewide high-stakes testing. Test scores are the main measure of student achievement under these systems, and labels based on those scores are given to each school (i.e. Highly Performing, Underperforming, etc.). The state education administration's interpretation and strict enforcement of Proposition 203 has ensured that nearly all ELL students in grades K-3 are instructed through the English-only Sheltered English Immersion (SEI) model. They claim that SEI has led to better test scores and increased achievement among ELLs, using as evidence improved test scores and the decrease in the number of schools labeled as "Underperforming." However, analyses of test data for

students in grades two through five and changes in the state accountability system revealed the contrary; they exposed serious achievement gaps between ELLs and their counterparts, and proved that positive looking improvements in school accountability labels mask test-score decline in a large number of elementary schools.

From 2002 to 2004, students in Arizona were required to take two standardized tests: Arizona Instrument to Measure Standards (AIMS), a test given in grades three, five, eight, and high school that is designed to measure student achievement against state standards; and the Stanford Achievement Test *Ninth Edition* (Stanford 9), a test given in grades two through nine that is designed to measure student achievement against the national average. The state has divided test score data into two categories: ALL (Category 1) and ELL (Category 2). The labels are misleading: The ALL category excludes the scores of ELL students who have been enrolled in public school for less than four years, thereby excluding the scores of the ELL students with the lowest levels of English proficiency. The report's analyses focus mostly on third grade AIMS test scores and the Stanford 9 test scores of elementary school students as they progressed from one grade to the next between 2002 and 2004. The key findings are:

- The overwhelming majority of third grade ELLs fail the AIMS test in contrast to ALL students, and ELLs score well below the 50th percentile on the Stanford 9 and well below students in the ALL category.
- There is a general pattern of higher test scores on AIMS in 2003, followed by decline in 2004 for both ALL and ELL students on the Reading and Math subtests.

- ELL student percentile rankings on the Stanford 9 rose slightly in 2003 followed by a decline in 2004 while ALL student rankings remained relatively stable.
- Improvement in test scores in 2003 corresponds with a period of greater flexibility for schools in offering ESL and bilingual education, while the decline of scores in 2004 corresponds to a period of strict enforcement of Proposition 203 and mandates for English-only instruction.
- The sudden increase in 2004 of ELLs passing the AIMS Writing subtest is questionable, as there was decline or no significant growth on all other subtests for both the AIMS and Stanford 9, and as similar gains were not evident for ALL students.
- In terms of the percent passing the AIMS test, ELL students trailed behind ALL students by an average of 33 percentage points in Math, 40 points in Reading, and 30 points in Writing.
- On the Stanford 9, ELL students trailed behind ALL students by an average of 28 percentile points in Language, 26 points in Math, and 33 points in Reading. The gap increased for all Stanford 9 subtests between 2003 and 2004.
- The narrowing of the achievement gap in AIMS Reading and Math is actually a function of ALL student scores decreasing at a higher rate than decreases in ELL scores.

- ALL students score lower on the AIMS and Stanford 9 in ELL-Impacted elementary schools (schools that test 30 or more ELL students in third grade) than they do in other elementary schools.
- Lack of reliable data: There are discrepancies in the number of ALL and ELL students tested on the AIMS and Stanford 9 within each year and across the three years that are inconsistent with the rapidly growing student population of Arizona. This raises questions on whether some student scores are missing from the data reported to the public, or if students were systematically excluded from taking specific tests.

This report also analyzes the changes in school labels under Arizona LEARNS and NCLB between 2002 and 2004. In 2002, the Arizona LEARNS labels were: Excelling, Maintaining, Improving, and Underperforming. In 2003, the labels were changed to: Excelling, Highly Performing, Performing, Underperforming, and Failing. These labels are based primarily on the test performance of students in the ALL category, which excludes most ELL scores. An analysis of the numbers of schools in each category throughout this time period along with the test data for the corresponding years revealed the following:

- There were increases in the number of "Performing" and "Excelling" schools in 2004 despite the general trend of flat or declining AIMS and Stanford 9 scores.
- Arizona LEARNS labels and NCLB AYP designations are not reflective of a school's success (or lack thereof) with ELL students as these labels and

designations are based on ALL score data which excludes most ELL test scores.

- Improvements in Arizona LEARNS labels and NCLB's AYP designations are masking the harm that current state language and testing policies are having on ELL students.

Close monitoring of ELL test scores is needed by policy makers and relevant stakeholders. A system is also needed for mutually exclusive categories of ELL and non-ELL students, and mechanisms are needed to track the progress of ELL students even after they are redesignated as fluent English proficient. State policy makers are encouraged to reconsider the narrow requirements and current strict enforcement of Proposition 203. In addition, rather than forcing ELLs to take English-only high-stakes tests only to exclude many of their scores from state and federal accountability formulas, state policy makers are encouraged to advocate for changes in the requirements of NCLB, or at the very least, heed the federal law's requirement to test ELLs in the *language* and *form* most likely to yield valid and reliable information about what students know and can do.

Academic Achievement of English Language Learners in Post Proposition 203 Arizona

Wayne E. Wright

and

Chang Pu

University of Texas, San Antonio

Introduction

State education leaders in Arizona claim that education is improving in Arizona for children classified as English Language Learners (ELLs). This success is attributed to strict enforcement of Proposition 203 which requires that ELL students be instructed only in English through Sheltered (Structured) English Immersion (SEI).[1] The current Superintendent of Public Instruction and his appointed leaders, who supervise ELL programs in the Arizona Department of Education (ADE), claimed that bilingual education programs in the state were failing to teach English and were preventing academic success.[2] They are enforcing their own interpretation of Proposition 203 which makes it difficult for any Arizona school to offer bilingual programs, and nearly impossible for ELL students in grades K-3 to qualify for waivers as outlined in the law.[3] These leaders have claimed that their strict enforcement of Proposition 203 has removed the obstacles to ELL student success, and that English-only programs are now ensuring that ELL students will "soar academically."[4]

Success is also attributed to Arizona's school accountability system (Arizona LEARNS), the state's testing and accountability program. At the intersection of Proposition 203, No Child Left Behind, and Arizona LEARNS, ELL students are required to fully participate in statewide high-stakes tests in English, and schools are held accountable for the results.[5] Evidence for claims of ELL student success include rising test scores and the significant decrease in the number of schools labeled as "Underperforming" under Arizona LEARNS. The decrease in schools designated as "failing" to make Adequate Yearly Orogress (AYP) under No Child Left Behind also provides evidence for these claims of improvement.

This report provides data and analyses that reveal the problems with these claims and the evidence used to support them. After a brief overview of language and assessment policy in Arizona, this report analyzes student achievement and accountability data. Data analyses focus primarily on third grade, as this is the grade when students must first take the AIMS test, and it is one of the primary grades most affected by Proposition 203.

Analyses of student achievement and school accountability data are presented in two parts. In Part 1, analysis focuses on comparisons of statewide test score data and on changes in school achievement labels. In Part 2, similar analyses are conducted just on "ELL Impacted" elementary schools, that is, schools that have a significant number of ELL students (see below for selection criterion). We chose to focus on these elementary schools because they have experienced the greatest impact of state policies related to Proposition 203 and the requirements for testing ELL students. Analyses will also

explore the relationships between test-score trends and changes in school accountability ratings for these ELL Impacted schools.

Overview of Language and Assessment Policy in Arizona

Proposition 203

Proposition 203, "English for the Children," was passed by voters in November of 2001 and took effect at the beginning of the 2002-2003 school year. Proposition 203 requires that English Language Learners (ELLs) "be taught English, by being taught in English" and that they be placed in "English language classrooms" and "educated through sheltered English immersion."[6] The law lacks a clear definition of Structured English Immersion (SEI) and, to date, no operational definition of SEI has been provided by the Arizona Department of Education (ADE) other than simply requiring that ELL students be taught in English and that all instructional materials must be in English.[7] At present it is unclear how SEI differs from the sink-or-swim mainstream instruction declared unconstitutional under Lau v. Nichols.

Proposition 203 also mandates that ELLs in grades 2 through 11 be assessed annually in English on a norm-referenced test. During the years 2002-2004, the Stanford 9 was used for this purpose.[8] Prior to Proposition 203, state policy allowed schools to exclude many ELL students from the Stanford 9, or allowed schools to administer the Aprenda 2—the Spanish-language version of the Stanford 9—for their Spanish-speaking ELL students.[9]

The law also includes waiver provisions for parents who want their children in bilingual education programs. While these provisions were designed to be intentionally

difficult for parents to obtain and easy for schools and districts to deny,[10] they nonetheless make bilingual education possible for those parents who want it. Schools were initially provided with some flexibility in interpreting the ambiguous language of the law, as long as they followed proper procedures in the granting of waivers.[11] However, state policy changed with the election of a new Superintendent of Public Instruction in 2003, who joined forces with local leaders of English for the Children during his campaign and ran on a platform accusing his predecessor of failing to enforce Proposition 203.[12] After taking office, he appointed the local chairperson of English for the Children as an Associate Superintendent overseeing ELL programs in the state.[13] Together they issued new waiver guidelines[14] and hired monitors to visit schools to ensure compliance.[15] These efforts have succeeded in ending nearly all bilingual programs for ELL students in grades K-3 in the state.[16]

Arizona LEARNS and the No Child Left Behind Act

Arizona LEARNS was authorized by Arizona Revised Statutes (A.R.S.) §15-241 in 2001. Arizona LEARNS is designed to hold schools accountable by utilizing student achievement data. Prior to 2005, accountability formulas used data from the Arizona Instrument to Measure Standards (AIMS) test, the Stanford Achievement Test-9th Edition (Stanford 9), and the Measure of Academic Progress (MAP), to calculate school achievement ratings and to assign school labels. These labels are used to provide a system of rewards and sanctions for schools and teachers.[17]

AIMS is designed to measure student achievement in terms of meeting state academic standards in Math, Reading, and Writing. AIMS was first administered in the 1998-1999 school year and prior to 2005, it was only administered in grades three, five,

eight, and once in high school. Initial efforts to create a Spanish-language version of the AIMS came to end with the passage of Proposition 203.[18]

The high school AIMS test also functions as a graduation test.[19] However, the use of AIMS as an exit exam has been postponed several times due to substantially high failure rates.[20] Testing experts found that the state rushed the development and use of AIMS resulting in numerous problems, including: overly difficult items, testing students on material they had not yet had the opportunity to learn, errors on the test, ambiguous questions, errors in scoring, and inappropriately set passing scores.[21] As a result, the AIMS test has undergone numerous changes in an effort to make it "more reasonable."[22] As it currently stands, the Class of 2006 will be the first that must pass AIMS to receive a high school diploma.[23]

The Stanford 9 is a norm-referenced test, meaning that results are reported as percentile ranks which indicate how well students performed in comparison to a nationally-representative sample group (i.e., the norming population). The Stanford 9 was used in Arizona from 1997-2004.[24] Arizona students took the Math, Language, and Reading subtests of the exam. Unlike AIMS, no changes have been made to the Stanford 9. The Measure of Academic Progress (MAP), first used in 2000, was calculated using Stanford 9 scores, and attempted to measure growth over time.[25] While viewed as a fairer measure of progress, particularly for schools in low socioeconomic neighborhoods, calculating MAP was problematic for many inner-city and charter schools which traditionally have high rates of student mobility.

Arizona LEARNS required ADE to use data from AIMS and Stanford 9 (via the Measure of Academic Progress) to compile an "annual academic achievement profile"

and assign a label for each public school.[26] The labels have changed over time (see below), but essentially consist of a hierarchy of five classifications ranging from "Underperforming" to "Excelling"; schools obtaining a label of "Underperforming" for two consecutive years receive the label of "Failing." Underperforming and Failing schools must undergo a school improvement process with some assistance from a state-assigned "solutions team."[27] If a school continually fails to improve it is subject to state takeover.[28]

Arizona LEARNS closely mirrors the requirements of the No Child Left Behind Act (NCLB); the state accountability program, however, has undergone and is continuing to undergo a number of changes to come into full compliance with the federal law. NCLB requires the full participation of ELLs in the state's testing and accountability system. However, the law states that students must be "assessed in a valid and reliable manner and be provided with reasonable accommodations," which may include "assessments in the language and form most likely to yield accurate data on what such students know and can do in academic content areas."[29] As described previously, Proposition 203 ended efforts to create native-language versions of AIMS. ADE has not provided school districts with clear guidance on what constitutes "reasonable accommodations" for ELL students. While practice varies widely, it appears that most ELL students in Arizona are required to take the state tests without the benefits of the accommodations called for in the federal law.[30]

NCLB also requires the "disaggregation," or separation, of student achievement data, and requires that all subgroups make Adequate Yearly Progress (AYP) on state tests. Failure of any single subgroup can result in the entire school being designated as

"failing" to make AYP. Like Arizona LEARNS, No Child Left Behind also mandates a series of sanctions for schools that consistently fail to make AYP, including eventual state takeover or privatization of the school. The complexities of and controversies surrounding AYP requirements are beyond the scope of this report,[31] but it is sufficient to state here that schools with ELL students are under immense pressure to prepare their ELL students for and to raise their scores on the AIMS test (in English).

