Oh, the Places They'll Go!

SCHOOL RATINGS: IMPROVING THE DATA IN DATA DRIVEN DECISION MAKING

EXECUTIVE SUMMARY

The report's findings are a combination of good news and bad news. Student Growth Percentiles appear to be a fairer method to measure a school's contribution to student learning. However, the current use of growth does not meaningfully alter the negative relationship between the level of poverty in the school and the final school rating.

The Center for Student Achievement's guiding principle is that all students deserve high quality public school options, whether they attend a district or charter school. In order to determine the extent to which we are near or far from our goal, that every Arizona student attend a high quality school, we find ourselves having conversations about what school quality looks like and how to measure it. Each conversation leads us to the state's A-F Letter Grades and whether or not these are valid and reliable measures of school quality.

The A-F Letter Grades, published each year by the Arizona Department of Education, have evolved over time—from the initial *Excelling* through *Failing* labels to the current *A, B, C, D,* and *F* letter grades—but their purpose has remained the same: to convey a judgment as to the school's quality or effectiveness at educating students. The message is quite simple: *A* schools are doing a great job at educating students and *D* schools are not. If a school continues to be a *D* school for three years, then it is an *F* school—and failing students altogether.

While the general message may be simple, the subsequent uses of these school labels are not. In Arizona, the school accountability designations are more than simply informational; they are an explicit tool for school reform. Presently, they are used to determine whether a school is in need of instructional improvement strategies and to inform parents' and students' choices in an educational marketplace. Recent policy proposals would use the ratings to provide incentives or sanctions to deserving schools. What each of these uses has in common is the assumption that the school accountability ratings are actually measuring what we think they are measuring: school quality.

This report examines whether the trust our public policy has placed on the school ratings is well-founded, independent of how school ratings are used. It makes explicit unspoken assumptions underlying the use of school designations in any way— primarily that the designations do, indeed, reflect school quality and that the distinctions made by the designations (an *A* vs. a *B*, for example) are meaningful.

THE CHALLENGE

Unless otherwise mitigated, poverty can have a significantly deleterious effect on learning. Students growing up in poverty often arrive at school with challenges that adversely affect their ability to learn or to learn at the same rate as their wealthier

CENTER FOR
STUDENT
ACHIEVEMENT

counterparts. The relationship between poverty and measures of achievement (e.g., percent of students passing a state's standardized test) has long been a limitation of measurement of student achievement in education. This is not to say that schools do not make a difference; they do, but it is often difficult to measure their effect.

This is, in part, the reason why so much attention has been given to measures of student "growth," such as Value-Added Models and Student Growth Percentiles. These new methods allow researchers to "control for," or set aside statistically, the relationship between poverty and indicators of achievement in order to measure the effect that a school or teacher has had on an individual student.

The challenge for the Arizona accountability model, and any other accountability model, is to adequately control for the effect of poverty on the final school rating. An accountability system that fails to properly account for this would not only fail to measure school quality; but could place certain schools or educators at a relative advantage or disadvantage within the system. Accurate measurement and fairness are particularly important in a policy environment that seeks to attribute the results to educators and attach significant rewards or consequences.

THE FINDINGS

The report's findings are a combination of good news and bad news. Student Growth Percentiles appear to be a fairer method to measure a school's contribution to student learning. However, the current use of growth does not meaningfully alter the negative relationship between the level of poverty in the school and the final school rating.

Despite the inclusion of growth, the practical result of the accountability formula is yet another measure of performance that is closely related to poverty. The model fails to adequately control for the effect of poverty on indicators of achievement in order to measure the school's contribution to learning, therefore limiting its utility as an indicator of school quality. Furthermore, the point differences resulting from the accountability formula, as currently implemented, are large enough to make a difference in the final rating of the school.

RECOMMENDATIONS

This report details a number of possible changes to the Traditional and Small school accountability models designed to improve them as measures of school quality. All of the suggestions come from components that already exist in other parts of the accountability system, thus requiring no new measures, only the expanded use or a re-configuring of existing data. The Alternative school model, in particular, offers some interesting opportunities for the expanded use of existing measures for all schools.

A decade after their first publication, school accountability ratings are deeply embedded in Arizona's public education landscape. Their widespread use, particularly by policymakers, suggests that they are here to stay. Therefore, it is in our best interest to ensure that they do indeed measure school quality and that the distinctions made by the various labels (*A* vs. *B*, for example) are meaningful. The current A-F Letter Grades fall short in both regards, despite the strengths of some of the key components in the formula.

While the implementation of any or all of this report's recommendations would improve the model's ability to measure school quality, the model remains heavily reliant on one assessment—the Arizona Instrument to Measure Standards. Although data from AIMS are configured in different ways such as percent passing and student growth percentiles, it remains a single and limited view of what occurs in a school.

In order to better inform our work, the Center for Student Achievement will continue our efforts to identify other valuable indicators that can contribute to a wider view of school quality. We encourage the community to join us.

Oh, the Places They'll Go!

By Anabel Aportela, Ph.D. and Ildiko Laczko-Kerr, Ph.D.

CENTER FOR
STUDENT
ACHIEVEMENT

INTRODUCTION

The Center for Student Achievement's guiding principle is that all students deserve high quality public school options, whether they attend a district or charter school. In order to determine the proximity to our goal, that every Arizona student attend a high quality school, we find ourselves having conversations about what school quality looks like and how to measure it. Each conversation leads us to the state's A-F Letter Grades and whether or not these are valid and reliable measures of school quality.

The A-F Letter Grades, published each year by the Arizona Department of Education, are the latest iteration of the school accountability law passed in 2000 as part of Proposition 301. The ballot measure, which led to a 0.6 percent increase in the state's sale tax, provided additional funding for education in exchange for increased accountability for schools and teachers; teachers became subject to performance pay plans and schools became subject to school accountability designations[1].

School accountability ratings have evolved over time—from the initial *Excelling* through *Failing* labels to the current *A, B, C, D,* and *F* letter grades—but their purpose has remained the same: to convey a judgment as to the school's quality or effectiveness at educating students. The message is quite simple: *A* schools are doing a great job at educating students and *D* schools are not. If a school continues to be a *D* school for three years, then it is an *F* school—and failing students altogether.

While the general message may be simple, the subsequent uses of these school labels are not. In Arizona, school accountability designations are more than simply informational; they are an explicit tool for school reform. In fact, to date they have been a meeting point for two rather distinct views of school improvement. Both views find utility in the accountability designations, but in different ways.

In the standards-based theory of school improvement, academic standards explicitly state the expectations for what students should know and be able to accomplish. In this theory, the student assessments measure the extent to which students have mastered those standards, and the school ratings serve as an instrument to identify schools that are meeting expectations and schools that are in need of additional supports and interventions. To this end, the original school accountability legislation included the creation of Solutions Teams, managed by the Department of Education, to be deployed to *Underperforming* schools as part of the state's school improvement strategy. In this case, the school label serves as a trigger for additional supports from the school, district, and state with focused attention on quick improvement. While continued failure to improve does, under this improvement model, lead to increasingly more onerous sanctions for schools, this happens after multiple years of reform efforts.

In the market theory of school improvement, in which students and parents are seen as consumers of education, the school designations serve as indicators of quality to the public at large, but particularly to parents and students making choices about enrollment. A positive school designation signals to parents that this school is where they want to enroll their child and a negative school designation signals to the educators at the school that they need to change and improve in order to compete for students. The school designations serve to apply market pressure on schools and to "hold accountable" the educators working in them for the quality of their schools.

While the evaluation of these theories of school improvement is outside of the scope of this paper[2], what we do know is that after ten years of school accountability ratings, between five and ten percent of Arizona schools received the lowest designation in any given year (*D/F* beginning in 2011 and *Underperforming/Failing* prior to 2011). However, the vast majority of these Arizona schools continue to enroll students[3].

