District Grade-Span And Math AIMS Scores

Daniel Hunting

Abstract
In April 2005, Arizona Governor Janet Napolitano recently signed Senate Bill 1068, authorizing the formation of the School District Redistricting Commission (SDRC) that will be charged with making recommendations about combining some of the state's 200 school districts into units of more logical size. This effort to combine districts is based on arguments of fiscal and logistical efficiency and local control. Conspicuously absent from the legislation is any mention of the educational impacts of such changes. This paper examines the effect of district grade-span on high school math Arizona's Instrument to Measure Standards (AIMS) scores in 2003 in an effort to determine whether Kindergarten through 12th grade (Unified) districts performed differently from 9-12th grade (Union) districts. Analysis of the data shows no correlation between district configuration and math AIMS performance. Policy makers can be reasonably sure that changes to district grade-span will have little or no effect upon student achievement.

Introduction

Arizona is somewhat unusual in that it allows three types of school districts: Kindergarten through 12th grade districts known as *Unified* districts; K through 8th grade, or *Elementary* districts and 9th through12th grade, or *Union high school* districts (*Arizona Republic*, Feb. 27, 2005). There are over 200 school districts in the state, some consisting of a single small school, and some with many very large schools. On April 25, 2005, Governor Janet Napolitano signed Senate Bill 1068, which authorized the formation of a commission to make recommendations about the possibility of combining some of these districts into more rational forms (ibid). This legislation was designed to increase fiscal efficiency through the elimination of duplicate services, such as administration, in areas that are served by separate elementary and high school districts. The educational effects of these changes were apparently not considered when drafting this legislation.

It is reasonable to think that there might be effects on student learning from changes in district grade configuration. Alspaugh (1998) demonstrated that standardized test scores in Missouri tended to decrease when district grade configuration required the students to attend more schools, i.e. middle and junior high schools in addition to primary and high schools (p. 24). It seems that a likely cause for these effects is the change in curriculum associated with changing schools. It also seems reasonable to think that these changes would be exacerbated by changing *districts* as well as *schools*, and that students who had to change districts upon entering high school would have lower math AIMS scores in high school. Maintaining a consistent curriculum between elementary and high school districts must be difficult, especially in cases where multiple elementary districts feed into one high school district. The Phoenix Union High School District, for example, accepts students from 13 different Elementary districts, each with a different curriculum. A district such as this might have a harder time getting their students to perform well compared to those districts that are able to maintain a consistent curriculum.

Although the Arizona Department of Education sets statewide standards for each grade level in Arizona schools (Arizona Department of Education, 2005), choice of actual curriculum is left up to the individual districts and/or schools. Compliance with the state requirement to align their curriculum to the standards is unknown, but personal observation suggests that some schools are reluctant to change their programs to meet these standards. Under the federal No Child Left Behind program, all states are obligated to set standards and regularly test that those standards are being met. Arizona has chosen to make mastery of those standards, as measured through the AIMS test, a high school graduation requirement.

Literature review

A literature review of the subject turned up surprisingly few objective articles analyzing what influences outcomes on high-stakes tests (HSTs). There are even fewer articles available about the impact of district grade configuration on learning. This is probably because Arizona is one of only three states that still maintain separate elementary and high school districts (*Arizona Republic*, Feb. 27, 2005). The first article I came across that looked promising was a long piece by Amrein & Berliner (2002). The abstract for *High Stakes Testing, Uncertainty, and Student Learning* promised a brief history of high-stakes testing and an analysis of 18 states to see how the tests affect actual learning. The authors have an apparent bias against high-stakes testing which makes evaluation of their work difficult. In the first sentence of the abstract, it warns of the 'severe consequences' associated with high-stakes testing, although the authors later admit that it is very difficult to answer the question of whether learning is increased by high-stakes tests. Throughout the paper, the authors seem intent on making assertions, in the absence of data, that these tests discriminate against ethnic minorities and, furthermore, that the tests are somehow *designed* to be oppressive to various segments of society.