The Nature and Use of Disaggregated Data in Arizona LEARNS

Before analyzing the student achievement data, it is important to understand the nature and use of disaggregated student achievement data in Arizona. Between 2002 and 2004, Arizona achievement data has been separated and reported into just two categories: Category 1 "All Students" (ALL) and Category 2 "English Language Learners" (ELL). School accountability ratings and labels are based on complicated formulas using Category 1 test scores mainly from the AIMS test.[32] Category 1 test scores from the Stanford 9 via the Measure of Academic Progress (MAP) are also used in school accountability formulas. However, the name of Category 1 is misleading, as not *all* students are included as the name of this category indicates. Test scores of ELL students with less than four years of enrollment in school are excluded from Category 1 (ALL).[33]

The exclusion of these ELL scores from Category 1 (ALL) represents some recognition on the part of state education leaders that these scores may not be valid and reliable indicators of student (and school) achievement, given the fact that these students are not yet proficient in the language of the test. For a school designated as "failing" to

make AYP, these same ELL test scores can also be excluded from school AYP designations under No Child Left Behind (NCLB) upon appeal. Furthermore, under NCLB, schools are not required to have, nor be held accountable for, a subgroup of ELL or "Limited English Proficient" (LEP) students if there are less than 30 ELL students at any given grade-level on an AIMS subtest.[34]

Thus, Category 1 (ALL) data, and the resultant school labels assigned under Arizona LEARNS and NCLB, may not be reflective of the progress and achievement of ELL students. Nevertheless, test scores published in the local newspapers, on websites such as Greatschools.net, and on the state-required school report cards sent to parents, only include Category 1 (ALL) student data. Given the rhetoric of "test all students" and "no child left behind," parents and other members of the public are likely under the false impression that these reported test scores include all the students in a given school.

Category 2 (ELL) data are reported by the Arizona Department of Education (ADE) for both the AIMS and Stanford 9 tests as required by Proposition 203 and NCLB, although these data are not used for any particular purpose.[35] To date, no analyses of these data have been completed, or at least publicly reported. Analyses of these data are difficult given the fact that they are spread across multiple databases and reported in many different formats (e.g., Excel spreadsheets, HTML, Adobe Acrobat (PDF) files, and tab delimited text files from on-line report generators). In order to create this report, it was necessary to compile data from 21 different datasets available on the ADE website containing student achievement data (AIMS, Stanford 9) and school accountability (Arizona LEARNS, NCLB) ratings for the years 2002, 2003, and 2004 (see Appendix A).

In the comparisons between Category 1 (ALL) and Category 2 (ELL) data below, it is important to understand that these categories are not mutually exclusive. ELLs who have been enrolled in school for four years or longer are included in both categories. Given the general recognition that it takes four to seven years to acquire proficiency in English,[36] Category 1 excludes those ELLs with the lowest levels of English language proficiency while retaining those with the highest. Also, given the fact that ELLs currently constitute approximately 14.9 percent of the total student population in Arizona, the effects of the remaining ELL scores in Category 1 data are likely to be minimal. Thus, even though Category 1 and Category 2 are not mutually exclusive, they nonetheless provide the best approximation for the differences in test scores between ELL and Non-ELL (i.e., English-proficient) students.

Lastly, in analyzing the comparisons between Category 1 and Category 2 data below, it should be noted that the belief that test scores are a valid and reliable indicator of actual student achievement is not universal. There are numerous psychometric problems related to the inclusion of ELL students on English-only high-stakes tests which call into question the validity and reliability of their test scores.[37] This report simply focuses on how well the available data support the claims that education has improved for ELL students in Arizona as a result of the state's language, education, and testing policies.

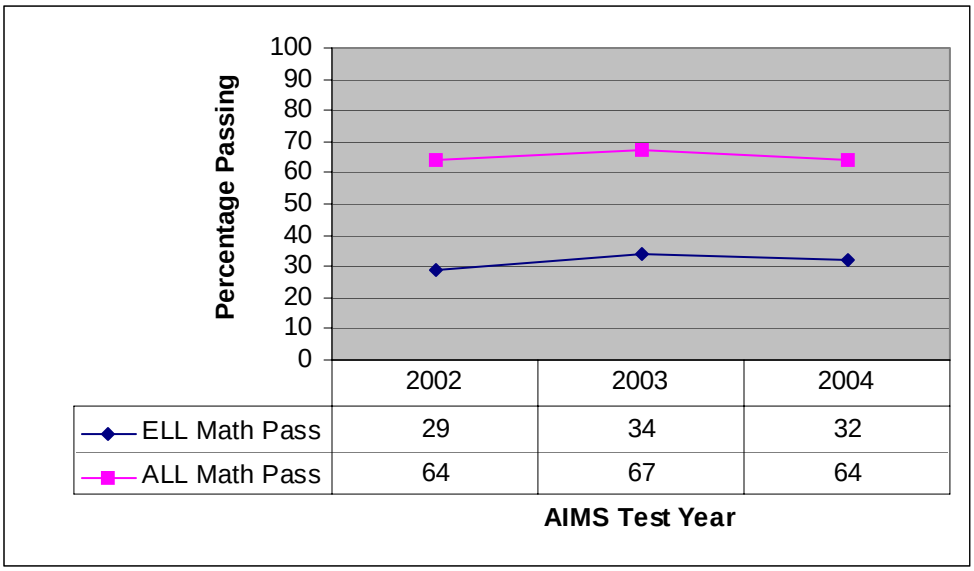
Comparison of Statewide ALL and ELL Student Achievement Data

Arizona Instrument to Measure Standards (AIMS)

Test results for AIMS are typically reported in the percentage of students whose scores fall into one of four categories: (a) *Exceeds* the Standards, (b) *Meets* the Standards, (c) *Approaches* the Standards, and (d) *Falls Far Below* the Standards. Students are considered as passing a subtest of the AIMS (i.e., Math, Reading, and Writing) if they *Meet* or *Exceed* the standards. In this report, we simply combine the *Meets* and *Exceeds* categories and report the percentage of students deemed as passing each subtest. It should be noted, however, that very few ELL students were deemed as *Exceeding* the standards; across the three years (2002-2004), on average, only 3 percent of English Language Learners (ELLs) *Exceeded* the standards in Writing, 4 percent in Reading, and 8 percent in Math.

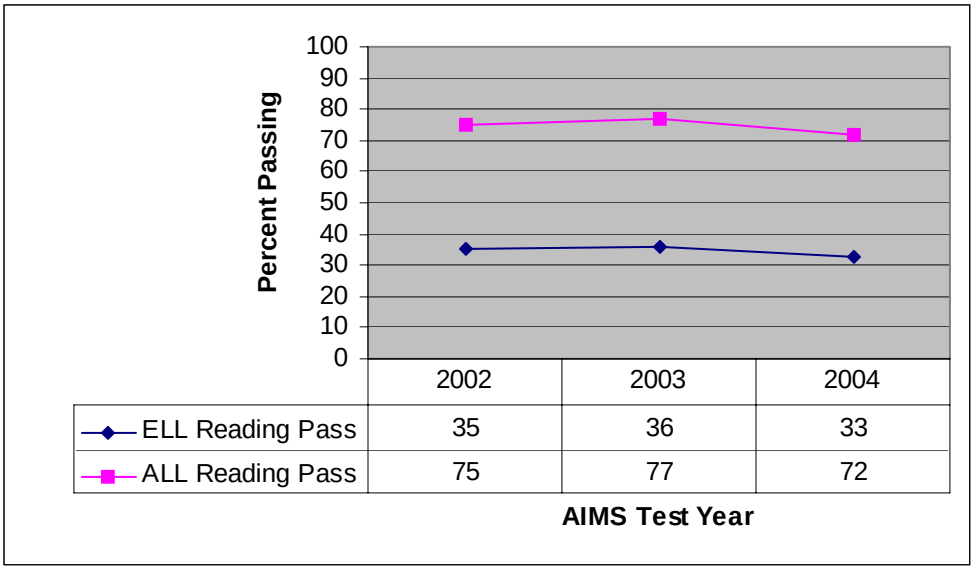
Figure 1 shows the percent of students in Category 1 (ALL) and Category 2 (ELL) who passed the third grade Math subtest of the AIMS. For students in both the ALL and ELL categories, scores rose slightly between 2002 and 2003, then decreased in 2004. However, a large gap is observed between ALL and ELL students across the three years, with 64 percent to 67 percent of ALL students passing, while only 29 percent to 32 percent of ELL attained passing scores. Thus, the majority of third grade students in the ALL category passed the Math subtest, while the majority of ELL students failed.

Figure 1: Statewide AIMS Third Grade Math 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percent Passing



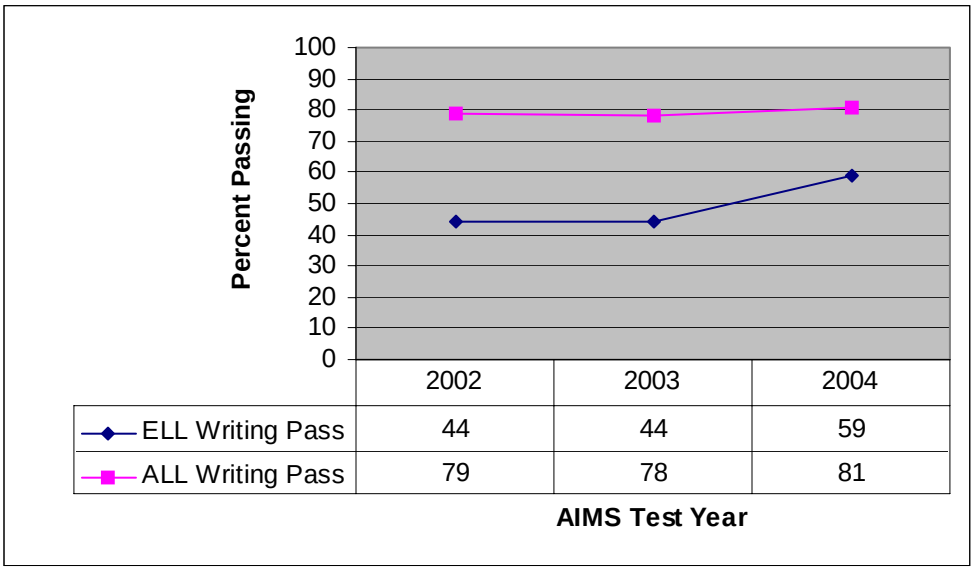
A similar trend is seen on the AIMS third grade Reading subtest (Figure 2): The majority of ALL students passed while the majority of ELLs failed. Pass rates for the ALL student group were between 72 percent and 77 percent, while pass rates for ELLs never exceeded 36 percent. As with the AIMS Math subtest, there is a slight increase for both ALL and ELL from 2002 to 2003, followed by a decrease in 2004. For both groups, a lower percentage of students passed the subtest in 2004 than two years earlier.

Figure 2: Statewide AIMS Third Grade Reading 2002-2004: Category 1 (ALL) and Category 2 (ELL) Percent Passing



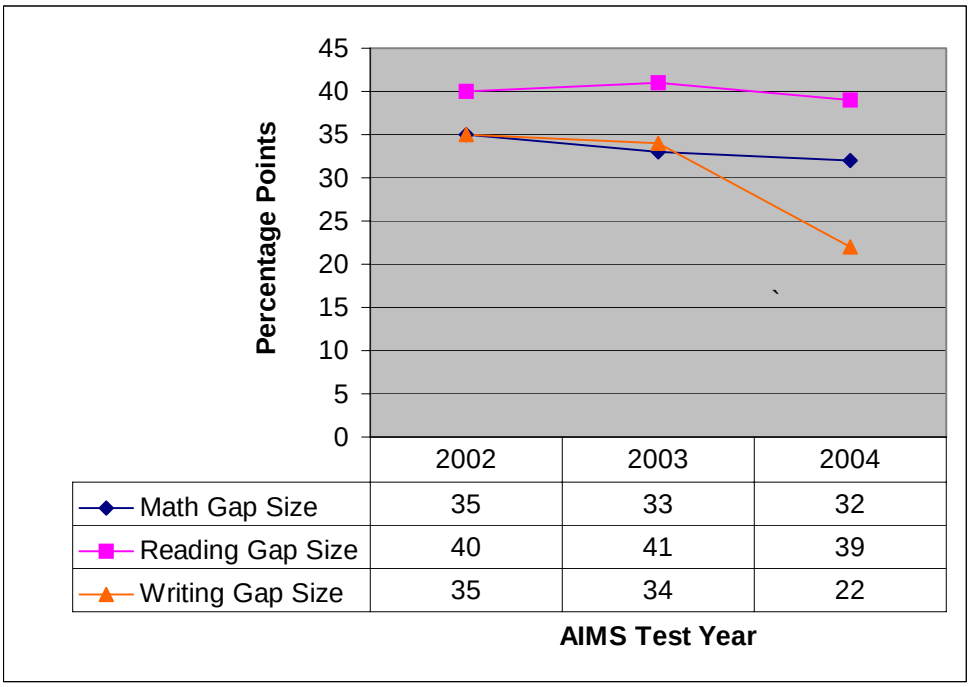
A surprisingly different trend is seen on the AIMS third grade Writing subtest (Figure 3). Between 79 percent and 81 percent of ALL students passed the writing test between 2002 and 2004. For ELL students, only 44 percent passed in both 2002 and 2003, but in 2004 the pass rate jumped dramatically to 59 percent—an increase of 15 percentage points. In contrast, there was an increase of only 3 percentage points for ALL students.

Figure 3: Statewide AIMS Third Grade Writing 2002-2004
Category 1 (ALL) and Category 2 (ELL) Percent Passing



In all three AIMS subtests, a wide gap is observed between students in the ALL and ELL categories. Figure 4 shows the gap size (differences in percentage passing) on the Reading, Writing, and Math subtests. In all three years, the widest gap between ALL and ELL occurred on the Reading subtest, with students in the ALL category scoring, on average, 40 percentage points higher than ELL students. The gap size increased slightly in 2003 and decreased slightly in 2004. The slight closing of the achievement gap between ELLs and ALL students on the Reading test in 2004 is actually due to the fact that the percent of ALL students passing decreased at a higher rate than the decrease in the percent of ELL students passing.