Perhaps these disappointing results help explain recent policy proposals that would introduce a third theory of school improvement into the accountability landscape—one categorized in this report as incentive-based. The incentive-based theory reflects a belief that educators can be or need to be incentivized (or threatened) to improve their performance. Two recent examples of incentive-based policy proposals are "parent trigger laws," where the school designation can "trigger" school takeover by parents or other community organizations and "performance funding" systems, in which a portion of school funding is determined by the school accountability rating. If faced with the possibility of more school funding or the potential loss of control of their school, the theory predicts, educators will behave in ways that will improve student achievement results.

Regardless of whether the underlying theory of improvement holds that educators need standards or market competition or sufficient motivation, each of these approaches relies on the school designation as a key instrument for school improvement. Each approach rests, in large part, on the belief that the school rating captures meaningful information about school quality and the publication of the rating will elicit the desired behavior from parents, students, teachers, and/or administrators.

This report examines whether the trust our public policy has placed on the school ratings is well-founded, independent of how school ratings are used. It makes explicit unspoken assumptions underlying the use of school designations in any way— primarily that the designations do, indeed, reflect school quality and that the distinctions made by the designations (an *A* vs. a *B*, for example) are meaningful.

The assumption that school labels accurately measure school quality is perfectly reasonable. School accountability legislation specifically calls for the designations to be determined "using a research based methodology, which is defined as the systematic and objective application of statistical and quantitative research principles to determine a standard measurement of acceptable academic progress for each school and LEA[4]." Clearly, the goal is that these designations be based on objective measures and that sound research methodologies are applied in their creation.

However, given the potential for increased expansion in the use of school designations coupled with the Center's own need to define and measure school quality, this report takes a look at the latest iteration of school designations, the A-F Letter Grades, and applies basic statistical analysis to better understand what the designations really tell us about schools, and what conclusions we safely draw from a school's designation.

We have used publicly available data from the 2012 A-F Letter Grades for schools to answer the following research questions:

1. Are the A-F Letter Grades measuring a school's contribution to learning or simply repackaging other indicators, such as poverty, and presenting them as measures of school quality?
2. What existing data and research based methods have the potential to improve the A-F Letter Grades as measures of school quality?

ANALYSIS

In an effort to create a school accountability system that is fair to all public schools and accounts for the fundamental differences in school types and student populations attending different types of schools, the Arizona Department of Education created four discrete accountability models: Traditional, Alternative, Small, and K-2 (see Definition of Terms on page 23 for a full explanation).

TABLE 1: DISTRIBUTION OF STUDENT ENROLLMENT BY SCHOOL TYPE AND GRADE CONFIGURATION, 2012

	ENROLLMENT	% OF ENROLLMENT	NUMBER OF SCHOOLS
TRADITIONAL			
ELEMENTARY	672,493	65.7%	1,255
HIGH SCHOOL	248,858	24.3%	218
COMBINATION	60,698	5.9%	99
SMALL			
ELEMENTARY	6,081	0.6%	118
HIGH SCHOOL	2,952	0.3%	71
COMBINATION	1,460	0.1%	31
ALTERNATIVE			
ELEMENTARY	996	0.1%	11
HIGH SCHOOL	17,932	1.8%	124
COMBINATION	9,110	0.9%	30
K2			
ELEMENTARY	3,326	0.3%	11
STATE	**1,023,905**	**100%**	**1,968**

Although there are ten discrete categories of schools, the vast majority of schools are Traditional schools. Altogether, they enroll 95 percent of all students in the state. For the purpose of discussion, this paper will focus on Traditional elementary schools (65 percent of state enrollment), but the Appendices lists results for all Traditional, Small, and Alternative school types. We did not include K-2 schools in our analysis due to their extremely small size in both number of schools and students.

ARE THE A-F LETTER GRADES MEASURING A SCHOOL'S CONTRIBUTION TO LEARNING OR SIMPLY REPACKAGING OTHER INDICATORS, SUCH AS POVERTY, AND PRESENTING THEM AS MEASURES OF SCHOOL QUALITY?

Unless otherwise mitigated, poverty can have a significantly deleterious effect on learning. Students growing up in poverty often arrive at school with challenges that adversely affect their ability to learn or to learn at the same rate as their wealthier counterparts. This relationship between poverty and learning has been widely known (and acknowledged) since the 1960's and has resulted in the federal Title I program providing additional resources to schools that serve a significant percentage of students in poverty. Several states across the country (not including Arizona) also include "at-risk" or "poverty" weights into their school funding systems[5]. These additional funds, triggered by the number of poor students, are an explicit acknowledgement that schools serving students in poverty have a greater challenge in educating students to levels that are comparable to those of wealthier students. In public education, poverty is typically measured using participation in the National School Lunch Program, which provides free or reduced-price lunches (FRL) to students from low-income families.
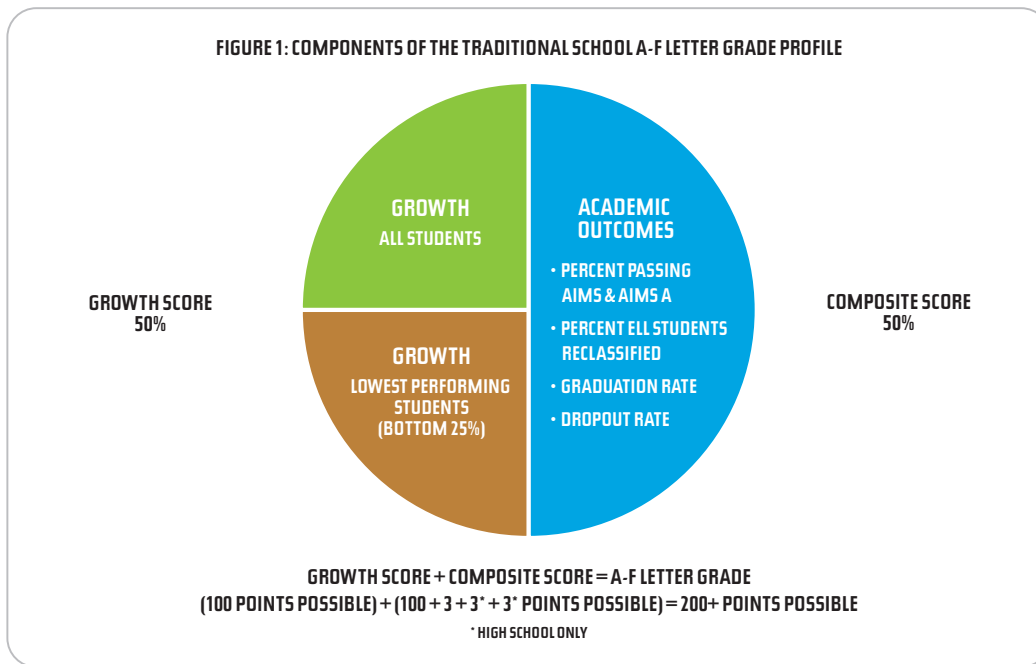
School funding is not the only place where this relationship is acknowledged. The relationship between poverty and measures of achievement (e.g., percent of students passing AIMS) has long been a limitation of measurement of student achievement in education. The contribution that a school itself makes to student learning is difficult to determine, when we know from the start that the wealthier the students in a school are, the higher that school's scores will be, simply as a function of the student population. In other words, all else being equal, we expect wealthier students to score higher on standards-based proficiency measures than low-income students[6]. This is not to say that schools do not make a difference; they do, but it is often difficult to measure their effect.

This is, in part, the reason why so much attention has been given to measures of student "growth," such as Value-Added Models (VAM) and Student Growth Percentiles (SGP)—a relatively new advancement in educational measurement made possible by data systems that allow student-level data to be linked across multiple years. VAM and SGP analyses allow the researcher to "control for," or set aside statistically, the relationship between poverty (and potentially other variables) and measures of achievement, in order to capture the effect that a school or teacher has had on an individual student's learning.

All of the A-F Letter Grade models utilize a combination of student achievement (percent passing AIMS) and student growth (most often using SGP) data for schools, though the exact calculations vary by model[7]. The Traditional model utilizes a 200-point scale where 100 points can be awarded for achievement (Composite Points) and 100 points can be awarded for student growth (Growth Points). Composite and Growth Points are added to arrive at the Total Points and the corresponding letter grade for each school.