The authors looked at student scores on four widely-used tests (SAT, ACT, NAEP and AP) to see how these scores changed after the introduction of high-stakes tests in 18 states. They speculated that, if the HSTs were effective at increasing student learning, then scores on the other four tests should increase as well. They failed to consider the fact that HSTs are minimum standards tests applied universally to all high-school students, while the SAT, ACT and AP tests are taken by a self-selected group of college-bound students. The National Assessment of Educational Progress (NAEP) test is a somewhat better comparison, but there are selection issues with this test as well and, like SAT and ACT, it is a norm-referenced test, not a criterion-referenced test like the HSTs the authors are comparing it to.

The statistical analysis that Amrein & Berliner apply to their data is quite elementary, simply tracking whether scores on the four tests in the 18 states went up or down in relation to national averages. They did not account for the magnitudes of any changes and used no regression, no analysis of variance, or any other measure whatsoever. They made no attempt to control for socio-economic status or any other variables and attributed all of the changes in the standardized test scores to the implementation of high-stakes tests.

Next I came across a response to Amrein & Berliner by Bruce Thompson. Thompson (n.d.) may have his own agenda to promote, but at least his biases are not quite so obvious. His rebuttal to Amrein & Berliner is focused on their flawed analysis of their data. He reviews their methodology and shows that their techniques lost a lot of information by converting continuous data into essentially binary data (2). Thompson then re-analyzes the data with what he claims are more appropriate tools and comes to the conclusion that there is no association between HST implementation and scores on the SAT, AP and NAEP, and a weak negative association with the ACT (8). Thompson's analysis is more sensible than Amrein & Berliner's, actually computing p-values and confidence levels. It is interesting to note that, while Amrein & Berliner saw a strong negative relationship between HST implementation and standardized test scores, Thomson basically saw no association, while Raymond & Hanushek (2003), of the Hoover Institution found a strong *positive* association from the exact same data (53).

An article by Slovacek, Kunnan & Kim (2002), while not as directly related to this topic, as the others, uses regression techniques that are very similar to the ones used here. They looked at how charter schools in California serving low-socio-economic status (SES) students compared to traditional public schools. In comparison to the other articles I looked at, this article was much more academically rigorous. Their methodology was explained thoroughly, and SPSS outputs were presented as tables in the final article. The dependent variable in this study was the Academic Performance Index (API), which was correlated with six independent variables; participation in Free or Reduced Price Lunch Program (as a proxy for SES), percentage of English language learners, percentage of student turnover in the school, percentage of fully credentialed teachers, percentage of teachers with emergency credentials, and number of students enrolled in the school.

Although it is clear that these authors are supporters of charter schools, their methodology is at least more reliable than that of Amrein & Berliner. One thing did strike me as curious: they produced a regression equation for all California schools using the above variables, and a separate, slightly modified equation, dropping the emergency certification variable, which was run on data for charter schools. Several variables that were thought to be significant for charter schools were considered, but ultimately not included in the final equation. I was left wondering why the authors didn't just include a dummy variable indicating 'charter / not-charter' and run one regression on the entire data set. That would better quantify the effects of charter schools in relation to overall outcomes.

Specification of the theoretical model

The dependent variable for this study is mean scale math scores from the 2003 high school AIMS test, averaged at the district level. AIMS is actually three separate tests, covering math, reading and writing. Since there is no composite score indicating overall school performance, I decided to choose the math test as the benchmark of district performance. Scores on the math portion of the AIMS test are reported in the ADE data as Mean Math Scale Score (MMSS). Math scores are the most appropriate measure for this examination for two reasons. First, the basic skills of reading and writing will have been taught in primary school. Subsequent years of schooling will refine these skills, but major new reading and writing skills will not be added at the high school level. The math portion of the AIMS test, however, tests skills that are generally introduced to students in middle and high school grades. If there is a positive benefit to be found in a Unified school district, perhaps from having a curriculum that is more continuous across grades, that benefit may be more observable in math scores. Secondly, the math portion of the AIMS test is more significant from a political perspective. There has been considerable controversy surrounding the imposition of a high-stakes test as a condition for high school graduation in Arizona. Passing rates for the math portion of AIMS are far lower than in reading and writing, so most of that controversy is centered on math scores (Bland, 2005). An analysis that showed that district configuration had a sizable effect on math AIMS scores would have genuine policy implications.