Figure 4: Statewide AIMS Third Grade Gap between ALL and ELL, 2002-2004



The gap between ALL and ELL was similar for both the third grade Math and Writing AIMS subtests in 2002 and 2003, with a gap size between 33 and 35 percentage points. The gap in Math lowered slightly in 2004, but as with Reading (above), the narrowing gap is a function of a lower percentage of ALL students passing Math in 2004. The most dramatic closing of the gap occurred on the Writing test between 2003 and 2004, with the gap size decreasing by 12 percentage points.

In summary, from the data from the third grade AIMS for 2002 to 2004, the majority of ELL students failed all three subtests each year, while the majority of ALL students passed each subtest each year. The only exception is in 2004 when 59 percent of ELLs passed the Writing subtest. There was a decrease in the percentage of both ALL and ELL students passing the Math and Reading subtests between 2003 and 2004. A lower percentage of ELLs passed the Reading subtest in 2004 than two years earlier in

2002. A large gap is evident between students in Category 1 (ALL) and Category 2 (ELL). While very slight gap size decreases were observed on the Math and Reading subtests, these were the result of lower percentages of students in the ALL category passing. Only the narrowing of the gap in Writing can be attributed to a dramatic increase in ELLs passing this subtest in 2004.

Stanford Achievement Test, 9th Edition (Stanford 9)

Third grade students in Arizona took three sections of the Stanford 9—Language, Math, and Reading. Test results for Stanford 9 are typically reported as aggregate (or averaged) percentile ranks. The Stanford 9 is a norm-referenced test, and no passing standard has been set by the state. Nonetheless, policymakers and educators typically expect students to score at least at or above the 50th percentile, which (purportedly) indicates that students are at or above the national average.[38] In analyzing the differences in percentile rankings below, it is important to point out that distances between percentiles are not equal.[39] It is also important to note that while the percentile rankings are reported to parents and published in local newspapers, these percentile rankings are not used in Arizona LEARNS school accountability formulas. Rather, ADE uses stanine scores (a statistic which assigns students a score between 1 and 9) for students across multiple test years in an effort to measure growth over time.[40] This is the basis for MAP (see above) which does get factored into the Arizona LEARNS school accountability formulas. Despite the limitations of the aggregate percentile rankings, they do provide some basis for comparisons between students in the ALL and ELL categories, particularly given the fact that, unlike the AIMS test, no changes have been made to the Stanford 9 test during these years.[41]

Figure 5 shows the results of the third grade Stanford 9 for the Language subtest. Third grade students in the ALL category consistently scored above the 50th percentile from 2002 to 2004, while ELLs never scored higher than the 34th percentile. In both groups, rankings improved from 2002 to 2003, and decreased from 2003 to 2004. The decrease for the ELL group, however, was greater.

Figure 5: Statewide Stanford 9 Third Grade Language, 2002-2004: Category 1 (ALL) and Category 2 (ELL) Percentile Rankings

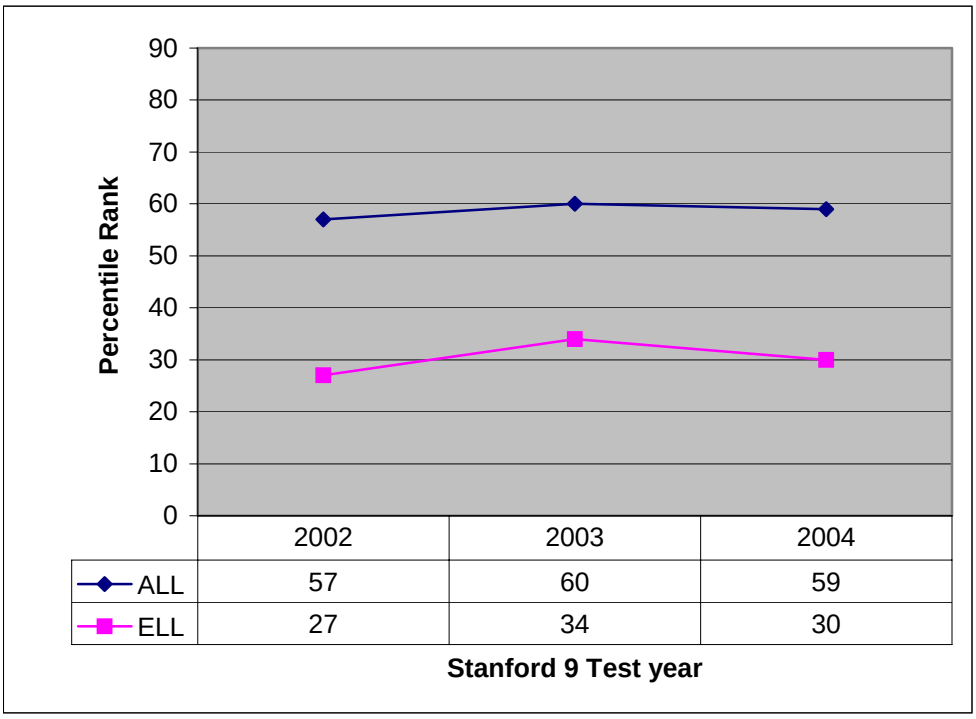


Figure 6 shows the results of the third grade Stanford 9 Math subtest. As with the Language subtest, students in the ALL category were above national average from 2002 to 2004 while ELL students did not score higher than the 35th percentile. While rankings increased for both groups from 2002 to 2003, the ranking for the ELL category declined from 2003 to 2004 while the ALL category remained stable.

Figure 6: Statewide Stanford 9 Third Grade Math, 2002-2004
Category 1 (ALL) and Category 2 (ELL) Percentile Rankings

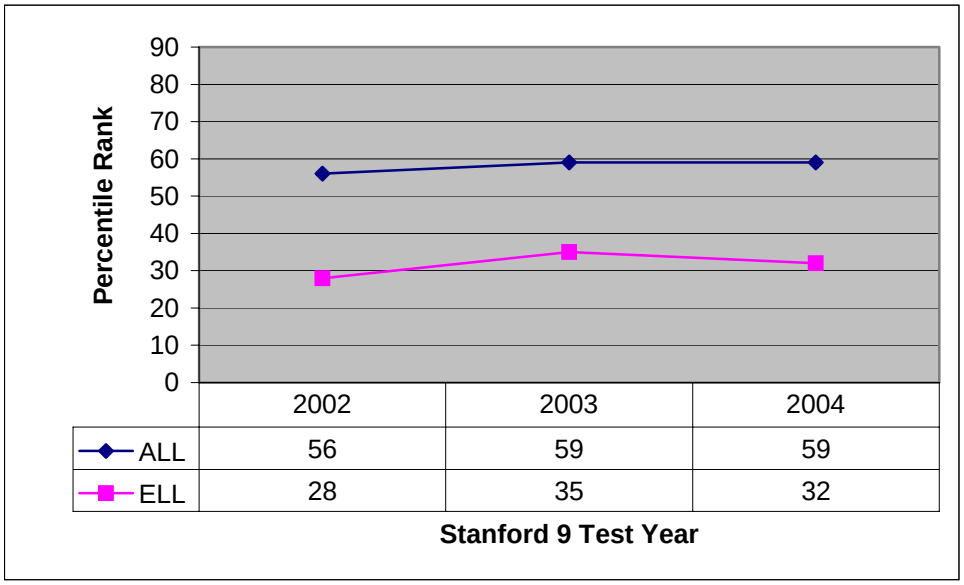
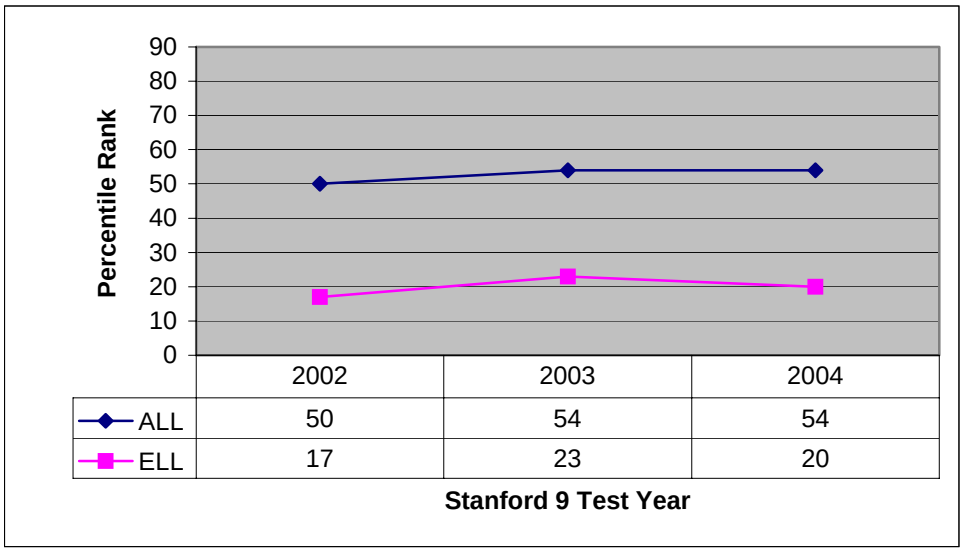


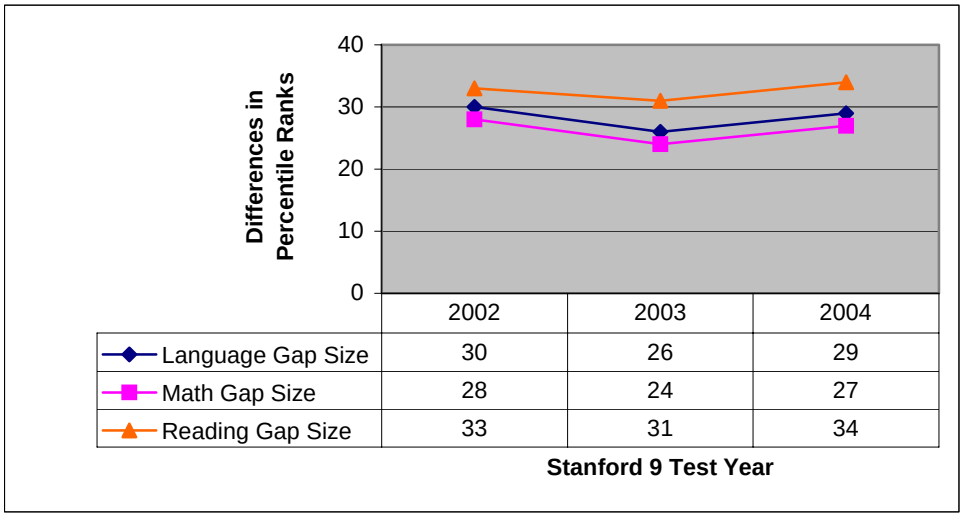
Figure 7 shows the results of the third grade Stanford 9 Reading subtest. Once again, students in the ALL category score at or above national average, while ELLs are well below average, scoring no higher than the 23rd percentile. While there was an increase for both groups from 2002 to 2003, the ELL subgroup decreased in 2004 while the ALL subgroup remained stable.

Figure 7: Statewide Stanford 9 Third Grade Reading, 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percentile Rankings



As on the AIMS test, a wide gap exists between ALL and ELL students on the Stanford 9. Students in the ELL category trailed far behind students in the ALL category by an average of 28 percentile points in Language, 26 percentile points in Math, and 33 percentile points in Reading. As shown in Figure 8, the gap narrowed slightly between 2002 and 2003, but then increased in 2004 to nearly the same levels as in 2002. In the case of Reading, the gap in 2004 was slightly higher than it was in 2002.

Figure 8: Statewide Stanford 9 Third Grade 2002-2004: Gap Size Between ALL and ELL



Simulated Stanford 9 Cohorts

Unlike AIMS, prior to 2005 the Stanford 9 was administered every year to elementary students in grades two and higher. Table 1 shows the percentile rankings for grades two through five for both Category 1 (ALL) and Category 2 (ELL). Without exception, in every grade level in every year for every Stanford 9 subtest, the average percentile rankings for the ELL group increased from 2002 to 2003, and then decreased in 2004. Average percentile rankings for the ALL student subgroup, in contrast, are more stable, with slight increases between 2002 and 2003, and no change or only slight decreases in 2004. In the case of fifth grade Math, there was an increase for the ALL category, in contrast to a decrease for the ELL category.

