

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

Purpose

A significant amount of public trust has been placed on school A-F ratings, in part, due to the fairness that is implied by the inclusion of growth scores, which are intended to mitigate for the impact of poverty and other student demographics. The use of growth is expected to credit schools, in particular those that serve at risk populations but measuring their “value-add” for their students. This study examines the impact of the use of median student growth percentiles and alternative methods of measuring growth in a high –stakes accountability model.

Perspective

With at least a dozen states utilizing accountability models that assign A-F Letter Grades to schools little research has been conducted about the validity of these designations and their methods for evaluating schools. A southwestern state, used in this study, adopted the A-F accountability system in 2011, modeled after Florida, as the latest iteration of a school accountability law initially passed in 2000 which 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 (X.R.S. § 15-241). Regardless of the state, it is clear that even though school accountability ratings may evolve over time their purpose remains 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 *F* schools are failing their students altogether.

While the general message may be simple, the subsequent uses of these school labels are not. In this southwestern state, school accountability designations are more than simply informational; they are an explicit tool for school reform with the school designation used to trigger the school improvement process, despite little research evidence that this improves student outcomes (Hanushek, E.A. & Raymond, M.E, 2005). School improvement policies have been developed based 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. Given this, it was necessary and appropriate to determine whether the state’s A-F rating system truly provided valuable information to stakeholders about school quality.

The relationship between poverty and measures of achievement (e.g., percent of students passing the state’s assessment) has long been a limitation of measurement of student achievement in education (Kane, T.J. & Staiger, D.O., 2002). 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.

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

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, we expect wealthier students to score higher on standards-based proficiency measures than low-income students. 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 (Betebenner, 2008). 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 measure the effect that a school or teacher has had on an individual student’s learning.

The state’s A-F Letter Grade models utilize a combination of student achievement (percent passing) and student growth (most often using SGP) data for schools, though the exact calculations vary by model (AZDOE, 2012). 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. 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 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.

Based on the findings reported in *Oh, the Places They’ll Go! School Ratings: Improving the data in data driven decision making*, a review of 2012 A-F results indicate:

“Student Growth Percentiles have the potential to be a much fairer method of determining the contribution of schools to student learning than percent passing on AIMS, the state’s standardized test. However, the current use 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. 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 model fails to control adequately 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” (pg. 2) (Aportela, A. & Laczko-Kerr, I., 2013).

Researchers have demonstrated that the current A-F model used in one southwestern state, which is similar to many other states’ models, fails to adequately control for the effect of

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

poverty on indicators of achievement in order to measure the school’s contribution to learning. Analyses previously conducted confirmed the negative relationship known to exist in student proficiency rates- as poverty increases, percent passing decreases. Additionally, researchers have confirmed that there is a weaker relationship between student growth and poverty- a poor school is as likely to demonstrate the same amount of growth as a more wealthy school. It has been observed that the effective weight based on the A-F accountability models is closer to 60/40 weight 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 (Aportela, A. & Laczko-Kerr, I., 2013).

This occurs not because of a limitation of the SGP model, but rather because of the methods utilized by the AZDOE, specifically the calculation of a schools’ median SGP. According to the AZDOE’s technical manual, a seven step calculation is used to compute the total growth points which are based on the calculation of four medians which are then averaged three times (AZDOE, 2012, pg. 23-27). Given the consequences associated with the A-F accountability results and the explicit policy expectations that these formula are to balance growth with proficiency, these results are alarming.

The researcher intends to develop an alternative method for the calculation of a growth score for the purpose of state A-F accountability determinations. This alternative method continues to use the SGP metric, as it is currently calculated at the student level, in a way that fully represents the range of growth at the school level by giving credit for schools that grow students across the full range of the growth percentile (1-99). These values will then be transformed into points that can be utilized in the overall formula.