The primary independent variable of interest for this study is district grade span, denoted by the dummy variable **K12**, where one represents a Unified district with a kindergarten through twelfth grade program and zero is a Union High School District covering ninth through twelfth grade only. Other independent variables are as follows:

- **FRL** = Percentage (10% = 10) of students on Free and Reduced Lunch Price Program in each district. Okpala, Okpala and Smith (2001, p. 111), and other researchers have used this as a proxy for socio-economic status.

- **DistSize** = Number of students tested in each district on the math portion of the AIMS test. Districts vary greatly in size, so heteroskedasticity is a possibility with this variable. The Park test will be used to evaluate whether heteroskedasticity is present.

- **Ethnicity** will be represented by five variables, each expressing ethnicity as a percentage (10% = 10) of the students in each district taking the test, with White as the omitted case. These are the same categories expressed in the AIMS data from ADE:
    1. **Hispanic**
    2. **Black**
    3. **American Indian**
    4. **Asian**
    5. **Other**

- **PPE** is per-pupil expenditures, in dollars, for each district, as reported in data on the ADE's web site.

- **Rural** is a dummy variable that is set to one if the school is in a rural area. Most districts in the state, except for those in metropolitan Phoenix and Tucson, plus those serving Flagstaff, Prescott and Yuma, were deemed to be rural in nature.

- **TeachExp** is the experience, in years, of the average teacher in the district.

Table 1: Sign hypotheses

Variable	Expected Sign	Rationale
K12	+	This is the primary variable of interest in this study. Assuming that AIMS success is at least partly reliant upon a curriculum that specifically teaches the standards that the test measures, and that a Unified district is more likely to maintain a rational curriculum flow across grades than split districts, this sign should be positive.
FRL	-	Fowler & Walberg, noted that socio economic status was negatively associated with student success, so this coefficient will be negative.
DistSize	-	Fowler & Walberg observed district size to be negatively correlated with student outcome.
Hispanic	-	Hispanic, Black, and American Indian students generally score lower on standardized tests (Drake & Forester, 2003).
Black	-	Okpala, Okpala & Smith observed negative relationship between African Americans and math achievement scores.
American Indian	-	
Asian	+	Asian students tend to score higher on standardized tests.
Other	?	Because this is a catch-all ethnicity, its sign is difficult to predict.
PPE	+	This variable represents per pupil expenditure, and it seems reasonable that money spend on students has a positive effect.
Rural	-	Issues such as hours-long bus rides for students and an inability to attract and retain well-qualified teachers indicate that rural schools will not perform as well on the AIMS test.
TeachExp	+	Teachers who are more experienced should have higher performing students.

Discussion of the data to be analyzed

Data for this study came from the Arizona Department of Education (ADE) web site that provides a rich source of information on Arizona schools. This data was downloaded and imported into SPSS for analysis. The specific data was drawn from the 2003 tenth grade AIMS results, aggregated at the district level for traditional public schools. According to the ADE's online 'frequently asked questions' page, private schools are exempt from AIMS, but charter schools, which receive state funding, must administer the test. Charter schools were omitted from this study for two reasons. First, they often do not follow the same grade-span configuration as traditional public schools. A charter school may choose to serve students from sixth through twelfth grade, for example. This makes comparisons to traditional districts inappropriate. More importantly, charter schools are largely immune from public policy control (ADE, 2006), so even if an effect were observed at this level it would be very difficult to do anything about it.