Table 1: Statewide Stanford 9 Results, Grades 2-5, 2002-2004

Stanford 9 Category 2 (ELL) - All District and Charter Schools									
	Subject	2nd Grade		3rd Grade		4th Grade		5th Grade	
		#	PR	#	PR	#	PR	#	PR
2002	Language	15,230	17	14,297	27	13,468	23	11,514	19
	Math	15,499	34	14,435	28	13,568	29	11,644	29
	Reading	14,373	24	13,940	17	12,741	19	11,109	17
2003	Language	17,485	22	16,946	34	15,642	28	14,786	24
	Math	17,655	40	17,083	35	15,836	37	15,005	36
	Reading	16,477	30	16,613	23	14,998	26	14,395	23
2004	Language	16,235	20	13,203	30	7,983	22	7,193	18
	Math	16,417	37	13,315	32	8,062	30	7,253	32
	Reading	15,469	27	12,992	20	7,607	18	6,949	17

Stanford 9 Category 1 (ALL) - All District and Charter Schools									
	Subject	2nd Grade		3rd Grade		4th Grade		5th Grade	
		#	PR	#	PR	#	PR	#	PR
2002	Language	54,081	48	59,339	57	60,603	50	61,770	47
	Math	54,237	61	59,473	56	60,898	58	62,187	59
	Reading	52,059	57	58,616	50	59,465	55	61,156	53
2003	Language	52,282	49	54,135	60	58,154	52	59,598	49
	Math	52,471	63	54,320	59	58,574	60	60,110	61
	Reading	50,100	57	53,597	54	57,076	57	59,038	54
2004	Language	55,954	48	57,077	59	62,241	50	62,499	48
	Math	56,281	63	57,357	59	62,609	59	62,937	63

	Reading	54,162	57	56,635	54	61,173	56	62,003	54

Note: # = number student tested; PR = Percentile Rank

Given the fact that the Stanford 9 is taken each year, it is possible to simulate cohorts of students from 2002 to 2004 (that is, tracking the test scores of a group of students as they go from one grade to the next). It should be noted, however, that this method assumes that the same students have moved together from grade to grade over the three-year period. While it can be argued that it is likely (especially with statewide data) that the majority of the children are the same, factors such as retention and student mobility (e.g., moving out of and into the state, moving back and forth between charter/private schools and public schools, home schooling, etc.), and the rapidly growing student population as a result of new families moving into the state, affect the stability of the subgroups. This is particularly problematic for the Category 2 (ELL) subgroup, as students who attain fluency in English are redesignated and removed from the group.[42] Also, the Category 1 (ALL) subgroup changes each year as previously-excluded ELL students are included after four years of attendance. Nevertheless, the trends in these cohorts may provide some evidence of the general test performance of students in these subgroups as they moved up in grade level each year.

Table 2 shows the aggregate percentile rankings for students who were in grades two to four, and in grades three to five from 2002 to 2004, respectively. In general for the two ELL cohorts, the percentile ranks increased slightly between 2002 and 2003, then decreased in 2004. The trends are less consistent for the two cohorts of students in the ALL student category, but in contrast to the ELL students, some cohorts of ALL students saw improvement between 2002, 2003 and 2004.

Table 2: Statewide Stanford 9 Cohorts, Grades 2-4 and 3-5, 2002-2004

	Language			Math			Reading		
ELL	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**
2nd-4th	17	34	22	34	35	30	24	23	18
3rd-5th	27	28	18	28	37	32	17	26	17
ALL	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**
2nd-4th	48	60	50	61	59	59	57	54	56
3rd-5th	57	52	48	56	60	63	50	57	54

In summary, while students in the ALL student category consistently perform (on average) above national average (except for second and fifth grade Language subtests where scores were slightly below the 50th percentile), students in the ELL subgroup are (on average) far below the national norm. The highest ranking is the 40th percentile on second grade Math in 2003, and the lowest is the 17th percentile in 2004 on fifth grade Reading in both 2002 and 2004. There is a wide gap between ALL and ELL, and the gap widened between 2003 and 2004. A pattern of general improvement is observed between 2002 and 2003. However, ELL scores declined in every grade level (second through fifth) and on every Stanford 9 subtest between 2003 and 2004.

Discrepancies in Number of Third Grade Students Tested on AIMS and Stanford 9

As indicated above, from 2002 to 2004, under both state and federal policy, all elementary students in grades three and five were required to take the AIMS test as mandated by both Arizona LEARNS and NCLB, and all elementary students in grades

two and higher were required to take the Stanford 9 as mandated by Proposition 203. Given these mandates, it would be expected that the number of students taking the AIMS and Stanford 9 would be (roughly) equal. However, as shown in Figures 9 and 10, there are discrepancies in the number of third grade students taking the AIMS and Stanford 9 for both Category 1 (ALL) and Category 2 (ELL). For the ALL student category (Figure 9), 1,293 more third grade students took the Stanford 9 test than the AIMS in 2002, while in 2003 and 2004, more third graders took the AIMS than the Stanford 9; in 2003, 2,996 more third grade students took the AIMS than the Stanford 9, while in 2004 the difference was only 785. Both the AIMS and Stanford 9 are given in the spring semester, typically within one or two weeks of each other. Nevertheless, it may be possible to attribute these discrepancies to student absences, or students moving into or out of state between administrations of the AIMS and Stanford 9.

The discrepancy between the numbers of students tested across the three years, however, is more difficult to explain. Arizona has one of the fastest growing student populations in the country. Therefore, it is difficult to understand why the number of third grade students tested on both AIMS and Stanford 9 in 2003 and 2004 are less than the number tested in 2002. On the Stanford 9, 5,153 fewer third grade students took the Math subtest in 2003 compared to 2002, and while the number of third grade students increased by 3,037 the following year, this is still 2,116 fewer students tested than in 2002. On the AIMS test, 1,856 fewer students took the Math subtest in 2003 than in 2002, and while the number increased by 1,818 the following year in 2004, it is still 38 students fewer than the number tested in 2002.

Figure 9: Number of Third Grade Category 1 (ALL) Taking the AIMS and Stanford 9 Math Tests, 2002-2004

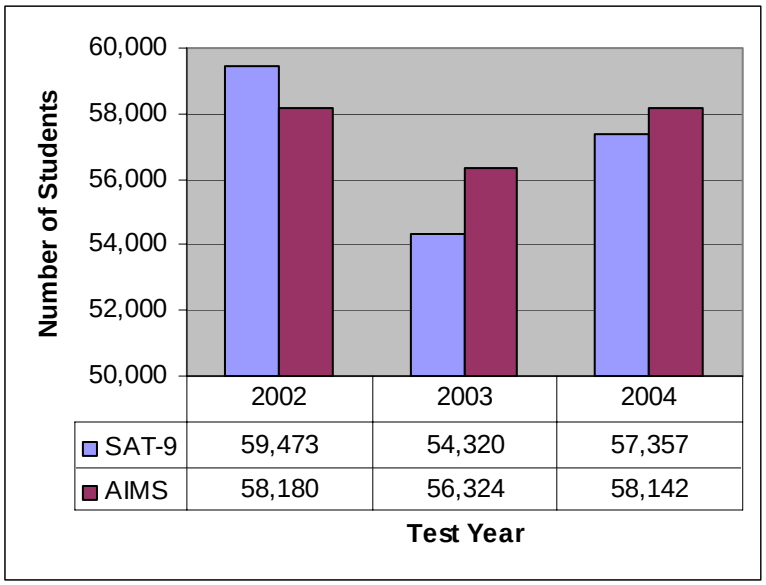
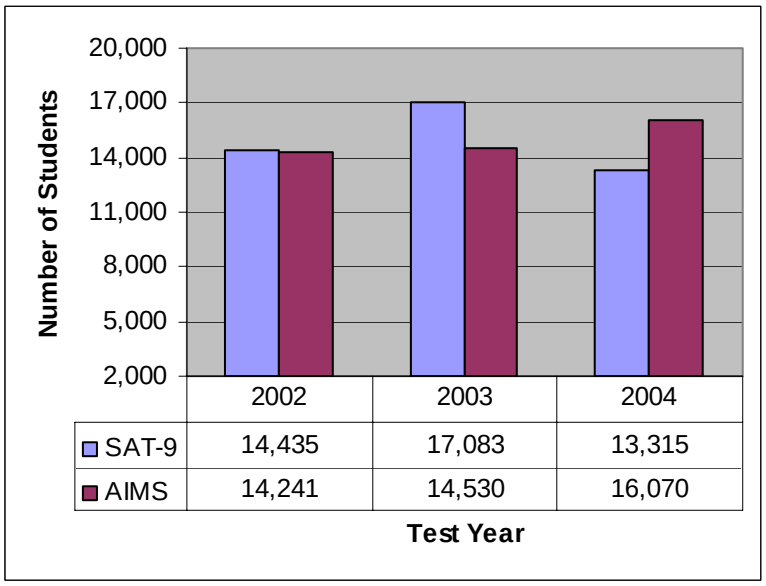


Figure 10: Number of Third Grade Category 2 (ELL) Taking the AIMS and Stanford 9 Math Tests, 2002-2004



A much different pattern is observed for the number of Category 2 (ELL) students tested on the third grade AIMS and Stanford 9 Math subtests. The number of students taking the AIMS tests steadily increased each year, with 289 more taking the test in 2003,

and 1,540 more taking it in 2004. While a sudden decrease is not observed in 2003 as with the ALL students, the increase is still quite small and appears inconsistent with the rapidly growing ELL student population in the state. Some evidence for this can be found in the contrast of the number of ELL test takers on the Stanford 9 which had a much larger increase in the number of students taking the Math subtest with 2,648 more students taking the Stanford 9 Math test in 2003 than in 2002. However, this was followed by a sharp decrease of 3,768 in 2004. While the discrepancies between the number of third grade ELLs taking the AIMS and Stanford 9 Math subtest in 2002 may be small enough to attribute to absences and student mobility, the discrepancies in 2003 and 2004 are too wide for this explanation to be feasible.

In summary, the numbers of third grade students tested in 2003 and 2004 is inconsistent with the rapidly growing student population. Therefore it appears that many test scores of third grade students are missing from the data reported by ADE. To date ADE has not publicly reported the number of students classified as ELLs for each school and grade level, therefore it is difficult to know just how many third grade ELLs should have been tested. As with the ALL student category, the decline in the number of ELLs taking the third grade Stanford 9 Math subtest between 2003 and 2004 is not consistent with the rapidly growing ELL student population in Arizona. The increases in the number of third grade ELLs taking the AIMS test provides evidence of the growing ELL student population, but the increase in the number taking the AIMS test is much less than the increase on the Stanford 9 between 2002 and 2003. Thus, it appears that many ELL student test scores are also missing from the publicly reported data. For both the ALL and ELL student categories, it is unclear at this point whether there were actually fewer

students taking the tests, if there were problems in reporting the data, or if certain scores were systematically excluded from public reporting. Regardless, these discrepancies and inconsistencies should be kept in mind in the next section on changes in Arizona LEARNS schools labels.

Changes in Arizona LEARNS Labels

Table 3 shows the number of schools attaining each label under Arizona LEARNS from 2002 to 2004. Notice the change in the names of the labels between 2002 and 2003. The labels "Improving" and "Maintaining" proved to be confusing to district and school administrators, teachers, parents, students, and the general public. More importantly, these labels did not carry politically symbolic weight in portraying a picture of success in improving education in Arizona.[43] These labels were replaced by more positive sounding labels in 2003—"Performing" and "Highly Performing." It should also be noted that not every school received a label each year. In fact, as shown in Table 3, there were fewer schools labeled in 2003 than in 2002. In some years as many as 40 percent of schools did not receive a label because the schools were designated as K-2, new, small, or alternative.

In 2002 only three schools in the state received the highest classification of "Excelling," while 276 schools (21.7%) were designated as "Underperforming." State leaders were uncomfortable with these results for a number of reasons, but particularly because of concern about the high costs involved in providing the assistance to these schools as required by the law.[44] The following year, the state made several changes to the formulas and procedures used to assign labels, making it easier for schools to obtain the "Excelling" label, and more difficult to obtain the "Underperforming" label.[45] As a

result, in 2003, 132 schools (12%) received the "Excelling" designation, while the number of Underperfoming schools was reduced to 135 (12.4%)—a decrease of over 50 percent.[46]

Table 3 – Statewide Arizona LEARNS Labels, 2002-2004

2002		2003		2004
Label	**# of Schools**	**Label**[47]	**# of Schools**	**# of Schools**
Excelling	3 (0.2%)	**Excelling**	132 (12.0%)	151 (9.1%)
Maintaining	547 (43.0%)	**Highly Performing**	167 (15.2%)	208 (12.6%)
Improving	446 (35.1%)	**Performing**	663 (60.4%)	1,173 (70.9%)
Underperforming	276 (21.7%)	**Underperforming**	135 (12.3%)	110 (6.6%)
		Failing	---	12 (0.7%)
Totals	*1,272*		*1,097*	*1,654*

Note: Percentages may not add up to 100 because of rounding.

A key component of the new formula was a change in how ELL test scores affect a school's designation. For schools' aggregate AIMS test scores and MAP calculations, scores for ELLs enrolled less than four years are excluded.[48] This policy change eliminated the scores of many ELL students, particularly and most importantly, those at the lowest levels of English language proficiency. Also between 2002 and 2003, several changes were made to the AIMS test in an effort to make it more "reasonable," and also to correct many of the problems with previous versions of the tests.[49]

The "Failing" label was first used in 2004 and assigned to 12 schools that had been designated as "Underperforming" in both 2002 and 2003. The current Superintendent of Public Instruction, in his 2005 State of Education speech, claimed that "the accountability system had no significant changes in its second year, so the only way the schools could escape failing status was to raise the student test scores."[50] He

described ADE's success in assisting 70 out of 81 schools improve their test scores enough to become "Performing" schools and thus avoided becoming "Failing schools."[51] Nonetheless, as shown in Table 3, there are still 110 Underperforming schools in the state. If 70 schools improved to Performing and 12 became Failing, then the current number of Underperforming schools include 53 schools labeled as "Underperforming" in both 2003 and 2004, and another 57 schools labeled as "Underperforming" for the first time in 2004. Thus, in these 110 schools, there has been no improvement or a decline in achievement (at least as measured by Arizona LEARNS) between 2002 and 2004.[52]

One other observation about the changes in the Arizona LEARNS labels should be made: In 2002, 43.2 percent of schools achieved the top two label categories (Excelling and Maintaining), however, in 2003 this declined to 27.2 percent of schools, followed by further decline in 2004 to only 21.7 percent of schools achieving the top two labels. The vast majority of schools, nearly 71 percent, are in the third category, "Performing." This also could represent a decline in student achievement, again, at least as measured in terms of test scores and Arizona LEARNS.