The 50/50 distribution of points between achievement and growth communicates to educators that both of these factors are equally important. Students must show growth, but they must also demonstrate proficiency on academic standards. Figure 1 is taken from page 13 the *A-F Letter Grade Accountability System Technical Manual* published by Arizona Department of Education and shows the components of the A-F Letter Grades for Traditional schools. The composition is the same for Small schools, with the exception that three years of pooled data are used in the calculations.

FIGURE 1: COMPONENTS OF THE TRADITIONAL SCHOOL A-F LETTER GRADE PROFILE

GROWTH
ALL STUDENTS

ACADEMIC OUTCOMES

· PERCENT PASSING AIMS & AIMS A
· PERCENT ELL STUDENTS RECLASSIFIED
· GRADUATION RATE
· DROPOUT RATE

GROWTH SCORE 50%

GROWTH
LOWEST PERFORMING STUDENTS (BOTTOM 25%)

COMPOSITE SCORE 50%

GROWTH SCORE + COMPOSITE SCORE = A-F LETTER GRADE
(100 POINTS POSSIBLE) + (100 + 3 + 3* + 3* POINTS POSSIBLE) = 200+ POINTS POSSIBLE
* HIGH SCHOOL ONLY

The inclusion of growth in the accountability formula is meant to give credit to schools for their effect on student learning, and to address the limitations that we know exist when using only proficiency results (the right side of the pie chart) to measure school quality. In other words, when combined with proficiency rates, growth is meant to erase, or mitigate, the negative relationship that exists between poverty and measures of student achievement.
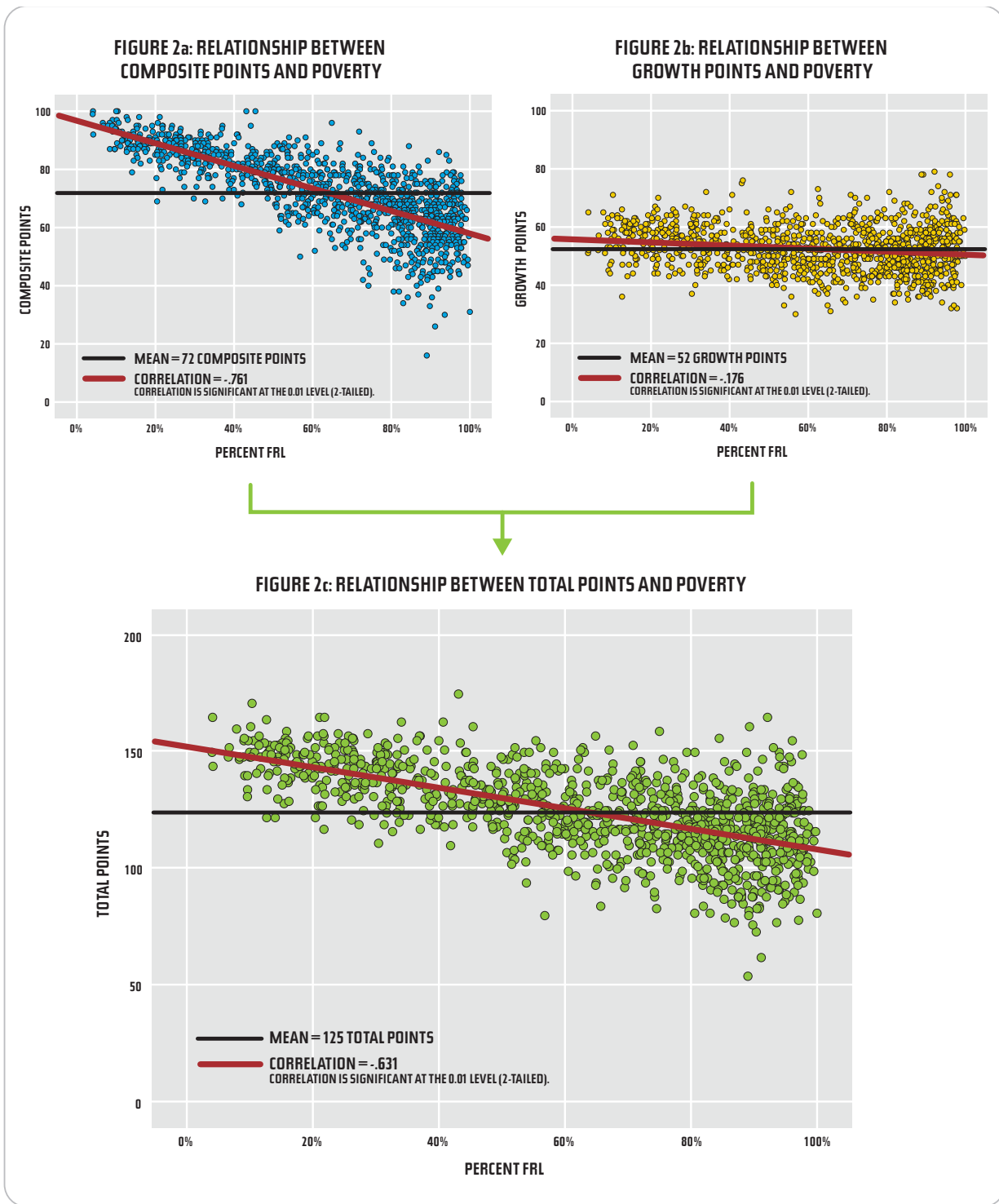
In order to determine whether this is actually the case, we performed a set of correlations to determine the relationship between poverty, using the percent of students eligible for the National School Lunch Program (FRL), and the different types of points (Composite, Growth, and Total) resulting from the accountability model.

CORRELATION

A correlation represents the direction and strength of a relationship between two variables.

· The direction is expressed as either positive or negative. For example, a negative correlation between poverty and student achievement means that as poverty increases, achievement decreases.

· The strength of the relationship is denoted by a number between 0 (no correlation) and 1 (perfect correlation). As a rule of thumb, a correlation between 0 and .39 is considered weak to moderate. A correlation of .40 to .69 is considered strong, and a correlation of .70 or above is considered to be very strong.

· A statistically significant finding is one that is very unlikely to happen by chance; in other words, we can be confident in the results. Statistically significant findings are indicated by a * or **, depending on the level of significance.

Figure 2 shows the result of this analysis for Traditional elementary schools (see Appendices A.1-A.3 for the correlation statistics for each Traditional, Small, and Alternative school type). The scatter plots graphically represent the relationship between poverty and Composite Points, Growth Points, and Total Points. In each case, points are graphed along the Y (vertical) axis and Percent FRL is graphed along the X (horizontal) axis. In this discussion we use "Composite Points" and "percent passing AIMS" interchangeably because virtually all (100 out of 103) Composite Points available to elementary schools are based on the percent of students passing AIMS.

FIGURE 2a: RELATIONSHIP BETWEEN COMPOSITE POINTS AND POVERTY

COMPOSITE POINTS

PERCENT FRL

MEAN = 72 COMPOSITE POINTS
CORRELATION = -.761
CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED).

FIGURE 2b: RELATIONSHIP BETWEEN GROWTH POINTS AND POVERTY

GROWTH POINTS

PERCENT FRL

MEAN = 52 GROWTH POINTS
CORRELATION = -.176
CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED).

FIGURE 2c: RELATIONSHIP BETWEEN TOTAL POINTS AND POVERTY

TOTAL POINTS

PERCENT FRL

MEAN = 125 TOTAL POINTS
CORRELATION = -.631
CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED).

The first scatterplot confirms the negative relationship we know exists—as poverty increases, Composite Points (or percent passing AIMS) decreases. The red line is the regression line, indicating the direction and strength of the relationship. This is a very strong statistical relationship at -.761.

The second scatterplot, however, shows a much different relationship. While there is still a negative relationship between poverty and Growth Points (as poverty increases, growth points decrease), this relationship is weak (-.176), indicating that there is no meaningful difference in the distribution of Growth Points among schools with varying degrees of poverty. A poor school is as likely to demonstrate the same amount of growth as a wealthier school.