Methods

The researcher used masked student level SGP data files from the 2013 academic year obtained from the AZDOE to model an alternative growth point calculation. These data were merged with the 2013 A-F accountability results: overall grade, points earned for growth and composite (a measure of proficiency); the A-F model (alternative, small or traditional), and the type of school (district, district sponsored charter or charter). These descriptive results are presented in Chart 1 below.

	Alternative	Small	Traditional	Grand Total
Charter school	114	18	371	503
District sponsored charter	3		62	65
District school	46	16	1273	1335
Grand Total	163	34	1706	1903

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

Rather than using medians and averages, the researcher created alternative growth points utilizing the percentage of students’ SGPs categorized based on the categorical determinations created by the R package that is commonly used to run student growth percentiles, see Chart 1.

VL= Very Low (SGP 1-19)
L= Low (SGP 20-39)
T= Typical (SGP 40-60)
H= High (SGP 61-80)
VH= Very High (SGP 81-90)

Chart 1. SGP growth categories

Alternative Growth Point calculation

The researcher next calculated the percentage of students at the school level, using all grades and content areas (reading and math), in each of the categorical growth bands. Given that the SGP calculation will rank order students within a cohort, a concern might be that if 20% of students have to be in the “very low” or the “very high” category, does this mean that all schools’ performance will mirror the state. In order to evaluate this, an analysis by school was conducted to determine whether schools’ SGP vary across growth category.

This calculation was conducted for “All students” as well as for the “Bottom 25%” student group as these are the student groups prescribed in statute for the A-F accountability model. This resulted in two sets of growth descriptors, in each of the categories, for each school in the state (N=1903).

In order to transform these data into growth points that can be used in the full A-F model, the researcher created several multipliers or weights for each growth band to determine points for the “All students” group and the “Bottom 25%” student group separately. Several weights were created to test the impact on this policy recommendation on the outcome of schools’ growth points (see weighting section below for a full description of this method). The formula created for this model is presented below:

$$\text{All students' growth points} = \%VL*a + \%L*b + \%T*c + \%H*d + \%VH*e$$

$$\text{Bottom 25\% students' growth points} = \%VL*a + \%L*b + \%T*c + \%H*d + \%VH*e$$

Total growth points are created by averaging the “All students” growth points and the “Bottom 25% students” growth points using the following formula:

$$\text{Total Growth Points} = \text{“All students”} + \text{“Bottom 25\% students”} / 2$$

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

Weighting

The determination of the weighting used in this model represents a significant policy decision by those entities responsible for approving A-F accountability models. There are many different weights that could be used to determine the multiplier for the points attributed to schools. The researcher will present two options; these are not exhaustive, with policy implications for consideration. These particular weights are designed to award credit for all students in a school, making growth across the spectrum of growth.

Option 1:

The following option, see Chart 2, represents a policy decision that provides no weight (or points earned) for the percentage of students in the “very low” category, SGP 1-19. This would be an incentive to have students in the higher growth categories. Additionally, this option would allow a school to earn 100% of their growth points if 100% of their school were making “typical growth”, SGP 40-60. This option is comparable to the use of state accountability assessments for percent passing, in that a school can earn 100% of their points if their students earn the average passing cut score. Lastly, this option provides the **most incentives** for the “high” and “very high” growth categories, given the significant amount of weight assigned to those groups. This is intentional as all schools have the potential for students to perform at the high range. In effect, the relative weight of 2.0 for “very high” is four times as much as the .5 for “low”. Because you can earn more than 100 points in this example, you can essentially "make" up for 4 low students with one very high student. The Board should consider capping the points earned at 100.

SGP Categories	Weight
VL= Very Low (SGP 1-19)	0.0
L= Low (SGP 20-39)	0.5
T= Typical (SGP 40-60)	1.0
H= High (SGP 61-80)	1.5
VH= Very High (SGP 81-90)	2.0

Chart 2. Option 1, SGP weight

Option 2:

This option, see Chart 3, also represents a policy decision that provides no weight (or points earned) for the percentage of students in the “very low” category, SGP 1-19. Again, this would be an incentive to have students in the higher growth categories. This option would allow a school to earn 100% of their growth points if 100% of their school were making “very high”, SGP 81-90. This option provides the most incentives for the “high” and “very high” growth categories given the significant amount of weight assigned to those groups. This would represent a high expectation for schools to growth students significantly better than other schools. Given

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

the weights used, they limit the likelihood that schools will earn more than 100 points. The use of this option would not likely result in a school’s ability to earn more than 100 points; therefore, it is unnecessary for the Board to consider capping the points earned at 100.