Data from 2003 was used because it is the most recent complete data set available. While much 2004 data is available, data is still missing for many schools. Of the 106 high school districts in this study, 91 are Unified (K-12) districts and 15 are Union high school (9-12) districts. The analysis was limited to Category 1 scores. According to ADE's online documents, Category 1 data contains scores only for those students who speak English as their first language. Category 2 data contains scores for all students whose first language is not English, regardless of their level of fluency. Because there was no way to gauge the English language proficiency levels of students recorded under the Category 2 data, it was decided that more accurate results would be obtained with Category 1 data. All of this data was imported into an Excel spreadsheet in preparation for regression in SPSS.

Category 1 AIMS data can be broken down by district. Within each district, the total number of students taking each portion of the test is reported, and those numbers themselves are further broken down by ethnicity. The mean scores for each district on the reading, writing and math portions of the test are reported.

The percentages of students enrolled in the free-and-reduced lunch program are available from the Child Nutrition Programs portion of the ADE website. October 2003 data was downloaded in Excel format. This data is aggregated at the school level, so it was necessary to combine these figures in the spreadsheet to obtain district-level data. Expenditures per pupil data were also brought into the spreadsheet from the ADE's *Annual Financial Report for Fiscal Year 2002 – 2003.* Teacher experience figures were obtained by downloading the Teacher Experience Index report for each of the districts in Adobe Acrobat format, reading the average teacher experience line, and inputting that number into the spreadsheet.

Table 2- Descriptive Statistics

Variable	Mean	Variation		Standard Deviation
		Min	Max	
MMSS	481	456	527	13.85
FRL (%)	56	07	100	20.94
DistSize (Students)	627	13	7294	1187.62
Asian (%)	1.20	0.00	6.95	1.60
Black (%)	2.05	0.00	12.78	2.91
Hisp (%)	25.47	0.00	95.97	23.06
Indian (%)	16.44	0.00	100.00	30.25
Other (%)	4.87	0.00	23.02	4.33
PPE ($)	3752	647	8294	992
TeachExp	8.86	4.90	11.40	1.40
n = 106				

Descriptive statistics for all variables are listed in Table 2, above. Of special note is the skewed distribution of district size. Arizona has many small districts with just a few hundred students, and a few that are extremely large. The implications of this distribution on academic achievement in relation to district grade span are unknown and may merit further study. It is also worth noting the relatively small standard deviation of the dependent variable, Mean Math Scale Score.

The **K12** dummy variable was encoded by looking at the name of each district. Those that had “Unified District” in their names received a '1' and those with “Union High School District” were labeled '0.' The **Rural** variable was somewhat more subjective. All districts were determined to be rural and coded as ‘1,’ except those in the metropolitan Phoenix and Tucson areas, Flagstaff, Prescott and Yuma (Table 3, below).

Table 3- Dummy Variables

Dummy Variable	1	0
K12	91	15
Rural	77	29
n =106		

Results

Ordinary least squares linear regression was run on this data using SPSS v12.0. The program was also instructed to produce a table of residuals, for later use in performing the Park test for heteroskedasticity. The following equation emerged from the regression run:

$$MMSS = 501.89 - 0.607K12 + 0.004FRL + 0.001DistSize + 2.131Asian - 1.047\ Black - 0.249Hisp - 0.214Indian - 0.349Other - 0.001PPE - 6.569Rural - 0.148\ TeachExp$$

To account for the possibility of heteroskedasticity, the table of residuals was immediately put in to Excel, along with data for the suspected proportionality factor, **DistSize**. Formulas were then applied to calculate the log of the squared residual and the log of district size for each observation. These figures were then imported back into SPSS and another regression run as the Park test. This test showed that the t-score of the suspected proportionality factor was –0.065, which is small enough that we can say that there is no evidence of heteroskedasticity of error in the original regression.

Full results of the estimated coefficients, standard errors and t-statistics are shown in Table 4, below.