Comparison of ELL Impacted Elementary School ALL and ELL Student Achievement

Schools with few English Language Learner (ELL) students typically were not affected much by Proposition 203, as few, if any, of these schools offered bilingual education programs. In fact**,** many schools reporting Category 2 data had very few ELL students. Thus, most of the ELL students in these schools were already in the type of English-only classrooms mandated by Proposition 203. Furthermore, test scores of ELL

students in these schools likely had minimal, if any, impact on the schools' aggregate (Category 1) test scores used for school accountability purposes under both Arizona LEARNS and No Child Left Behind (NCLB).

In an effort to assess whether these policies are indeed leading to improved academic achievement of ELL students, this section will focus on "ELL Impacted" elementary schools, that is, those schools with large ELL student populations. We identified a of total 190 schools that tested 30 or more third grade ELLs on the 2004 AIMS Math subtest (see Appendix B). These 190 schools provide instruction to 71 percent (N=11,091) of third grade ELLs in Arizona. As describe above, the minimum group size for the LEP (Limited English Proficient is the federal government's label for ELLs) subgroup under NCLB is 30. Thus, schools with 30 or more ELLs in third grade (and higher) are more likely to be affected by their ELL student test scores. Furthermore, these schools were more likely to have had the types of bilingual and/or English as a Second Language programs that Proposition 203 restricts, and thus have had to make substantial changes to their programs for ELL students to provide the English-only education mandated by the law.

The analyses below mirror those above for the statewide data. The first analysis examines the differences between Category 1 (ALL) and Category 2 (ELL) test scores on the AIMS test, followed by differences in performance on the Stanford 9, between 2002 and 2004. We then analyze the changes in Arizona LEARNS labels and Adequate Yearly Progress (AYP) designations.

Arizona Instrument to Measure Standards (AIMS)

Figure 11 shows the percentage of students in ELL Impacted elementary schools in Category 1 (ALL) and Category 2 (ELL) who passed the AIMS third grade Math subtest. As observed in the statewide data, scores for both groups rose slightly between 2002 and 2003, and then dropped slightly in 2004. Between 2003 and 2004, slightly more than half of the students in the ALL category passed the Math subtest, while the number of ELLs passing never exceeded 35 percent. Thus, in ELL Impacted elementary schools, the majority of third grade ELLs failed the AIMS Math subtest, with a higher percentage failing in 2004 than the previous year.

Figure 11: ELL Impacted Elementary Schools AIMS Third Grade Math 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percent Passing

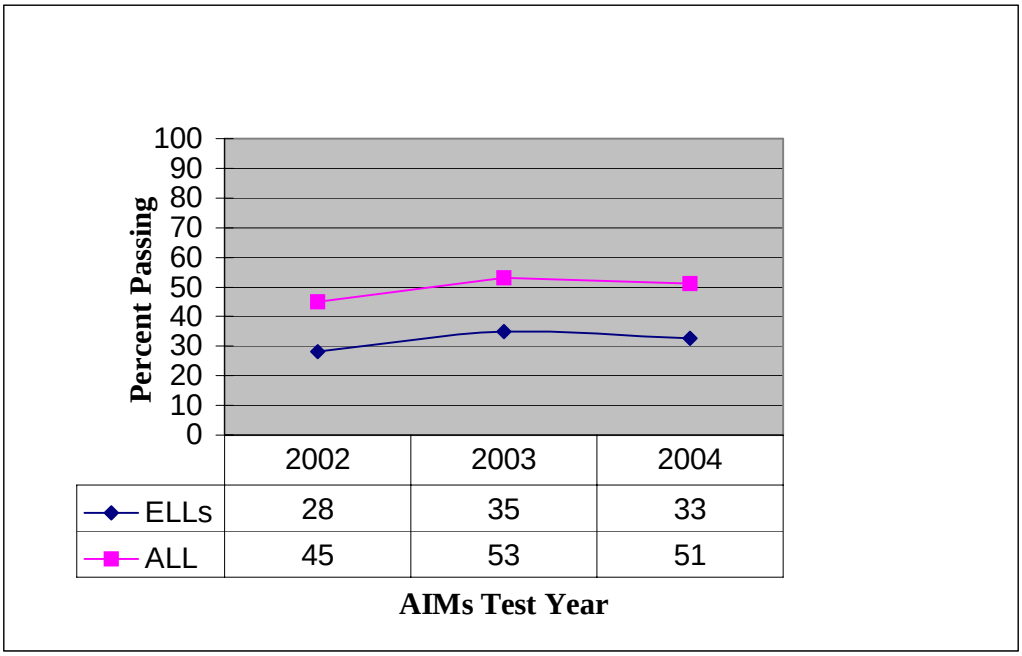
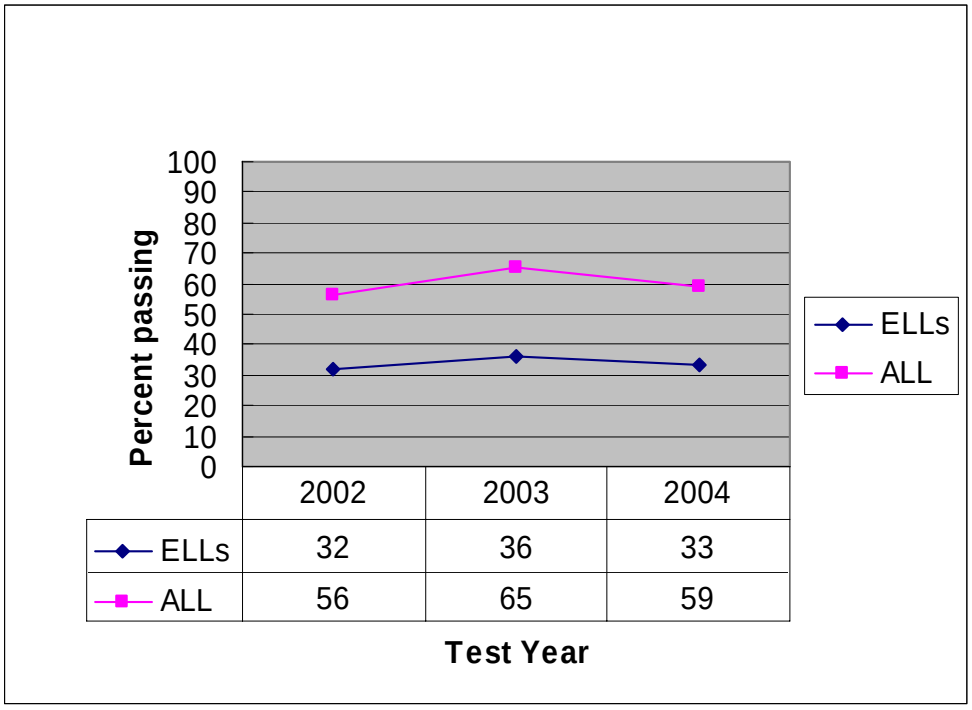


Figure 12 shows the percentage of students in both categories who passed the third grade AIMS Reading subtest. As with the Math subtest above, test scores for both

groups improved slightly between 2002 and 2003, and then decreased slightly in 2004. The majority of students in the ALL category passed, while the majority of ELL students failed.

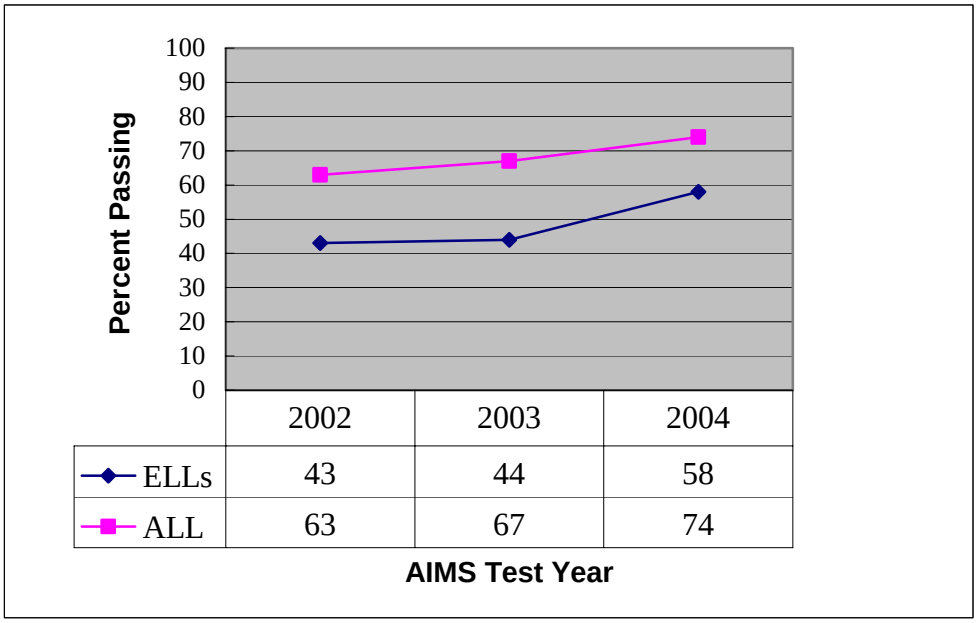
Figure 12: ELL Impacted Schools AIMS Third Grade Reading 2002 2004 Category 1 (ALL) and Category 2 (ELL) Percent Passing



	2002	2003	2004
ELLs	32	36	33
ALL	56	65	59

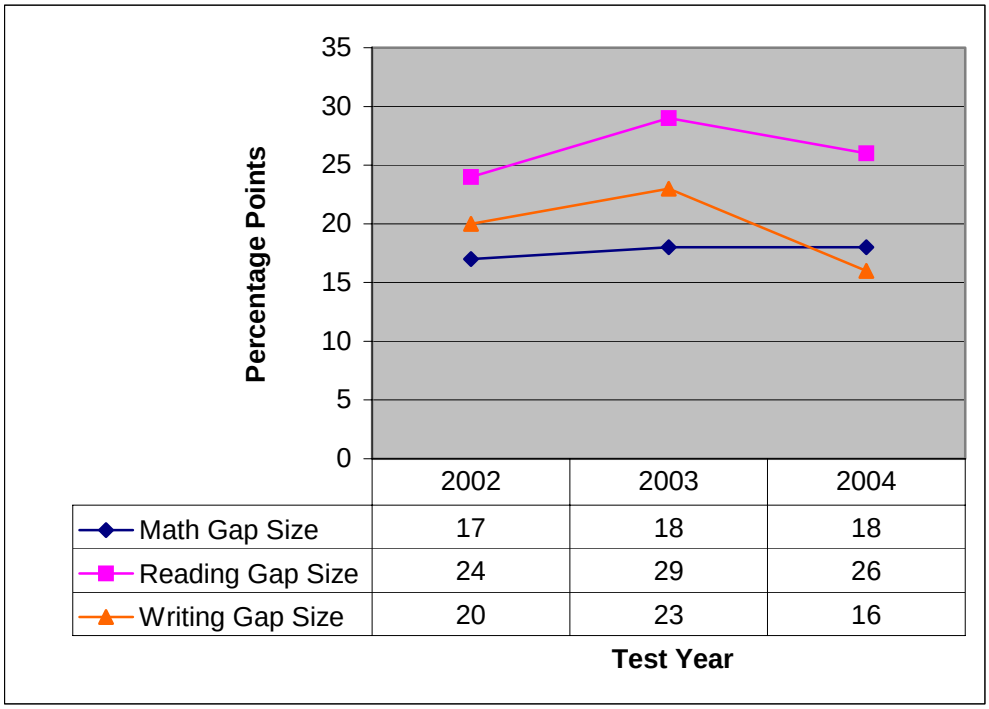
As with the statewide data, a much different pattern is observed on the AIMS Writing subtest (Figure 13). Pass rates for both the ALL and the ELL groups increased slightly from 2002 to 2003, followed by a more dramatic increase in 2004. Additionally, students in the ELL category showed even greater improvement on the Writing subtest than students in the ALL category between 2003 and 2004. By 2004, a little over half of the ELL students (58%) had passed the AIMS Writing subtest.

Figure 13-ELL Impacted Elementary Schools' AIMS Third Grade Writing 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percent Passing



As with the statewide data, large gaps are observed in the percentage of students in the ALL and ELL categories passing the various AIMS subtests (Figure 14). The largest gaps are observed on the Reading subtest with ELLs trailing behind students in the ALL category by an average of 26 percentage points, followed by Writing with a gap of 20, and Math with a gap of 18. The gap size increased for all three subtests between 2002 and 2003. In 2004, the gap size appears to have decreased slightly for both Reading and Writing, while remaining the same for Math. However, as with the statewide data, the closing of the gap on Reading scores is a function of a greater decline in the number of students passing in the ALL category in 2004. Only the decrease in the Writing score gap can be attributed to higher ELL student test scores.

Figure 14: ELL Impacted Elementary Schools AIMS Third Grade Gap between ALL and ELL, 2002-2004



In summary, in the ELL Impacted elementary schools between 2002 and 2004, the majority of third grade students in the ALL category passed the AIMS test, while the majority of ELLs failed. On the Reading and Math subtests, scores declined for both groups between 2003 and 2004. The majority of third grade ELLs failed the Writing subtest in 2002 and 2003. This changed in 2004 following a sudden increase of 14 percentage points in the number of ELLs passing the Writing subtest. Large gaps between the performance of students in the two categories are observed with ELL students trailing far behind students in the ALL category.