The third scatterplot shows the relationship between poverty and Total Points—the end result of combining Composite and Growth Points. While the negative relationship between poverty and Total Points (-.631) is less dramatic than the -.761 of poverty to Composite Points, it is still strong and statistically meaningful. The bottom line: despite the attempt to "level the playing field" through the use of growth, schools with greater degrees of poverty are less likely to do well using the current school accountability system.

Our finding here, then, is a combination of good news and bad news. The second scatterplot is good news; unlike Composite Points, Growth Points cannot be accurately predicted, simply by the level of poverty in the school. This finding indicates that growth appears to be a fairer method of determining the contribution of schools to student learning than just percent passing. The bad news, then, is the fact that the simple inclusion of growth in the accountability model does not meaningfully alter the negative relationship between the level of poverty in the school and the final school rating (Total Points).

Why, if Composite and Growth Points have equal weight in the accountability model and growth is meant to mitigate the negative relationship, isn't the relationship between a school's Total Points and their poverty rate now significantly weaker? The answer is in the effective (actual) versus nominal (theoretical) distribution of points. The horizontal black lines in the scatterplots indicate the mean, or average, number of points the group of schools earned. On average, Traditional elementary schools earned 72 Composite Points, 52 Growth Points, and 125 Total Points.

The difference between the average number of Composite and Growth Points is meaningful because it illustrates that schools are getting most of their points from the Composite side of the equation—the one with the strongest relationship to poverty.

Table 2 shows some descriptive statistics for Traditional schools (see Appendices B.1-B.3 for descriptive statistics of all Small and Alternative school types).

TABLE 2: DESCRIPTIVE STATISTICS OF TRADITIONAL MODEL SCHOOLS

	NUMBER OF SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
		MINIMUM	MAXIMUM		
ELEMENTARY					
GROWTH POINTS	1,249	24	79	52	8.011
COMPOSITE POINTS	1,249	16	101	72	13.447
TOTAL POINTS	1,249	54	175	125	18.493
ENROLLMENT	1,254	79	1,657	536	248.230
PERCENT FRL	1,156	4%	100%	64%	0.261
PERCENT MINORITY	1,151	6%	99%	56%	0.268
HIGH SCHOOL					
GROWTH POINTS	215	5	85	52	10.219
COMPOSITE POINTS	215	24	106	70	16.161
TOTAL POINTS	215	33	191	122	23.184
ENROLLMENT	218	79	3,469	1,142	869.570
PERCENT FRL	188	4%	95%	54%	0.225
PERCENT MINORITY	200	11%	98%	53%	0.256
COMBINATION					
GROWTH POINTS	94	16	79	52	11.281
COMPOSITE POINTS	94	31	106	77	18.503
TOTAL POINTS	94	50	174	129	26.464
ENROLLMENT	99	1	4,781	613	749.393
PERCENT FRL	52	6%	99%	57%	0.270
PERCENT MINORITY	89	0%	96%	43%	0.222

The minimum and maximum values depict the range of scores found in the group of schools. For Traditional elementary schools, Growth Points range from 24 to 79, compared to a Composite Point range of 16 to 101. This indicates that the highest number of Growth Points that any elementary school earned is 79—or 21 points shy of the theoretical 100 points possible. We see a similar distribution of scores in the Traditional high schools and combination schools, which are schools that offer any grades that span across elementary and high school.

Figure 3 illustrates the distribution of points for Traditional elementary schools grouped by their level of poverty. Growth points decrease slightly as poverty increases, but all scores are relatively flat, near 50 points. This is likely due to the use of the school level median, which is used to measure growth; it should be noted that the state median is also 50. Composite Points, however, definitely trend downward as poverty increases and the differences are significant—90 points for schools with fewer than 25 percent of students in poverty to 63 points for schools with 75 percent or greater poverty (see Appendices C.1-C.3 to see similar results for each Traditional, Small, and Alternative school type).

These differences are not negligible; the Total Point difference between the schools with less than 25 percent poverty and the schools with 75 percent or more poverty is 32 Total Points (146-114). This is enough of a difference to place the former group of schools in the *A* category (*A* point range is 140-200) and the latter group in the *C* category (*C* point range is 100-119). **Given this finding, can we reliably determine that C schools are lower quality schools, or is it more appropriate to assume that they serve a greater proportion of students in poverty?**

FIGURE 3: AVERAGE POINTS BY PERCENT FRL — TRADITIONAL ELEMENTARY SCHOOLS

90 POINTS COMPOSITE	83 POINTS COMPOSITE	72 POINTS COMPOSITE	63 POINTS COMPOSITE
56 POINTS GROWTH	54 POINTS GROWTH	51 POINTS GROWTH	52 POINTS GROWTH
<25% 146 TOTAL POINTS	25% TO 49% 137 TOTAL POINTS	50% TO 74% 123 TOTAL POINTS	75%+ 114 TOTAL POINTS
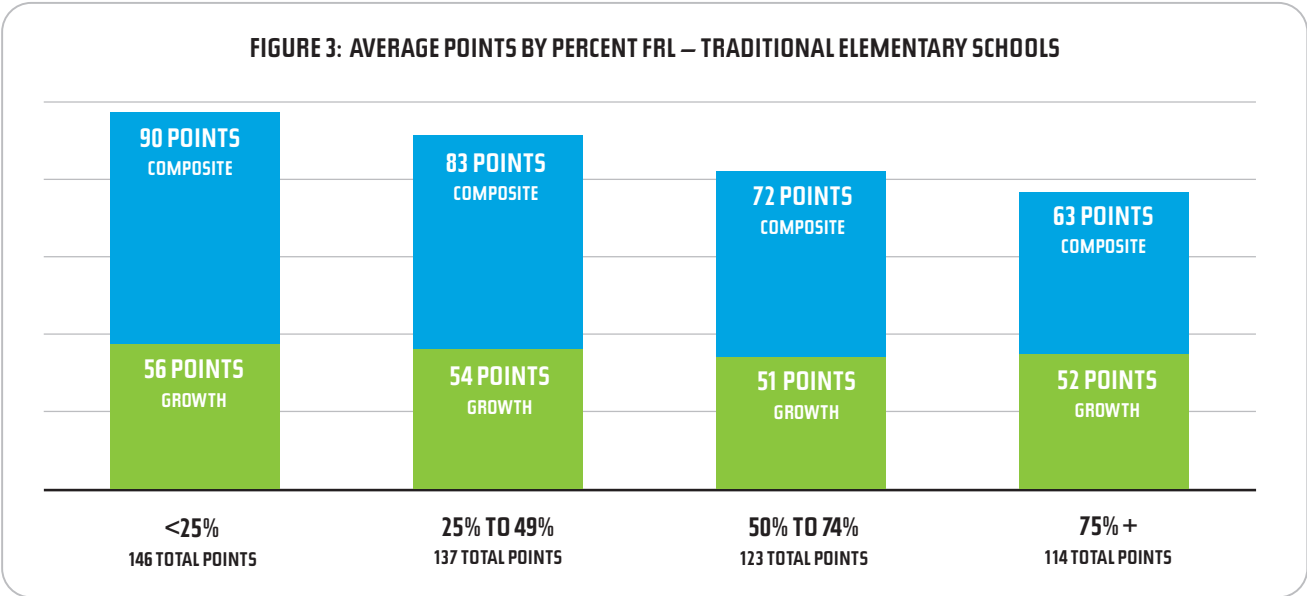
When using the actual distribution of points, we see that what began as 50/50 weight to percent passing and growth turns into something approaching a 60/40 weight, skewed toward percent passing—clearly not the intention of the accountability model. Furthermore, the point differences resulting from the accountability formula as implemented are large enough to make a difference in the final rating of the school. Despite the inclusion of growth, the practical result of the accountability formula is yet another measure of performance that is closely related to poverty. The current model fails to adequately control for the effect of poverty on indicators of achievement in order to measure a school's contribution to learning.

WHAT EXISTING DATA AND RESEARCH BASED METHODS HAVE THE POTENTIAL TO IMPROVE THE A-F LETTER GRADES AS MEASURES OF SCHOOL QUALITY?