Table 4 - Results of the regression

Variable	***Coefficient***	***Std. Error***	***t***
(Constant)	501.886	8.774	57.202
K12	-0.607	2.980	-0.204
FRL	0.004	0.064	0.061
DistSize	0.001	0.001	1.129

Table 4 - Results of the regression

Variable	Coefficient	Std. Error	t
Asian	2.131	0.821	2.595*
Black *	-1.047	0.423	-2.478*
Hisp	-0.249	0.052	-4.812
Indian	-0.214	0.048	-4.489*
Other	-0.349	0.218	-1.599
EPP	-0.001	0.001	-1.006
Rural	-6.569	3.321	-1.978
TeachExp	-0.148	0.786	-0.188
** Statistically significant at or above the 5% level, 1 tailed*			

The equation as a whole has a respectable R^2 of 0.599, an adjusted R^2 of 0.552 and an F-statistic of 12.745, which is well above the critical value required for this equation. However, the primary independent variable of interest, **K12**, had a very small magnitude with an unexpected sign, and was of no statistical significance. A minuscule t-statistic and standard error several times the magnitude of the coefficient, indicates that district configuration has little impact on the equation. Surprisingly, most of the other variables also had small coefficients and poor levels of significance.

Socio-economic status, as measured by participation in the Free and Reduced Price Lunch program (**FRL)** had almost no magnitude or significance. Size of district (**DistSize),** expenditures per pupil (**EPP**) and teacher experience (TeachExp) were all similarly trivial in their contribution to math AIMS scores. With both magnitude and significance levels so low, it is impossible to say that consideration of these variables adds anything to our understanding of math AIMS scores.

Racial and ethnic variables showed stronger levels of significance, with **Asian, Black** and **Hisp** being significant at or above the 5% level. But, there were very low levels of magnitude for each of these coefficients. The maximum effect was seen with Asian students, where a 1% increase in the number of Asian students would raise the district average score only two points on a test where the statewide mean score is 481, all other factors being equal.

The coefficient tied to the dummy variable **Rural** had the largest magnitude of any in the equation, and was significant at nearly the 5% level, one-tailed. The equation indicates a likely drop in AIMS math scores of 6.6 points for a rural district, compared to a similar urban or suburban district.

Conclusion

The weakness of this regression equation, with low magnitudes and confidence in nearly all the independent variables, can be explained by noting the very narrow variation of the dependent variable. High school Math AIMS scores in 2003 had a mean value of 481 and a standard deviation of less than 14. There is simply not enough difference in the scores to show much effect from any combination of variables. The scores are clustered tightly enough that any effects from the independent variables considered here are likely to be very small. But this is not to say that there are no policy implications from these findings. First, and most directly related to the passage of SB 1068 and the possibility of combining districts, it appears that the commission does not need to consider the educational effects of their decisions very much. They can feel free to base their choices on fiscal and other concerns, as the legislature intended.

There are a number of secondary inferences that may be reached from an analysis of this data as well. The relatively tight clustering of math AIMS scores, with such a small standard deviation, may indicate that districts have already substantially aligned their curricula to the state standards. If all districts were teaching to the same standards, and doing so with similar effectiveness, then one would expect the scores to be tightly grouped.

Although there are significant variations in per pupil funding in this data set, this variation appears to have no effect on AIMS scores. This may indicate that the funding formulas for school districts, which have been criticized for being inequitable, are actually quite fair. Wealthy suburban districts with computers in every classroom performed about as well as poorer districts with lower funding levels. Similar reasoning applies to the free and reduced price lunch program: it could be argued that the program has had its desired effect of boosting the scores of poor students to meet those of richer students. One could also infer that the lack of significance and magnitude in the teacher experience variable indicates that the districts are doing a good job of recruiting able young teachers who are as capable in the classroom as their veteran colleagues. More directed research will, of course, be required to reach firm conclusions about any of these hypotheses. The troubling relationship between rural districts and lower AIMS scores should also be investigated further to ascertain the nature and cause of this variation.