In comparison with the statewide data, the pass rates of Category 2 (ELL) third grade students in ELL Impacted elementary schools on each AIMS subtest is nearly identical. This is likely due to the fact that students in the ELL Impacted elementary schools account for 71 percent of the statewide Category 2 data. In contrast, Category 1

(ALL) students in ELL Impacted elementary schools trailed behind ALL students in the statewide data by an average of 15 percentage points in Math and Reading, and 11 percentage points in Writing. This gap is indicative of the fact that ELL Impacted schools are typically in lower socioeconomic neighborhoods.[53] Also, ELL Impacted elementary schools are more likely to have a greater number of ELL and former ELL students in the ALL category.

Stanford 9

Figure 15 shows the results of the ELL Impacted schools' third grade Stanford 9 Language subtest. Students in both categories scored below the 50th percentile in all three years, but ELL scores were lower than students in the ALL category. Students in the ALL category scored (on average) above the 40th percentile from 2002 to 2004, while students in the ELL category never scored higher than the 32nd percentile. Both groups increased their percentile ranking between 2002 and 2003, but ELL scores declined slightly in 2004 while ALL scores remained stable.

Figure 15: ELL Impacted Elementary Schools Stanford 9 Third Grade Language, 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percentile Ranking

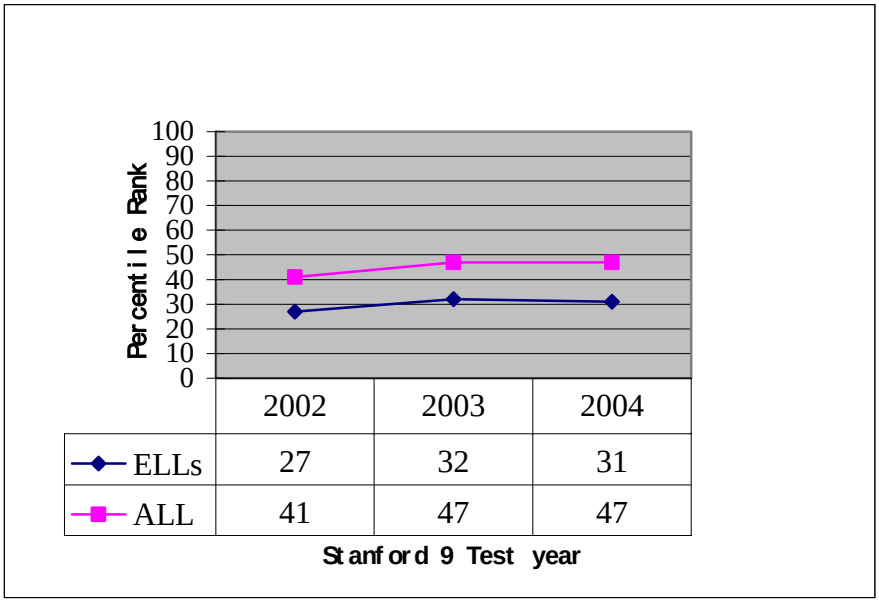


Figure 16 shows the results of the ELL Impacted schools' third grade Stanford 9 Math subtest. As with the Language subtest, both groups scored below the 50th percentile, and ELL scores were lower than ALL scores. The third grade students in the ALL category still ranked at or above the 40th percentile, while third grade students in the ELL category never exceeded the 33rd percentile. Rankings increased for both groups from 2002 to 2003. In 2004, the ranking for ALL students increased slightly while the ranking for ELLs remained the same.

Figure 16: ELL Impacted Elementary Schools Stanford 9 Third Grade Math, 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percentile Rankings

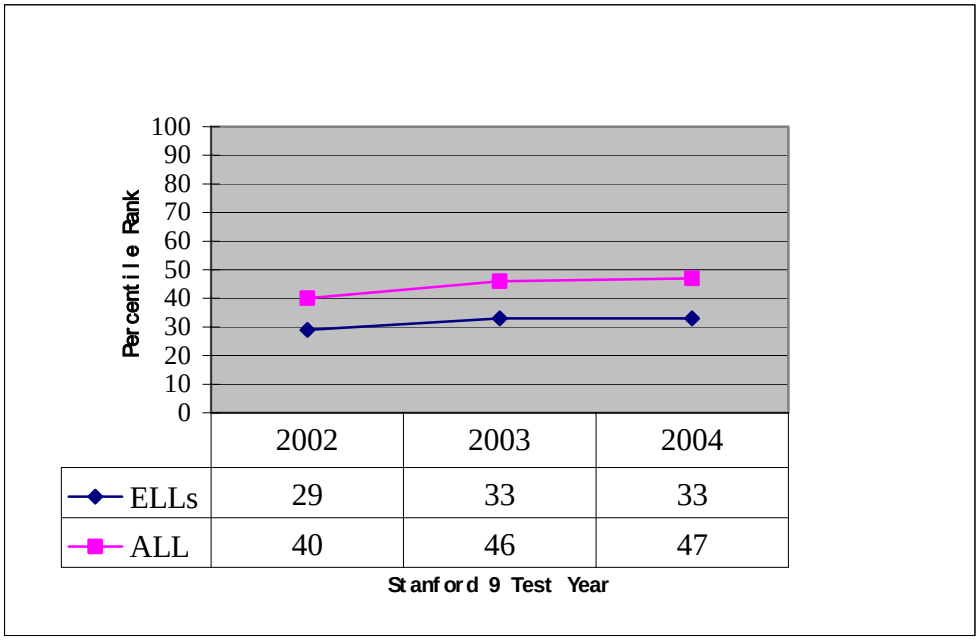
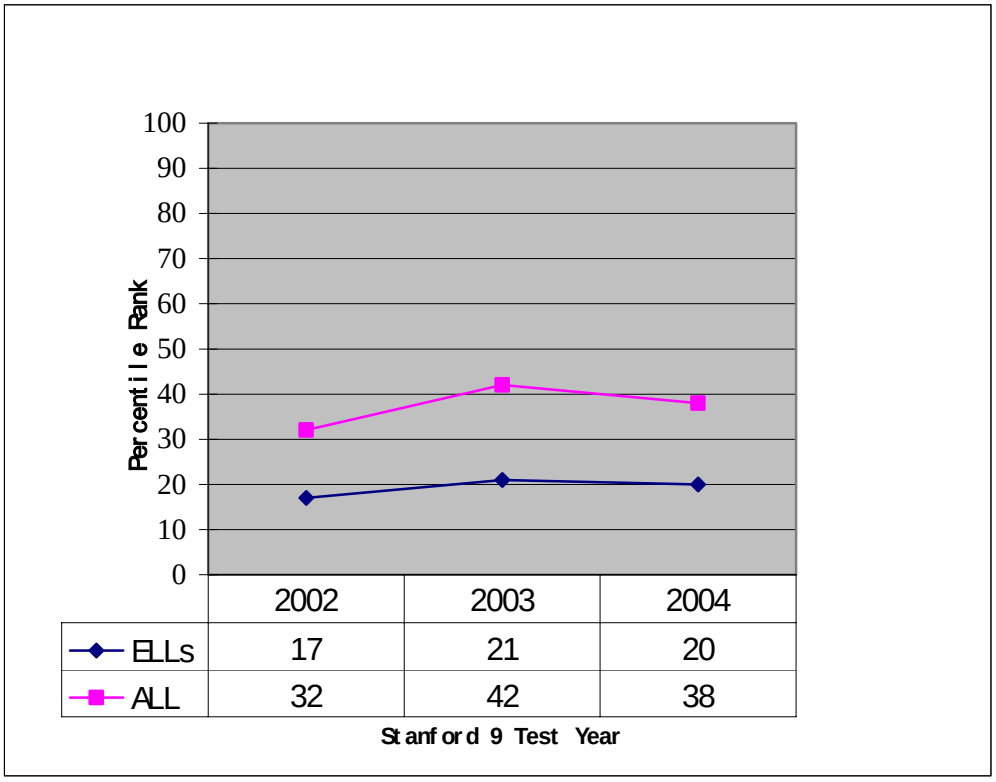


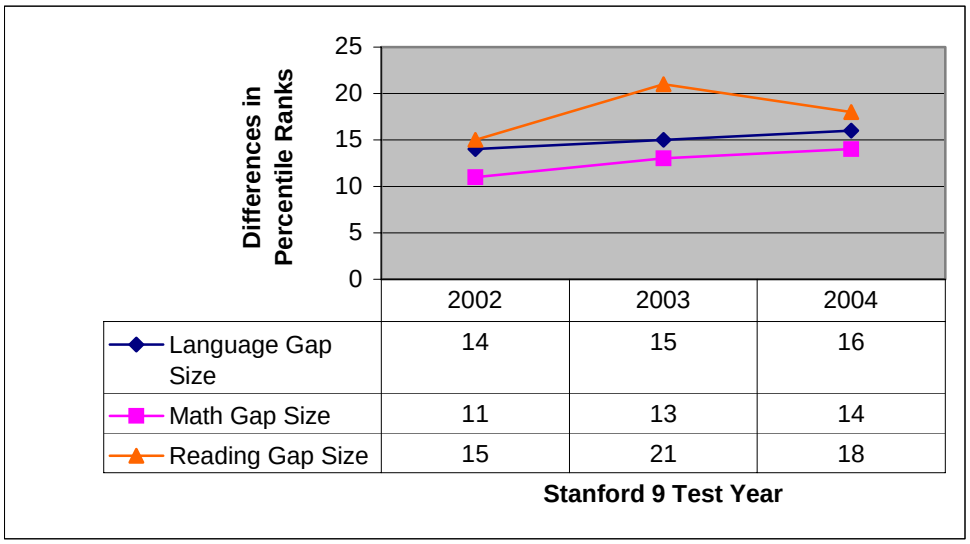
Figure 17 shows the results of the ELL Impacted schools' third grade Stanford 9 Reading subtest. Once again, both groups scored below the 50th percentile, with ELL students scoring lower than ALL students. Students in the ALL category scored as high as the 42nd percentile, while ELL students never exceeded the 21st percentile. For both groups there was a slight increase from 2002 to 2003, followed by a decline in 2004.

Figure 17: ELL Impacted Elementary Schools Stanford 9 Third Grade Reading, 2002-2004 Category 1 (ALL) and Category 2 (ELL) Percentile Rankings



The figures above also reveal a large gap between the performance of ELL and ALL students in ELL Impacted elementary schools, with ELL students trailing behind students in the ALL category. Figure 18 shows the size of these gaps for each of the Stanford 9 subtests between 2002 and 2004. Across all three years, ELLs trailed furthest behind on the Reading subtest, followed by the Language subtest and then the Math subtest. The gap size increased slightly each year for the Language and Math subtests. While the gap size appears to have closed slightly for the Reading subtest, this change is due to the fact that scores for students in the ALL category declined at a higher rate than ELL students between 2003 and 2004.

Figure 18: ELL Impacted Elementary Schools Stanford 9 Third Grade 2002-2004 Gap Size Between ALL and ELL



As with the statewide data, simulated cohorts of students were created for grades two through four and grades three through five from 2002 to 2004 for students in ELL Impacted elementary schools (see earlier discussion on limitations of simulated cohorts, pg. 18). In general, average percentile ranks for the ELL subgroup declined between 2003 and 2004 as students moved from grade three to four and from grades four to five (except Reading for the third to fifth cohort where the ranking remained the same). In the case of Math and Reading, scores for ELL students in the second to fourth grade cohort consistently declined as they moved up in grade level. A similar decline is observed for ELL students in Math in the third to fifth grade cohort. Declines are also observed for students in the ALL category in ELL Impacted elementary schools between 2003 and 2004 with the exception of third to fifth grade Math subtest scores. As with the AIMS scores, the Stanford 9 scores for ELL students in ELL Impacted schools are similar to the statewide data reported above. The average percentile ranks for ALL students in these schools, however, trail behind ALL students in the statewide data by 9 to 25 percentile

points. Once again, this may be indicative of the typically low socioeconomic status of ELL Impacted Elementary Schools, and the fact that both current and former ELL students make up a greater percentage of students in the ALL category than at schools with few or no ELLs.

Table 4: ELL Impacted Elementary Schools Stanford 9 Cohorts, Grades 2-4 and 3-5, 2002-2004

	Language			Math			Reading		
ELL	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**
2nd-4th	18	33	30	36	33	30	25	21	18
3rd-5th	17	21	18	27	24	19	29	32	32
ALL	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**	**2002**	**2003**	**2004**
2nd-4th	34	47	37	48	46	45	43	42	36
3rd-5th	32	37	36	40	44	47	41	37	36

In summary, the Stanford 9 data from ELL Impacted elementary schools reveal that ELL students score far below the 50th percentile, and far below their peers in the ALL category in their schools. Despite some small gains between 2002 and 2003, third grade ELL student scores declined across all three subtests in 2004. ALL students in ELL Impacted schools also scored below the 50th percentile, saw some declines in 2004, and fall far below their peers in schools with low populations of ELL students. In simulated cohorts, scores, in general, declined for students in both the ELL and ALL categories as they moved up in grade level.