SGP Categories	Weight
VL= Very Low (SGP 1-19)	0.0
L= Low (SGP 20-39)	.25
T= Typical (SGP 40-60)	.50
H= High (SGP 61-80)	.75
VH= Very High (SGP 81-90)	1.0

Chart 3. Option 2, SGP weight

Results

The following data are presented to determine the validity of the alternative SGP calculation for A-F letter grades. These data were calculated based on all students in the state, enrolled in all schools that received A-F letter grades in 2013 (N=1903). Given the three A-F models, this paper focuses on the analyses of school evaluated using the traditional accountability model (N=1706), see Chart 4. In this analysis, it is important to note that we should expect roughly 20% of students in each category given the nature of percentiles at the state level.

State Distribution	Mean Math	SD Math	Mean Reading	SD Reading
SGP Very Low= 1-19	20%	10%	20%	7%
SGP Low= 20-39	20%	6%	20%	5%
SGP Typical= 40-60	21%	5%	21%	4%
SGP High= 61-80	20%	5%	20%	5%
SGP Very High= 81-99	19%	9%	19%	7%

Chart 4. State distributions, reading and math for traditional A-F model

An analysis by school was conducted to determine whether schools’ SGP vary across growth category. The findings indicate that there is variation by school in the percentage of students within each growth category (see Chart 5). The range in the categories is significant in some schools. Therefore, while it is true that there will always be 20% of students statewide in each of the growth categories, there is no reason to assume that they are all clustered in particular schools. It is possible to both provide incentives for schools to improve growth as well as reward schools for making significant growth with their students.

Math	Min. Percentage	N Count	Max. Percentage	N Count
Very Low= 1=19	0	17	100	5
Low= 20-39	0	18	100	1

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

Typical= 40-60	0	19	45	1
High= 61-80	0	26	50	1
Very High= 80-99	0	33	100	4
Reading	Min. Percentage	N Count	Max. Percentage	N Count
Very Low= 1=19	0	17	100%	1
Low= 20-39	0	17	100%	2
Typical= 40-60	0	11	67%	1
High= 61-80	0	21	100%	1
Very High= 80-99	0	19	100%	1

Chart 5. School distributions, reading and math

Calculating Alternative SGP Points

The weight chosen by the policy makers in a state will have a significant impact on the growth points earned. Both weighting methodologies were analyzed to determine the impact of the potential policy implications. Option 1 (see Chart 2) weighting is presented below to demonstrate the possible impact of this change. This analysis was conducted by content area, for all grades for illustration purposes. However, the methodology recommended would combine the content areas and grade level data prior to the calculation of the categorical percentages. Again, these data are limited to only the calculation for all students, as bottom 25% data indicators were not available at the time of this analysis. Chart 6 presents the results for a selected number of charter and district schools, for the purposes of illustrating the variance between methodologies.