Finally, there is opportunity to refine and further this research. Since averaging the data at the district level absorbed much of the variability in the scores, it might be productive to run a similar regression on school-level data. School-level data is available from the ADE website, so it would just be a matter of matching schools to what type of district they are operating under and running the same model. It is possible that there are more apparent effects at this level. The same regression could also be run on the reading and writing AIMS scores to see if differences emerge. The district-level reading scores are just as tightly clustered as math, and the writing scores have only a slightly higher standard deviation, so I would expect the results to be similar, but if the data is analyzed at the level of the school, interesting differences may be detected.

References

Alspaugh, J.W. (1998) Achievement losses associated with the transition to middle and high school. *The journal of educational research. 92(1). pp 20-25.*

Amrein, A. L. & Berliner, D. C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10(18).* Retrieved March 4, 2005 from http://epaa.asu.edu/epaa/v10n18/ .

Arizona Department of Education (n.d.) *Frequently Asked Questions Regarding Arizona's Instrument to Measure Standards (AIMS).* Retrieved May 1, 2005 from http://www.ade.state.az.us/standards/gradtestfaq.asp.

Arizona Department of Education. (2004) *Annual Financial Report for Fiscal Year 2002 – 2003.* Retrieved April 23, 2005 from http://www.ade.az.gov/annualreport/annualreport2003/PerPupilExpenditures/CurrentExpenditures.aspx

Arizona Department of Education. (2004) *Child Nutrition Programs.* Retrieved April 23 2005 from http://www.ade.az.gov/health-safety/cnp/frpercentages/

Arizona Department of Education. (2005) *Academic K-12 Standards.* Retrieved October 19, 2005 from http://www.azed.gov/standards/contentstandards.asp.

Arizona Department of Education. (2006) *Charter Iinformation.* Retrieved June 12 2006 from http://www.ade.state.az.us/charterschools/info/faqs.asp .

Arizona Department of Education. (2005) *Definition of Categories 1 and 2 in 2002, 2003 and 2004 AIMS results* Retrieved May 1, 2005 from http://ade.az.gov/Definition_of_Cat1and2_AIMS_02_04.pdf

Arizona Republic. (2005, February 27) Unifying school districts will mean better education for kids. Editorial. p. V1-V2s

Arizona State Senate (2005), *Fact Sheet for SB 1068: School District Redistricting Commission.*

Bland, Karina, Anne Ryman and Carrie Watters (July 15, 2005) Aims Results Studied; Improvements Sought, *Arizona Republic.* Retrieved June 12, 2006, from Access World News.

Drake, Tonya, M., and Forester, Christine, A., *(2003). Arizona Minority Student Success Report*. Arizona Minority Education Policy Analysis Center, Spring 2003.

Fowler, W.J. & Walberg, W.J. (1991) *School Size, Characteristics and Outcomes.* Educational Evaluation and Policy Analysis. 13(2) pp. 189-202

Okpala, C.O., Okpala, A.O, & Smith, F.E. (2001) *Parental Involvement, Instructional Expenditures, Family Socio Economic Attributes, and Student Achievement.* The Journal of Educational Research, 95(2). p. 110-115

Raymond, Margaret E. & Hanushek, Eric A. (2003). *High-stakes research.* Education Next, 3(3). Retrieved March 7, 2005 from http://www.educationnext.org/20033/pdf/48.pdf.

Slovacek, Simeon P., Kunnan, Antony j. & Kim, Hae-Jin. (2002). *California charter schools serving low-SES students: an analysis of the academic performance index.* Retrieved March 7, 2005, from http://www.calstatela.edu/academic/ccoe/c_perc/rpt1.pdf

Thomson, Bruce, (Undated). *High Stakes Testing and Student Learning:A Response to Amrein & Berliner*. Retrieved March 7, 2005, from http://richardphelps.net/BTArticle-HighStakesTestingx.htm.