Based on this analysis, we have developed five suggestions for changes to the A-F Letter Grades that we believe have the potential to improve them as measures of school quality. The suggestions are to maximize the utility of the student growth calculations, incorporate elements from other models into the Traditional and Small school model, and use consistent measures of school performance across all of the models.

MAXIMIZE THE UTILITY OF STUDENT GROWTH CALCULATIONS

One initial question we asked ourselves when considering these findings was whether schools are not able, based on poor performance, to access more Growth Points or is it that the "ceiling" we see in Growth Points is a function of the way in which individual student SGPs are combined to arrive at an overall Median SGP for a school. While a student-level analysis of SGPs is outside of the scope of this paper, we do know from our work with schools that individual student SGPs have a much wider range (1-99) than the ranges we see in Table 2 above. Therefore, we know growth is happening beyond the 79 SGP; however, it is not showing up in the aggregate school-level measures used in the accountability models.

Suggestion #1: Use SGPs in a way that preserves the amount of information they contain.

As Figure 1 on page 8 indicates, Growth Points are computed based on the growth of all students in a school combined with the growth of the lowest performing 25 percent of students. The seven-step calculation is performed as follows:

1. Calculate the median of all individual student SGPs in Reading, across all grade levels (Median 1)
2. Calculate the median of all individual student SGPs in Math, across all grade levels (Median 2)
3. Calculate the median of all individual student SGPs in Reading, across all grade levels, for the bottom 25 percent only (Median 3)
4. Calculate the median of all individual student SGPs in Math, across all grade levels, for the bottom 25 percent only (Median 4)
5. Calculate the average of Medians 1 and 2 (Average 1)
6. Calculate the average of Medians 3 and 4 (Average 2)
7. Calculate the average of Averages 1 and 2 (Average 3)
8. Average 3 = Growth Points

Medians and averages are quite useful as quick summaries of information—on average, how does a group perform?—but they inevitably result in a loss of detailed information. For example, it is possible that the same median SGP hides two very different distributions of individual student SGPs. What we may have encountered in the current A-F model is that seven layers of summation dilute student growth information to the point where nearly all schools are about average.

This seven-step calculation stands in stark contrast to the relatively straightforward calculation of Composite Points—the percentage of students passing AIMS. The full range of performance across all schools is maintained using that simple treatment of results. The same could be done on the Growth side of the accountability model.

Rather than using medians and averages, we propose that individual student SGPs be categorized as *low*, *typical*, or *high* growth. An individual SGP of 1-33 is considered *low*, 34-66 is *typical*, and 66-99 is *high*. The number of categories can be changed, if necessary; the point here is to categorize degrees of growth.

In this case, Growth Points can be earned based on the percentage of students who achieve high growth (or high and typical growth), much in the same way that schools earn points based on the percentage of students

who pass AIMS. Using growth in such a way gives schools credit for all of the students in their school with *high* growth and it would likely result in more variability in Growth Points (a larger range of scores), which helps distinguish the schools that are truly performing better than others.

Another potential dynamic that may be at work in the Growth side of the model is the composition of the bottom 25 percent of students. Currently, every school has a bottom 25 percent of students, but that group of students may present very different challenges for instruction from one school to the next. Growth Points are not currently reported out separately for the "all students" group and the "bottom 25 percent of students" group, so we are not able to determine if the relationship between poverty and growth looks different for those two components of the model. We think this is a worthy avenue of investigation and request that the Arizona Department of Education report these two components separately in the future.

INCORPORATE OTHER SCHOOL ACCOUNTABILITY ELEMENTS

While the discussion in the report has focused on the Traditional accountability model, the full analysis of results suggests that there is much to learn from the other school accountability models (see Appendices A.1-B.3 for the full set of descriptive statistics and correlations for all Traditional, Small, and Alternative school types). Each of the other models have components that are unique to that model, but that have the potential to provide additional information to other models and a more consistent analysis of school quality for all schools.

Keeping in mind that the number of alternative schools for which we have both school rating and poverty information is small[8], we see no relationship between the school's level of poverty and the Total Points earned. In fact, the only statistically significant correlation we find is a moderately positive relationship between poverty and Growth Points (.346, significant at the .05 level), meaning that as poverty increases, Growth Points also increase. Unlike the Traditional school model, the Growth Points for Alternative schools include a measure of improvement on AIMS—the percentage of students who improved a performance level on AIMS from one test administration to the next. Also, instead of the 50/50 split between percent passing and student growth of the Traditional and Small models, the Alternative model utilizes a 70/30 split weighted toward growth.

Suggestion #2: Incorporate the Alternative schools' additional growth measure into the Traditional and Small high school models.

We see no reason why improvement on AIMS, as defined in the Alternative school model, could not be used for all high schools and included as part of the Growth Points calculation, particularly since high schools have only one year of SGP data. The current SGP for the Traditional high school model is based on growth from the grade 9 Stanford 10 assessment to the grade 10 AIMS. There is no AIMS test beyond grade 10 (other than retesting for those students who have not passed) and, as a result, the entire growth component is based on the performance of a single grade cohort of students (those moving from grade 9 to 10). The inclusion of an additional measure of growth—the percentage of students who re-test and improve—is valuable information for all high schools, not just Alternative high schools. It is also worth exploring the potential impact of a similarly weighted formula (70/30) for all schools. Rather than the 50/50 Growth and Composite split that results in a 60/40 distribution of points, a higher weight toward Growth may result in a 50/50 practical distribution of points.

Suggestion #3: Incorporate persistence (re-enrollment) as a function of school quality for all schools.

Similarly, the use of persistence points could be extended to all schools. Persistence points are used in the Alternative school model to "reward schools for keeping their students enrolled in school[9]." A potential modification is to reward schools for re-enrollment in their school—an indicator continued of parent and student confidence in the school. The relationship of persistence to poverty needs to be explored, however, since the mobility of families will have an impact on this indicator.

USE CONSISTENT MEASURES OF SCHOOL PERFORMANCE

The first three suggestions above are intended to address specific limitations that exist in the current Traditional and Small school models, but they also highlight a dynamic that is worthy of consideration. Each accountability model was developed in order to address the specific characteristics of certain types of schools. For example, Alternative schools serve a distinct population of students and therefore require a distinct set of indicators. The informative value of the resulting indicators, however, is not limited to the unique type of school. If an indicator provides valuable information in one accountability model, it likely provides the same valuable information in the other accountability models. When dealing with a limited number of performance measures, it would be sensible to take full advantage of all available indicators.

Beyond the additional information that multiple indicators can provide, a consistent set of indicators across all school types makes the school ratings more comparable across different types of schools. Used in this way, quality is defined using the same constructs across all school types. In a policy environment where consequences are attached to schools based on school ratings, a consistent definition of quality helps to minimize the potential for under- or over-identification of a particular type of school for rewards or sanctions. This is not to say that different types of schools should be measured in the same way or that all indicators are valuable for all school types. For example, elementary schools do not have dropout and graduation rates, so it makes no sense to use them within those models. However, a consistent set of indicators can be combined in unique ways to appropriately measure the effectiveness of a particular type of school. Persistence, for example, can make up a different proportion of the Total Points, depending on the accountability model in which it is used.

Suggestion #4: Explore the potential of the On Target indicator as a measure of growth for all elementary schools.

Our analysis did not include the K-2 accountability model because only eleven schools have this designation, making the sample too small for analysis; however, we encourage the Arizona Department of Education to explore the indicator's potential for all schools. The K-2 accountability model uses the On-Target indicator as a proxy for growth. The indicator is the result of the need for a growth indicator in K-2 schools that, due to their grade configuration, lack the required two years of data to calculate growth. "The On-Target Score is a measure indicating the degree to which students in Grade 2 are on-track to meet grade-level proficiency on AIMS Reading and Mathematics in Grade 3[10]." In the context of Move on When Reading, this is a very valuable piece of information currently available only to eleven schools.

Move on When Reading legislation requires schools to retain students in third grade if they score in the Falls Far Below performance level on the third grade AIMS[11]. While the On-Target indicator should not replace SGP as the measure of growth in the accountability model, it can provide an additional measure of performance and at the same time provide valuable predictive information that can help guide intervention strategies prior to the third grade. Furthermore, it uses an existing assessment, the Stanford 10, which currently has limited use in schools.