Arizona LEARNS Labels

The labels received under Arizona LEARNS by the ELL Impacted elementary schools from 2002 to 2004 are shown in Table 5. It should be noted that not every ELL Impacted elementary school received a label each year (see earlier discussion for reasons why some schools are excluded). Also, the "Failing" label was not applicable until 2004 (only applied to those schools labeled as "Underperforming" in both 2002 and 2003). From 2002 to 2003 the number of ELL Impacted schools labeled "Underperforming" decreased only slightly, but by 2004, the number was substantially reduced to just eight schools. The number of ELL Impacted schools receiving the "Improving/Performing" label more than doubled between 2002 and 2003. Despite these apparent "successes," it should be noted that seven of the ELL Impacted schools were labeled as "Failing," and no ELL Impacted elementary school achieved the second-highest label "Highly Performing" in 2003 or 2004. This marks a decline in the number of ELL Impacted schools attaining the highest labels.

Table 5: ELL Impacted Elementary Schools Arizona LEARNS Labels, 2002-2004

2002		**2003**		**2004**
Label	**# of Schools**	**Label**	**# of Schools**	**# of Schools**
Excelling	0	**Excelling**	1 (0.6%)	1 (0.6%)
Maintaining	28 (16.1%)	**Highly Performing**	0	0
Improving	78 (44.8%)	**Performing**	112 (64.7%)	164 (91.1%)
Underperforming	68 (39.1%)	**Underperforming**	60 (34.7%)	8 (4.4%)
		Failing	n/a	7 (3.9%)
Totals	*174*		*173*	*180*

NCLB Adequate Yearly Progress Designations

Table 6 shows the number of ELL Impacted elementary schools which made or failed to make Adequate Yearly Progress (AYP) as defined by NCLB. In general, there was a substantial increase in the number of schools making AYP. However, Table 7 provides a more detailed view as to the AYP designations. Fourteen of the ELL Impacted elementary schools went from making AYP in 2003 to failing to make AYP in 2004, and an additional 21 schools failed to make AYP in both 2003 and 2004. Thus, for these schools, representing 20 percent of the ELL Impacted schools, there has been little to no improvement or decline in "academic achievement" as defined by NCLB.

Table 6: ELL Impacted Elementary Schools' Adequate Yearly Progress (NCLB) Designations, 2003-2004

	Yes ***Made AYP***	**No** ***Failed to Make AYP***	**Pending**	**None**
2003	111 (58%)	69 (36%)	2	8
2004	147 (77%)	37 (19%)	0	6

Note: More schools are accounted for in this table than in Table 5 because some schools did not receive Arizona LEARNS labels.

Table 7: ELL Impacted Elementary Schools Changes in Adequate Yearly Progress Designations, 2003-2004

Changes in AYP	**# of Schools**
Made AYP in both 2003 and 2004	96 (54%)
Failed to make AYP in 2003, made AYP in 2004	48 (27%)
Made AYP in 2003, failed to make AYP in 2004	14 (8%)
Failed to make AYP in both 2003 and 2004	21 (12%)

Note: Figures above are for the 179 schools which received an AYP designation in both 2003 and 2004. Percentages do not total 100 because of rounding.

Inconsistencies in the Number of Students Tested

As with the statewide data, we found inconsistencies in terms of the number of third grade ELL students tested on AIMS and Stanford 9 in 2004.[54] The differences in the number of ELL students taking these tests in 2004 are shown in Table 8. In 2004, over 2,000 more students took the AIMS Reading subtest than the Stanford 9 Reading subtest. A similarly large discrepancy is also observed for the AIMS and Stanford 9 Math subtests. As mentioned above, these tests are administered within one to two weeks of each other, thus is doubtful that over 2,000 ELL students had moved or were absent between the administrations of these two tests. Ironically, a change in NCLB allows the exclusion of newcomer ELLs from the AIMS Reading test (but not the Math test). No such allowances are made for the Stanford 9 as Proposition 203 requires all ELLs to be included. Thus, it would have been expected to see a lower number of ELLs tested on AIMS than on Stanford 9. However, the data reveal that the case is the exact opposite. As with the statewide data, it is not clear if scores have been systematically excluded, or if students were actually excluded from taking the tests.

Table 8: Number of Third Grade ELLs from ELL Impacted Elementary Schools Tested on 2004 AIMS and Stanford 9 Reading Tests

Subtest	AIMS	Stanford 9	Difference
Math	11,091	9,159	1,932
Reading	11,012	8,935	2,077

Discussion and Conclusion

Current state education leaders in Arizona have strongly supported Proposition 203 and have been strictly enforcing their own interpretation of this law. They claim that bilingual and English as a Second Language programs in the state were failing to help English Language Learners (ELLs) learn English and that these programs were a barrier to ELL student academic success. These leaders claimed that English-only education would help increase ELL students "soar academically." These same leaders have also been strong supporters of state (Arizona LEARNS, Proposition 203) and federal (No Child Left Behind) high-stakes testing policies, including mandates to include ELL in statewide (English-only) standards-based (criterion) and norm-referenced tests.

The aim of this report has been to analyze available student achievement and school accountability data to determine whether there is any evidence that ELL students are now "soaring academically" as a result of English-only and high-stakes (English-only) testing policies in the state of Arizona. As the data and analyses reveal, there is no evidence that ELL students are experiencing greater academic success (as measured by state tests). Rather, the data show the contrary. The overwhelming majority of third grade ELLs fail the Arizona Instrument to Measure Standards (AIMS) test in contrast to ALL students, and ELLs score well below the 50th percentile on the Stanford 9 and well below students in the ALL category. In addition, overall, test score performance for students has declined between 2003 and 2004, and the gap between ELL and ALL students has failed to close, and in some cases, has even widened. Positive-looking improvements in school accountability labels mask test-score decline in a large number of elementary schools, particularly those with the greatest number of ELL students.

More specifically, on the AIMS test, there was a general pattern evident in both the statewide and the ELL Impacted schools data of higher test scores in 2003, followed by decline in 2004 for both ALL (Category 1) and ELL (Category 2) students on the Reading and Math subtests. In the statewide data for the Stanford 9—a much more stable testing instrument—ELL student percentile rankings rose slightly in 2003 followed by a decline in 2004 while ALL student rankings remained essentially the same.

These data raise the question: To what can these general trends of increases in 2003 and declines in 2004 be attributed? We argue that it is important to understand these changes within the political context of Arizona's educational and school accountability policies. Prior to 2003, under previous State Superintendents of Public Instruction, districts and schools were given much greater flexibility in terms of educational programs schools could offer ELL students. In other words, state policy made it clear that bilingual education was allowed for ELL students through the waiver provisions of Proposition 203, and thus many elementary schools continued (or even expanded) their bilingual programs up through the administration of AIMS and Stanford 9 in 2003. The current Superintendent of Public Instruction and his appointed leaders of state ELL programs began their strict enforcement of Proposition 203 at the beginning of the 2003-2004 school year. Hence, the 2004 AIMS and Stanford 9 scores reflect the first year of strict enforcement of English-only education programs for ELLs. Stated more directly, the improvements in test scores from 2002 to 2003 correspond with a period of greater flexibility for schools in offering ESL and bilingual education, while the decline of scores in 2004 correspond to a period of forced closure for most bilingual programs and mandates for English-only instruction for ELL students.

The one exception to the overall decline in test scores is the sudden jump in the percentage of third grade ELL students passing the AIMS Writing subtest in 2004. Even more unusual is the fact that ELL scores increased at a much higher rate than those in the ALL student category. Indeed, the 2004 third grade AIMS Writing subtest is the only AIMS subtest across the three years that a majority of ELL students passed. This sudden jump in achievement would strike most experienced educators and researchers of ELL education as highly unusual, given that out of the four traditional language skills (listening, speaking, reading, and writing), writing is usually the most difficult skill for ELL students to master (especially younger ELLs).

We would have expected to see ELL students perform higher on the AIMS Math subtest, as, arguably, the language demands of a math test are typically less demanding than those on reading and writing tests. Indeed, as the data above show for the Stanford 9 tests, in all grades and in all years, ELLs scored higher on Stanford 9 Math than all other subtests. While it may be feasible to attribute this jump in Writing scores to English-only education and excellent instruction by teachers, the fact that these types of gains are not evident on any other AIMS or Stanford 9 subtests casts doubt on this explanation. A more logical explanation is that changes were made to items and/or scoring of the AIMS Writing subtest which proved advantageous for ELL students. The only other possible explanations are errors in scoring and/or reporting, or systematic exclusion of scores of lower performing students.

Another possible explanation for the general trend in rising test scores between 2003 and 2004 could be the inconsistencies described above in terms of the number of students tested. As shown in Figures 9 and 10, fewer students were tested (or at least

fewer test scores were publicly reported) in 2003 than in 2002. This is highly unusual given rapid growth of the student population in Arizona. Though it is not clear what (if any) scores are actually missing, it is feasible that exclusion of large numbers of lower scores resulted in the artificial inflation of the 2003 test scores. Indeed, when the number of tested students increased in 2004, most scores declined. Further evidence for missing test score data was shown in Table 7 where even within the same school year (2004), there were large discrepancies in the number of ELL students tested on the Stanford 9 versus the AIMS test in ELL Impacted elementary schools.

Given the complexities of test score data from two different tests (AIMS and Stanford 9) for different subgroups (Category 1 ALL and Category 2 ELL), many have come to rely on the Arizona LEARNS labels and No Child Left Behind (NCLB) Adequate Yearly Progress (AYP) designations. These labels provide easy to understand descriptions of a school's success (or lack thereof) and the public understands these labels to be based on the schools' test scores. Therefore, when policy makers pointed out the declining number of Underperforming schools and the increasing number of Performing and Excelling schools, the public likely assumed this meant that schools had improved their test scores.

However, in 2003 the formula for calculating Arizona LEARNS school labels changed substantially. One of the changes included the exclusion of test scores of ELLs with less than four years of enrollment from the Category 1 (ALL) student data, which are used to determine the labels. The dramatic improvements in Arizona LEARNS labels in the statewide data between 2002 and 2003 are likely due in large part to this exclusion of large numbers of ELL test scores. Hence, Arizona LEARNS labels are no longer

representative of schools' success (or failure) in helping ELL students learn English and academic content. ELL student test scores can also be eliminated from NCLB AYP designations. Many schools avoided having to have an LEP (ELL) subgroup if they had less than 30 ELL students tested at a given grade level on a given AIMS subtest. Many other schools successfully appealed their Failing designation, and were deemed as "making AYP" by excluding AIMS scores for ELL students with less than four years of enrollment.[55] As the data analyzed above shows, while schools were receiving better sounding labels through both Arizona LEARNS and NCLB, a lower percentage of ELLs were passing the AIMS test, and percentile rankings for ELLs on the Stanford 9 were declining on all three subtests.

Ironically, in 2004, while the number of "Underperforming" elementary schools further decreased, and the number of Performing, Highly Performing, and Excelling schools increased, a lower percentage of students in the ALL category passed the AIMS, and there were few changes in the percentile rankings of ALL students on the Stanford 9. This fact highlights the complexity of Arizona's school accountability formula which successfully creates the illusion of educational improvement even in the face of overall declining test scores.

Further evidence for declining academic achievement can be found among the students in the ALL category in the ELL Impacted schools. These students, while ahead of ELL students in their schools, trailed far behind their peers in the statewide data. In other words, a lower percentage of Category 1 (ALL) students in ELL Impacted elementary schools pass the AIMS test, and these students also score at lower percentile ranks on the Stanford 9 than Category 1 (ALL) students in schools with few or no ELL

students. Furthermore, ALL students in ELL Impacted elementary schools declined in their Stanford 9 percentile rankings in Reading in 2004, compared to ALL students statewide where scores remained stable. As suggested above, in ELL Impacted elementary schools, there are likely a much higher percentage of ELLs and former ELLs represented in the Category 1 data. ELL Impacted schools are also typically in lower socioeconomic neighborhoods. The relatively low scores and the decline in performance of ALL students in ELL Impacted elementary schools provides further evidence that education has not improved in those schools with the largest number of ELL students.

In conclusion, there is no evidence to support the claim that ELL students are now "soaring academically" as a result of Proposition 203's requirement for English-only education and the inclusion of ELLs in high-stakes (English-only) testing programs. With bilingual programs for ELLs effectively eliminated in grades K-3, bilingual education can no longer be blamed for low or declining test scores. Rather, there is now growing evidence that English-only education has contributed to these declines in ELL test scores and is contributing to lower levels of academic achievement (as measured by tests), especially in ELL Impacted elementary schools.

As long as federal and state policies mandate the participation of ELL students in high-stakes tests, we encourage the close monitoring of Category 2 (ELL) test scores by policy makers and relevant stakeholders. A system is also needed for mutually exclusive categories of ELL and non-ELL students, and mechanisms are needed to track the progress of ELL students even after they are redesignated as fluent English proficient. Little confidence can be placed on the Arizona LEARNS or NCLB AYP designations as they relate to a school's success in helping ELL students learn English and academic

content. In fact, these labels appear to be masking the harmful affects of the English-only education mandated by Proposition 203. Based on these and other emerging data, we encourage state policy makers to reconsider the narrow requirements and current strict enforcement of Proposition 203. In addition, rather than forcing ELLs to take high-stakes English-only tests only to exclude many of their scores from state and federal accountability formulas, we encourage state policy makers to advocate for changes in the requirements of NCLB, or at the very least, heed NCLB's requirement to test ELLs in the *language* and *form* most likely to yield valid and reliable information about what students know and can do.

Notes & References

[1] Horne, T. (2004). *First annual state of education speech.* Phoenix, AZ: Arizona Department of Education.

[2] Horne, T. (2002, August 10). *Real learning means learning English*. *Arizona Republic.* Retrieved March 5, 2005 from NewBank database.

[3] Horne, T. & Dugan, M.G. (2003). *Implementation of Arizona English language immersion laws*. Phoenix, AZ: Arizona Department of Education.