Suggestion #5: Consider incorporating multiple years of data into all of the accountability models in order to better assess the consistent quality of schools.

The Small school accountability model uses three years of pooled data in the calculations in order to address the measurement limitations that result when too few data are available. When data are limited, for example, an extreme score (outlier) can have a disproportionate effect on the results. The benefits of using multiple years of data, however, extend beyond small schools.

Multiple years of data allow for the determination of school quality to be made on the basis of trends, rather than one-year snapshots of performance. Furthermore, the more data used in a calculation, the more reliable the calculation, and the more confidence we can have in the results.

However, an argument against using multiple years of data is that large changes in performance cannot be measured immediately. Schools that make significant improvements in one year would only see a corresponding improvement in the school rating if the improvement is sustained over multiple years. However, the inverse is also true; schools that have an anomalous bad year of performance would only see a corresponding lowering of the school rating if poor performance is maintained. In a policy context that attaches numerous and significant consequences to the school designations, the use of multiple years of data seems a prudent course of action.

CONCLUSION AND POTENTIAL POLICY IMPLICATIONS

The Center for Student Achievement undertook this analysis as the result of internal conversations about school quality. What is it? How do you measure it? Having two former Arizona Department of Education Research Directors on staff, we also have a particular vantage point from which we view the school ratings that are published each year. The truth of the matter is that there are a finite number of indicators that are available to be used and each has its own list of virtues and limitations. The key is to put the limited number of indicators together in such a way that maximizes the amount of information available to make a determination or judgment about a school's performance. This is a tough enough challenge on its own, but the current policy environment ups the ante on the school designations.

Recent policy proposals such as "parent trigger laws" and "performance funding" would add to the already-long list of consequences tied to school ratings. For these levers to be effective, rather than simply punitive, in motivating and rewarding schools to improve, they must be based on valid and reliable measures of school quality and not simply on a byproduct of the student demographics that a school serves.

What this analysis finds is that the school ratings used for the 2011-12 school year are, to a large extent, tied to the degree of poverty in a school. With the exception of Alternative schools, we find that the higher the degree of poverty, the lower the school rating. Are high poverty schools, then, simply poor quality schools? A separate analysis of Composite and Growth Points suggests that this is not the case. All schools, regardless of poverty, demonstrate an equal ability to earn Growth Points. The hitch in the system is that there appears to be a "ceiling" to the number of Growth Points schools can earn, due to the methods utilized, resulting in a greater proportion of Total Points coming from the Composite side of the equation—where we see the strongest relationship to poverty.

This report details a number of possible changes to the Traditional and Small school accountability models designed to improve them as measures of school quality. All of the suggestions come from components that already exist in other parts of the accountability system, thus requiring no new measures, but the expanded use or a re-configuring of existing data. The Alternative school model, in particular, offers some interesting opportunities for the expanded use of existing measures such as improvement on AIMS in high schools and persistence points for all schools.

The State Board of Education and the Department of Education are currently working to finalize modifications to the current accountability system. These changes, which are slated to go into effect for the 2013 evaluations, are only focused on modifying the Composite side of the formula, not the Growth side. We recommend that the Board and Department consider analyzing the effect that these changes will have on the relationship between poverty and letter grades, so that the current negative relationship is not strengthened.

ENDNOTES

1 For the full text of the ballot initiative, see the Arizona Secretary of State at http://www.azsos.gov/election/2000/General/ballotmeasures.htm and for the current accountability language, see Arizona Revised Statutes §15-241 at http://www.azleg.gov/FormatDocument.asp?inDoc=/ars/15/00241.htm&Title=15&DocType=ARS

2 For a study of school accountability and choice, see Zimmer, R., et al. (2007) *State and Local Implementation of the No Child Left Behind Act: Volume I -- Title I School Choice, Supplemental Educational Services, and Student Achievement.* RAND Corporation. http://www.rand.org/pubs/reprints/RP1265. For a study of school ratings and choice, see Klute, M. (2012) *Understanding How Parents Chose Schools: An Analysis of Denver's SchoolChoice Form Questions.* Buechner Institute for Governance. http://www.dkfoundation.org/sites/default/files/files/How%20Parents%20Choose%20Schools%20Report%20FINAL.pdf

3 Five charter schools administered by the Arizona State Board for Charter Schools have been closed since 2006 due to their "failure to meet academic standards." We do not have access to data to determine the number of district schools that have been closed due to failure to meet academic standards.

4 Arizona Revised Statutes §15-241.G

5 Verstegen, D. A. (2011) *Public education finance systems in the United States and funding policies for populations with special educational needs.* Education Policy Analysis Archives, 19 (21). http://epaa.asu.edu/ojs/article/view/769

6 While poverty creates an achievement gap in student achievement, poverty is not destiny. Students from poverty backgrounds who are well-prepared for life after high school can perform on par with their wealthier counterparts. See, for example, http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/03/27-high-achievement-college-students-hoxby

7 For a full description of all accountability models and their calculations, see the *A-F Letter Grade Accountability System Technical Manual* (2012) available on the Arizona Department of Education website. http://www.azed.gov/research-evaluation/files/2013/02/final-af_tech_manual020613.pdf.

8 Many Arizona Alternative schools do not participate in the National School Lunch program even though they typically target students at-risk of not succeeding academically, including many students living in poverty.

9 *A-F Letter Grade Accountability System Technical Manual*, page 38.

10 *A-F Letter Grade Accountability System Technical Manual*, page 28.

11 Arizona Revised Statutes §15-701.2

APPENDICES

APPENDIX A.1: CORRELATION STATISTICS OF TRADITIONAL MODEL SCHOOLS

		PERCENT FRL	GROWTH POINTS	COMPOSITE POINTS	TOTAL POINTS
ELEMENTARY					
PERCENT FRL	PEARSON CORRELATION	1	-0.176**	-0.761**	-0.631**
	SIG. (2-TAILED)		0.000	0.000	0.000
	N	1,156	1,153	1,153	1,153
TOTAL POINTS	PEARSON CORRELATION	-0.631**	0.761**	0.922**	1
	SIG. (2-TAILED)	0.000	0.000	0.000	
	N	1,153	1,249	1,249	1,249
HIGH SCHOOL					
PERCENT FRL	PEARSON CORRELATION	1	-0.151*	-0.739**	-0.600**
	SIG. (2-TAILED)		0.038	0.000	0.000
	N	188	188	188	188
TOTAL POINTS	PEARSON CORRELATION	-0.600**	0.803**	0.926**	1
	SIG. (2-TAILED)	0.000	0.000	0.000	
	N	188	215	215	215
COMBINATION					
PERCENT FRL	PEARSON CORRELATION	1	-0.144	-0.658**	-0.525**
	SIG. (2-TAILED)		0.31	0.000	0.000
	N	52	52	52	52
TOTAL POINTS	PEARSON CORRELATION	-0.525**	0.813**	0.935**	1
	SIG. (2-TAILED)	0.000	0.000	0.000	
	N	52	94	94	94

** CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED). * CORRELATION IS SIGNIFICANT AT THE 0.05 LEVEL (2-TAILED).

APPENDIX A.2: CORRELATION STATISTICS OF ALTERNATIVE MODEL SCHOOLS

		PERCENT FRL	GROWTH POINTS	COMPOSITE POINTS	TOTAL POINTS
ELEMENTARY					
PERCENT FRL	PEARSON CORRELATION	1	0.356	0.066	0.819
	SIG. (2-TAILED)		0.644	0.934	0.181
	N	5	4	4	4
TOTAL POINTS	PEARSON CORRELATION	0.819	0.857**	0.043	1
	SIG. (2-TAILED)	0.181	0.007	0.919	
	N	4	8	8	8
HIGH SCHOOL					
PERCENT FRL	PEARSON CORRELATION	1	0.346*	-0.207	0.266
	SIG. (2-TAILED)		0.042	0.232	0.122
	N	39	35	35	35
TOTAL POINTS	PEARSON CORRELATION	0.266	0.977**	0.622**	1
	SIG. (2-TAILED)	0.122	0.000	0.000	
	N	35	98	98	98
COMBINATION					
PERCENT FRL	PEARSON CORRELATION	1	0.190	0.111	0.160
	SIG. (2-TAILED)		0.449	0.661	0.527
	N	19	18	18	18
TOTAL POINTS	PEARSON CORRELATION	0.16	0.969**	0.722**	1
	SIG. (2-TAILED)	0.527	0.000	0.000	
	N	18	24	24	24

** CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED). * CORRELATION IS SIGNIFICANT AT THE 0.05 LEVEL (2-TAILED).

APPENDIX A.3: CORRELATION STATISTICS OF SMALL SCHOOL MODEL SCHOOLS

		PERCENT FRL	GROWTH POINTS	COMPOSITE POINTS	TOTAL POINTS
ELEMENTARY					
PERCENT FRL	PEARSON CORRELATION	1	-0.266	-0.447**	-0.459**
	SIG. (2-TAILED)		0.065	0.001	0.001
	N	56	49	49	49
TOTAL POINTS	PEARSON CORRELATION	-0.459**	0.798**	0.918**	1
	SIG. (2-TAILED)	0.001	0.000	0.000	
	N 49	101	101	101	
HIGH SCHOOL					
PERCENT FRL	PEARSON CORRELATION	1	0.204	-0.364	-0.093
	SIG. (2-TAILED)		0.416	0.138	0.714
	N	21	18	18	18
TOTAL POINTS	PEARSON CORRELATION	-0.093	0.843**	0.801**	1
	SIG. (2-TAILED)	0.714	0.000	0.000	
	N 18	43	43	43	
COMBINATION					
PERCENT FRL	PEARSON CORRELATION	1	0.052	-0.051	-0.004
	SIG. (2-TAILED)		0.886	0.890	0.992
	N	11	10	10	10
TOTAL POINTS	PEARSON CORRELATION	-0.004	0.887**	0.867**	1
	SIG. (2-TAILED)	0.992	0	0	
	N 10	16	16	16	

** CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL (2-TAILED). * CORRELATION IS SIGNIFICANT AT THE 0.05 LEVEL (2-TAILED).

APPENDIX B.1: DESCRIPTIVE STATISTICS OF TRADITIONAL MODEL SCHOOLS

	NUMBER OF SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
		MINIMUM	MAXIMUM		
ELEMENTARY					
GROWTH POINTS	1,249	24	79	52	8.011
COMPOSITE POINTS	1,249	16	101	72	13.447
TOTAL POINTS	1,249	54	175	125	18.493
ENROLLMENT	1,254	79	1,657	536	248.230
PERCENT FRL	1,156	4%	100%	64%	0.261
PERCENT MINORITY	1,151	6%	99%	56%	0.268
HIGH SCHOOL					
GROWTH POINTS	215	5	85	52	10.219
COMPOSITE POINTS	215	24	106	70	16.161
TOTAL POINTS	215	33	191	122	23.184
ENROLLMENT	218	79	3,469	1,142	869.570
PERCENT FRL	188	4%	95%	54%	0.225
PERCENT MINORITY	200	11%	98%	53%	0.256
COMBINATION					
GROWTH POINTS	94	16	79	52	11.281
COMPOSITE POINTS	94	31	106	77	18.503
TOTAL POINTS	94	50	174	129	26.464
ENROLLMENT	99	1	4,781	613	749.393
PERCENT FRL	52	6%	99%	57%	0.270
PERCENT MINORITY	89	0%	96%	43%	0.222

APPENDIX B.2: DESCRIPTIVE STATISTICS OF ALTERNATIVE MODEL SCHOOLS

	NUMBER OF SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
		MINIMUM	MAXIMUM		
ELEMENTARY					
GROWTH POINTS	8	63.0	127.4	91.7	20.177
COMPOSITE POINTS	8	0.0	27.6	15.2	10.440
TOTAL POINTS	8	71.0	130.0	109.6	17.832
ENROLLMENT	8	6	342	124	131.219
PERCENT FRL	5	82%	96%	88%	0.061
PERCENT MINORITY	5	43%	96%	74%	0.240
HIGH SCHOOL					
GROWTH POINTS	98	43.4	140.0	99.7	26.731
COMPOSITE POINTS	98	3.0	40.2	19.5	7.174
TOTAL POINTS	98	59.0	185.0	123.7	30.943
ENROLLMENT	120	3	1,199	149	155.654
PERCENT FRL	39	36%	99%	81%	0.145
PERCENT MINORITY	78	9%	97%	63%	0.236
COMBINATION					
GROWTH POINTS	24	51.8	130.2	87.3	24.496
COMPOSITE POINTS	24	4.8	40.8	18.3	8.015
TOTAL POINTS	24	62.0	169.0	110.1	30.123
ENROLLMENT	28	0	5,413	325	1,006.752
PERCENT FRL	19	26%	96%	71%	0.179
PERCENT MINORITY	19	30%	96%	61%	0.215

APPENDIX B.3: DESCRIPTIVE STATISTICS OF SMALL SCHOOL MODEL SCHOOLS

	NUMBER OF SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
		MINIMUM	MAXIMUM		
ELEMENTARY					
GROWTH POINTS	101	11	75	43	12.761
COMPOSITE POINTS	101	10	100	60	19.427
TOTAL POINTS	101	23	163	102	28.019
ENROLLMENT	118	2	201	52	27.695
PERCENT FRL	56	0%	100%	75%	0.193
PERCENT MINORITY	80	0%	86%	41%	0.221
HIGH SCHOOL					
GROWTH POINTS	43	17	83	49	19.903
COMPOSITE POINTS	43	10	89	50	17.907
TOTAL POINTS	43	45	155	99	31.128
ENROLLMENT	71	0	95	42	30.476
PERCENT FRL	21	43%	97%	79%	0.146
PERCENT MINORITY	30	5%	86%	44%	0.214
COMBINATION					
GROWTH POINTS	16	3	82	40	21.014
COMPOSITE POINTS	16	21	88	52	19.487
TOTAL POINTS	16	44	158	91	35.537
ENROLLMENT	31	0	108	47	34.002
PERCENT FRL	11	52%	92%	73%	0.136
PERCENT MINORITY	21	7%	84%	46%	0.237

APPENDIX C.1: AVERAGE POINTS BY PERCENT POVERTY FOR TRADITIONAL MODEL SCHOOLS

PERCENT FRL	POINTS	SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
			MINIMUM	MAXIMUM		
ELEMENTARY						
<25%	GROWTH PTS.	127	36	71	56.1	6.332
	COMPOSITE PTS.	127	69	100	89.7	5.127
	TOTAL POINTS	127	117	171	145.8	9.875
25% TO 49%	GROWTH PTS.	213	37	76	53.9	6.801
	COMPOSITE PTS.	213	68	100	83.0	6.040
	TOTAL POINTS	213	110	175	136.9	11.340
50% TO 74%	GROWTH PTS.	303	30	73	51.0	7.840
	COMPOSITE PTS.	303	40	96	72.3	9.287
	TOTAL POINTS	303	80	157	123.3	14.676
75%+	GROWTH PTS.	510	32	79	51.5	8.095
	COMPOSITE PTS.	510	16	89	62.6	10.486
	TOTAL POINTS	510	54	165	114.1	16.079
HIGH SCHOOL						
<25%	GROWTH PTS.	23	38	74	56.6	8.072
	COMPOSITE PTS.	23	80	100	90.3	5.300
	TOTAL POINTS	23	120	170	147.0	11.412
25% TO 49%	GROWTH PTS.	56	33	75	52.3	8.093
	COMPOSITE PTS.	56	57	92	75.7	8.477
	TOTAL POINTS	56	93	162	128.0	14.260
50% TO 74%	GROWTH PTS.	68	26	73	52.0	9.624
	COMPOSITE PTS.	68	28	99	66.0	11.583
	TOTAL POINTS	68	56	147	118.0	17.883
75%+	GROWTH PTS.	41	32	79	51.0	9.209
	COMPOSITE PTS.	41	24	81	56.1	11.985
	TOTAL POINTS	41	57	160	107.1	18.864
COMBINATION						
<25%	GROWTH PTS.	7	55	67	60.7	4.271
	COMPOSITE PTS.	7	86	100	92.4	4.614
	TOTAL POINTS	7	146	165	153.1	8.194
25% TO 49%	GROWTH PTS.	15	30	63	47.5	8.560
	COMPOSITE PTS.	15	60	99	77.7	11.317
	TOTAL POINTS	15	98	158	125.2	18.284
50% TO 74%	GROWTH PTS.	11	30	60	46.4	7.839
	COMPOSITE PTS.	11	41	93	67.6	14.773
	TOTAL POINTS	11	71	153	114.0	21.799
75%+	GROWTH PTS.	19	32	66	51.2	8.921
	COMPOSITE PTS.	19	38	81	60.7	12.902
	TOTAL POINTS	19	72	138	111.8	19.112

APPENDIX C.2: AVERAGE POINTS BY PERCENT POVERTY FOR ALTERNATIVE MODEL SCHOOLS

PERCENT FRL	POINTS	SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
			MINIMUM	MAXIMUM		
ELEMENTARY						
75%+	GROWTH PTS.	4	79.8	105.0	95.9	11.287
	COMPOSITE PTS.	4	4.8	27.6	17.4	9.859
	TOTAL POINTS	4	110.0	122.0	116.3	5.679
HIGH SCHOOL						
25% TO 49%	GROWTH PTS.	1	70.0	70.0	70.0	.
	COMPOSITE PTS.	1	21.6	21.6	21.6	.
	TOTAL POINTS	1	92.0	92.0	92.0	.
50% TO 74%	GROWTH PTS.	9	46.2	134.4	88.7	36.137
	COMPOSITE PTS.	9	12.6	40.2	23.9	8.071
	TOTAL POINTS	9	65.0	167.0	118.4	41.049
75%+	GROWTH PTS.	25	65.8	140.0	111.7	19.051
	COMPOSITE PTS.	25	9.0	39.0	19.5	7.104
	TOTAL POINTS	25	82.0	185.0	136.6	23.109
COMBINATION						
25% TO 49%	GROWTH PTS.	3	57.4	71.4	63.0	7.408
	COMPOSITE PTS.	3	9.6	14.4	12.2	2.425
	TOTAL POINTS	3	76.0	87.0	81.3	5.508
50% TO 74%	GROWTH PTS.	6	63.0	121.8	96.4	20.698
	COMPOSITE PTS.	6	16.8	40.8	24.3	8.906
	TOTAL POINTS	6	85.0	169.0	125.8	27.694
75%+	GROWTH PTS.	9	51.8	130.2	82.8	29.644
	COMPOSITE PTS.	9	4.8	27.0	17.1	8.638
	TOTAL POINTS	9	62.0	158.0	104.0	36.473

APPENDIX C.3: AVERAGE POINTS BY PERCENT POVERTY FOR SMALL SCHOOL MODEL SCHOOLS

PERCENT FRL	POINTS	SCHOOLS	RANGE		MEAN	STANDARD DEVIATION
			MINIMUM	MAXIMUM		
ELEMENTARY						
25% TO 49%	GROWTH PTS.	3	35	57	48.0	11.533
	COMPOSITE PTS.	3	54	82	69.3	14.189
	TOTAL POINTS	3	89	134	117.3	24.664
50% TO 74%	GROWTH PTS.	16	26	69	45.1	12.201
	COMPOSITE PTS.	16	40	97	63.2	15.592
	TOTAL POINTS	16	71	149	108.3	23.669
75% +	GROWTH PTS.	30	19	64	40.2	12.918
	COMPOSITE PTS.	30	19	85	52.1	16.154
	TOTAL POINTS	30	59	139	92.4	21.500
HIGH SCHOOL						
25% TO 49%	GROWTH PTS.	1	21	21	21.0	.
	COMPOSITE PTS.	1	50	50	50.0	.
	TOTAL POINTS	1	71	71	71.0	.
50% TO 74%	GROWTH PTS.	7	19	64	41.7	16.347
	COMPOSITE PTS.	7	41	75	58.3	12.526
	TOTAL POINTS	7	60	128	100.0	22.241
75% +	GROWTH PTS.	10	17	72	43.9	18.900
	COMPOSITE PTS.	10	10	70	41.6	17.399
	TOTAL POINTS	10	45	121	85.5	30.318
COMBINATION						
50% TO 74%	GROWTH PTS.	5	23	52	38.6	11.803
	COMPOSITE PTS.	5	21	88	50.2	26.205
	TOTAL POINTS	5	51	140	88.8	37.131
75% +	GROWTH PTS.	5	21	82	37.0	25.525
	COMPOSITE PTS.	5	22	76	44.4	19.983
	TOTAL POINTS	5	44	158	81.4	44.506

DEFINITION OF TERMS: SCHOOL ACCOUNTABILITY MODELS AND GRADE LEVEL CONFIGURATIONS

SCHOOL ACCOUNTABILITY MODELS

The state's A-F Letter Grades are given to all public schools according to one of four school accountability models. The models are designed to provide a fair assessment of different types of schools that often enroll different groups of students. The four accountability models are:

Traditional School: A school that is not considered to be an Alternative, Small, or K-2 school. Traditional schools offer standard curricular options in a typical school setting. Most Arizona public schools are Traditional schools.

Alternative School: Alternative schools are held to a different accountability model due to the unique composition of their schools and the students they serve. Alternative status is granted to schools that apply to the Arizona Department of Education. Alternative high schools must award degree-granting credit and grant high school diplomas. In order to qualify as an alternative school, a school must serve one or more of the following student populations:

- Students with behavioral issues (documented history of disruptive behavior)
- Students identified as dropouts
- Students in poor academic standing who are either severely behind on academic credits (more than one year) or have a demonstrated pattern of failing grades
- Pregnant and/or parenting students
- Adjudicated youth

Small School: A school serving 100 or fewer students. A separate accountability model was defined for this group of schools due to the small number of students and the disproportionate impact that one or two students can have on the overall rating of a school.

K-2 School: A school serving only Kindergarten through grade 2. A separate accountability model was defined for this group of schools because of the limited amount of data available (i.e., AIMS state assessment is given in grades 3-8 and 10).

For a full description and additional information about the A-F Letter Grades, please visit the Arizona Department of Education at http://www.azed.gov/research-evaluation/a-f-accountability/.

GRADE LEVEL CONFIGURATIONS

Anyone who has been inside of an elementary school and a high school can see the significant differences in how the schools operate, their scheduling, curricular options, etc. In order to take these differences into account, this report uses the following three school configuration categories:

Elementary School: A school that offers any grades Kindergarten to 8th grade.

High School: A school that offers any grades 9-12.

Combination School: A school that offers any grades that span across the elementary and high school grades. For example, a school that offers instruction in grades 6-12 is a combination school.

For the purpose of analysis, each school is assigned to a single accountability model/grade configuration category. For example, a school that serves adjudicated students in grades 6-12 is considered an "Alternative combination school." Similarly, a school that serves students in grades K-5 in a traditional school environment is considered a "Traditional elementary school" in this analysis.

About the Center for Student Achievement

All students deserve a quality education, no matter where they attend school. Founded on that principle, the Center for Student Achievement was recently launched to support school leaders and teachers. The mission of the Center is to improve student achievement in all schools focusing our efforts on high-quality school improvement initiatives, with a particular focus on professional development and the publication of rigorous and transparent research and evaluation.

From this foundation, the Center developed a suite of supporting programs for school leaders and teachers, currently in use by a growing number of district and charter schools in Arizona. Nationally, the Center also provides consulting services to educators in other states about how to make student data analysis applicable and actionable. The Center has invested significantly in professional talent, hiring a team of experienced trainers and researchers who are capable of delivering customized, high-quality, practical and actionable data and professional development for teachers and leaders. The Center's goal is that these partnerships will expand the number of high quality public school choices available to all students.