[4] Wright, W.E. (2004). Intersection of language and assessment policies for English language learners in Arizona. (Doctoral Dissertation, Arizona State University, 2004). *Dissertation Abstracts International, 65-02*, 389.

Wright, W.E. (2005). The political spectacle of Arizona's Proposition 203. *Educational Policy, 19*(4), 1-40.

[5] Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

[6] Arizona Revised Statutes §15-752 English Language Education. Available: http://www.azleg.state.az.us/FormatDocument.asp?inDoc=/ars/15/00752.htm&Title=15&DocType=ARS

[7] Wright, W.E. (2005). The political spectacle of Arizona's Proposition 203. *Educational Policy, 19*(4), 1-40.

[8] Beginning in 2005, the state used a new test—a combined criterion and norm referenced test called the AIMS Dual Purpose Assessment (AIMS-DPA). This combined test has helped to reduce the amount of testing. Unfortunately, it will be difficult to make comparison with previous norm-referenced and criterion-referenced test scores.

[9] Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

Few school districts actually utilized these allowances. Hence, the majority of ELLs still took the Stanford 9 test in English.

[10] Wright, W.E. (2005). The political spectacle of Arizona's Proposition 203. *Educational Policy, 19*(4), 1-40.

[11] Molera (2001). *Guidance regarding the implementation of A.R.S Section 15-751—755 and Flores Consent Order* (CIV 92-596 TUC ACM). Phoenix, AZ: Arizona Department of Education.

[12] Horne, T. (2002, August 10). *Real learning means learning English*. *Arizona Republic.* Retrieved March 5, 2005 from NewsBank database.

[13] Horne, T. (2003). *Media advisory: Superintendent of Public Instruction Tom Horne begins year with new leadership team.* Phoenix, AZ: Arizona Department of Education. Retrieved March 5, 2005, from http://www.ade.state.az.us/pio/Press-Releases/2003/pr01-13-03.pdf

[14] Horne, T. & Dugan, M.G. (2003). *Implementation of Arizona English language immersion laws*. Phoenix, AZ: Arizona Department of Education. Available http://www.ade.state.az.us/pio/Press-Releases/2003/Englishimmersion2-13-03.pdf

[15] Horne, T (2003, July 22). *Superintendent Tom Horne vindicated on bilingual guidelines: 45 monitors to police bilingual*. [Press release]. Phoenix, AZ: Arizona Department of Education. Available: http://www.ade.state.az.us/pio/Press-Releases/2003/pr07-22-03.pdf

[16] Wright, W.E. (2004). Intersection of language and assessment policies for English language learners in Arizona. (Doctoral Dissertation, Arizona State University, 2004). *Dissertation Abstracts International, 65-02*, 389.

Wright, W.E. (2005). The political spectacle of Arizona's Proposition 203. *Educational Policy, 19*(4), 1-40.

This does not mean that bilingual programs no longer exist in Arizona in grades K-3. Some schools have continued to offer dual-language immersion programs, but these programs are only available to native English-speaking students, and students from non-English language backgrounds who are fluent in English or who were former ELL students but have been redesignated as fluent English proficient (FEP). Some bilingual programs in grades 4-12 have also survived and these programs may include ELLs.

[17] Arizona Department of Education (2003). *School improvement in Arizona*. [Press Release]. Phoenix, AZ: Author. Retrieved March 4, 2004 from www.ade.az.gov/services/pio/Press-Releases/2001/pr01-26-01.htm

Arizona Department of Education. (2003). *Arizona LEARNS: A step-by-step guide to calculating an achievement profile*. Phoenix, AZ: Author, Research and Policy Division.

[18] Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

[19] Associated Press. (1998, October 27). High school AIMS test to be in English, board says. *Arizona Republic.* Retrieved March 5, 2005 from NewsBank database.

[20] Kossan, P. (2000, November 21). Keegan puts AIMS on hold indefinitely, cites high failure rate. *Arizona Republic.* Retrieved March 5, 2005, from NewsBank database.

Kossan, P. (2001, August 24). Tamer AIMS on the way; Molera wants 4-year delay, eased graduation rates. *Arizona Republic.* Retrieved March 5, 2005 from NewsBank database.

Barrett, J. & Pearce, K. (1999, November 17). Some items not covered in class. *Arizona Republic.* Retrieved March 5, 2005, from NewsBank database.

[21] Barrett, J. & Pearce, K. (1999, November 17). Some items not covered in class. *Arizona Republic.* Retrieved March 5, 2005, from NewsBank database.

Kossan, P. (2000, November 21). Keegan puts AIMS on hold indefinitely, cites high failure rate. *Arizona Republic.* Retrieved March 5, 2005, from NewsBank database.

Kossan, P. (2004, April 11). Standardized test changing. *Arizona Republic*, p. B1.

Pearce, K. (2000, May 10). AIMS test has errors, teachers, students say. *Arizona Republic*, p. B1.

Pearce, K. (2000, March 18). Mistake made in AIMS scoring. *Arizona Republic*, p. A1.

[22] Arizona State Board of Education. (2004). *State Board of Education minutes (January 26)*. Phoenix, AZ: Arizona Department of Education.

Kossan, P. (2004, January 25). 8th-graders' AIMS math to get easier; State admits error in scoring. *Arizona Republic*, p. B1.

Horne, T. (2003, October 13). Arizona Education not dumbing down. *Arizona Republic*, p. B9.

Horne, T. (2004). *First annual state of education speech*. Phoenix, AZ: Arizona Department of Education.

[23] High failure rates, continuing opposition from students, parents, and educators, and community leaders, and even legal challenges from within the state Legislature may ultimately lead to another delay in the AIMS test as a graduate requirement.

[24] The norm-referenced portions of the new AIMS-DPA exam were not developed by Harcourt Education Measurement, the producer of the Stanford 9 exam. Second and ninth grade students who do not

take the AIMS test, but are nonetheless required to take a norm-referenced exam by Proposition 203, now take the Terra-Nova exam. This change in norm-referenced tests will make comparisons with prior years problematic.

[25] Garcia, D. & Aportela, A. (2000). *Arizona measure of academic progress: A first look at growth in Arizona* schools (Technical document). Phoenix, AZ: Arizona Department of Education.

Pearce, K. (2000a, May 10). AIMS test has errors, teachers, students say. *Arizona Republic*, p. B1.

[26] Arizona Department of Education. (2003). *Arizona LEARNS: A step-by-step guide to calculating an achievement profile*. Phoenix, AZ: Author, Research and Policy Division.

Arizona Revised Statutes (A.R.S.) §15-241 School Accountability. Available: http://www.azleg.state.az.us/FormatDocument.asp?inDoc=/ars/15/00241.htm&Title=15&DocType=ARS.

[27] Arizona Department of Education. (2003). *Arizona LEARNS: A step-by-step guide to calculating an achievement profile*. Phoenix, AZ: Author, Research and Policy Division.

Arizona Revised Statutes (A.R.S.) §15-241 School Accountability. Available: http://www.azleg.state.az.us/FormatDocument.asp?inDoc=/ars/15/00241.htm&Title=15&DocType=ARS..

[28] Arizona Department of Education. (2003). *Arizona LEARNS: A step-by-step guide to calculating an achievement profile*. Phoenix, AZ: Author, Research and Policy Division.

[29] No Child Left Behind, Public Law 107-110, §1111(b)(3)(C)(ix)(III).

[30] Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

[31] See the following for details on problems with AYP for English language learners:

Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

Wiley, T.G., & Wright, W.E. (2004). Against the undertow: The politics of language instruction in the United States. *Educational Policy*, *18*(1), 142-168.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

[32] Arizona Department of Education. (2003). *Arizona LEARNS: A step-by-step guide to calculating an achievement profile*. Phoenix, AZ: Author, Research and Policy Division.

[33] Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

[34] *Ibid*

[35] It could be argued that Category 2 data are used for the purpose of calculating AYP for the LEP subgroup under NCLB. But, as pointed out in this report many ELL scores within the Category 2 data are excluded before these calculations are conducted.

[36] Hakuta, K., Butler, Y.G., & Witt, D. (1999). How long does it take English learners to attain proficiency? *University of California Linguistic Minority Institute Newsletter, 9*(1), 1.

[37] Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Wright, W.E. (2005). English language learners left behind in Arizona: The nullification of accommodations in the intersection of federal and state language and assessment policies. *Bilingual Research Journal*, *29*(1), 1-30.

Wiley, T.G., & Wright, W.E. (2004). Against the undertow: The politics of language instruction in the United States. *Educational Policy*, *18*(1), 142-168.

[38] See the following for critiques of the use of norm-referenced and other tests as indicators of student achievement:

Kohn, A. (2000). The case against standardized testing: Raising the scores, ruining the schools. Portsmouth, NH: Heinemann.

Sacks, P. (1999). Standardized minds: The high price of America's testing culture and what we can do to change it. Cambridge, MA: Perseus Books.

[39] The mean scale score makes for better comparisons as increments between different scores are equal:

Thompson, M., DiCerbo, K., Mahoney, K., & MacSwan, J. (2002). Exito en California? A validity critique of language program evaluations and analysis of English learner test scores. *Education Policy Analysis Archives, 10*(7), Retrieved January 26, 2002, from http://epaa.asu.edu/epaa/v2010n2007/

Unfortunately, at the time of this study, mean scale scores were not made publicly available for the Stanford 9 for all years and cohorts of interest.

[40] A student is considered as making one year's growth if their stanine score remains constant or improves on the subsequent year's Stanford 9 test.

[41] Despite the stability of the Stanford 9, there are still some problems in using it for year-to-year comparisons. Research has shown that the longer a norm-reference test is in use, the more familiar teachers and students become with it, and scores typically increase each year and then eventually plateau. When this happens, test makers typically re-norm the test. Harcourt Measurement has already produced an updated test with new norms—the SAT-10.

See, e.g,:

Tindal, G. & Haladyna, T. H. (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 27-48). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Butler, Y.G., Orr, J.E, Gutierrez, M.B., & Hakuta, K. (2000). Inadequate conclusions from an inadequate assessment: What can SAT-9 scores tell us about the impact of Proposition 227 in California? *Bilingual Research Journal*, *24*(1 & 2), 141-154.

[42] Proficiency in English is determined by an English language proficiency test administered each year. ELLs who meet the test publisher's criterion for fluency are redesignated as "fluent English proficient," and are no longer counted as ELL students. While many students make progress in learning English each year, only a small percentage of ELLs are redesignated each year.

[43] Wright, W.E. (2004). Intersection of language and assessment policies for English language learners in Arizona. (Doctoral Dissertation, Arizona State University, 2004). *Dissertation Abstracts International, 65-02*, 389

[44] Kossan, P. (2003a, September 17). Arizona relaxes school rankings; More lenient formula will save state money. *Arizona Republic*, p. A1.

[45] Kossan, P. (2003b, October 12). Doubts cloud school rankings; Criteria eased under new formula. *Arizona Republic*, p. A1

[46] *Ibid*.

[47] In 2003, there were 832 schools (43%) that did not receive a label because the school was new, small, an alternative school, or only had students in grades K-2. Many of these types of schools were also

excluded in 2002. In 2004, a system was created to assign these schools a label, hence the large increase in number of schools between 2003 and 2004.

[48] Arizona Department of Education. (2004). *Arizona's school accountability system technical manual. (Volume 1 and 2): Arizona Learns achievement profiles*. Phoenix: Author

[49] Horne, T. (2003, October 13). Arizona education not dumbing down. *Arizona Republic*, p. B9.

Horne, T. (2004). *First annual state of education speech*. Phoenix, AZ: Arizona Department of Education

Kossan, P. (2004, April 11). Standardized test changing. *Arizona Republic*, p. B1

Kossan, P. (2004, January 25). 8th-graders' AIMS math to get easier; State admits error in scoring. *Arizona Republic*, p. B1.

[50] Horne, T. (2005), State of Education 2005 (p. 4). Phoenix, AZ: Arizona Department of Education. Available: http://www.ade.state.az.us/administration/superintendent/2005StateofEducation.pdf

[51] The number of "Underperforming" schools cited by the Superintendent of Public Instruction in his State of Education speech appear inconsistent with those reported by the ADE as shown in Table 3. However, the Superintendent was only referring to those schools that had been "Underperforming" for two years in a row and thus were facing the "failing" label if they did not improve.

[52] It is possible that many of the 110 Underperforming schools only received a label for the first time in 2004, given the large number (40%) of schools that did not receive a label in 2003. But as the next section on ELL Impacted schools shows, there are schools that were previously "Performing" and are now "Underperforming."

[53] This in no way suggests that students in lower socioeconomic neighborhoods are less capable or "intelligent" as students in higher SES neighborhoods. Schools in lower SES neighborhoods typically receive less funding, pay lower teacher salaries and thus have difficulty recruiting and retaining experienced credentialed teachers, serve a greater percentage of language minority students, and have high student mobility rates. All of these factors can affect test scores.

[54] We only focus on a single year (2004) for the ELL Impacted elementary schools due to the high mobility rate these schools typically experience.

[55] These strategies to exclude ELL test scores from NCLB AYP designations may not work for long. While they successfully helped to lower the number of schools "failing" to make AYP between 2003 and 2004, the cut score for determining AYP (called the Annual Measurable Achievement Objective) is set to rise, meaning schools will have to greatly increase the number of students (including ELLs) passing the AIMS. Furthermore, these strategies delay rather than prevent the negative impact of ELL test scores. After four years, these students' scores can no longer be excluded. At this point, the inadequacies of their English-only education may become even more apparent.