Reading SGP Alternative Calculation	VL	VL Points	L	L Points	T	T Points	H	H Points	VH	VH Points	Total	2013 Growth	Diff.
Challenge Charter School	11%	0	16%	0.082	22%	0.2245	28%	0.4184	22%	0.44898	117	60	57
Kingman High School	27%	0	15%	0.074	19%	0.1939	19%	0.2908	20%	0.392857	95	48	47
Kingman Middle School	35%	0	21%	0.103	17%	0.1744	17%	0.2536	10%	0.202847	73	35	38
Redbird Elementary School	26%	0	29%	0.144	18%	0.1779	15%	0.2295	12%	0.24911	80	42	38
Yuma High School	30%	0	26%	0.131	17%	0.1732	13%	0.1912	13%	0.267974	76	41	35
Math Alternative SGP Calculation	VL	VL Points	L	L Points	T	T Points	H	H Points	VH	VH Points	Total	2013 Growth	Diff.
Challenge Charter School	10%	0	19%	0.097	26%	0.2551	23%	0.352	22%	0.435374	114	60	54
Kingman High School	23%	0	19%	0.093	17%	0.1732	20%	0.2992	21%	0.414698	98	48	50
Kingman Middle School	39%	0	22%	0.11	18%	0.1833	13%	0.1922	8%	0.153025	64	35	29
Redbird Elementary School	33%	0	19%	0.096	23%	0.2278	14%	0.2028	12%	0.234875	76	42	34
Yuma High School	25%	0	20%	0.101	26%	0.255	18%	0.2718	11%	0.221477	85	41	44

Chart 6. Alternative SGP calculation results

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

Lastly, these data were used to estimate the final overall results for schools A-F Letter Grades, given the alternative methodology. Chart 7 presents these results for the same sample of schools presented earlier.

Recalculated A-F Results	RSGP	MSGP	Avg SGP	Composite	New Total	Actual 2013 Pts.	New A-F	Actual A-F
Challenge Charter School	117	114	116	98	213.6	158	A	A
Kingman High School	95	98	97	67	163.6	115	A	C
Kingman Middle School	73	64	69	57	125.6	92	B	D
Redbird Elementary School	80	76	78	67	145.1	109	A	C
Yuma High School	76	85	81	53	133.6	94	B	D

Chart 7. Recalculated A-F Letter Grade Results

Significance

Given the public nature of school accountability, it is clear that the A-F Letter Grade data is intended for stakeholders to use in order to make their decisions- sanctions, rewards or selection. Therefore, this finding is significant given the implications, positive and negative, that these letter grades have on schools. As the results indicate, these letter grades do not accurately represent the performance of schools, in particularly those schools that serve a high population of students in poverty. However, no school- rich or poor- is receiving the appropriate allocation of points associated with student growth. This clearly violates the intent of the legislature and the State Board of Education’s desires for a fair and balanced accountability system.

Additionally, given that over a dozen states utilize similar A-F letter grading systems that use some measure of student growth to “accurately and fairly” evaluate schools this study serves as a critical lens into potential limitations in their own models. For those states considering A-F systems, this study serves as a warning during the development phase to ensure that better methodological decisions are made to evaluate schools more effectively. Lastly, these results evidence that school accountability results are tied directly to methodological decisions made by agencies; therefore, it is critical that agencies model multiple methods to determine the most appropriate models.

Beyond the use of median SGPs to measure growth:
Using categorical analysis to determine the real “value add” of a school for all students

References

- Aportela, A., & Laczko-Kerr, I. (2013). Oh, the places they'll go! School Ratings: Improving the data in data driven decision making. Retrieved from https://azcharters.org/ckeditor_assets/attachments/1161/oh_the_places_they_ll_go_data_driven_decisions_final.pdf?1367380808
- AZ Department of Education (AZDOE). (2012). *A-F Letter Grade Accountability System Technical Manual* [Everything else pertaining to this citation has been removed to ensure blind peer review].
- Betebenner, D. W. (2008). *Norm- and criterion-referenced student growth*. Report of the National Center for the Improvement of Educational Assessment. Retrieved from Colorado Department of Education website: http://www.cde.state.co.us/cdedocs/Research/PDF/betebenner_norm_crit_measuresofgrowth.pdf
- Hanushek, E.A. & Raymond, M.E. (2005). *Does school accountability lead to improved student performance?* *Journal of Policy Analysis and Management*, 24:2, pp. 297-327. doi: 10.1002/pam.20091
- Kane, T. J. & Staiger, D.O. (2002). *The promises and pit-falls of using imprecise school accountability measures*. *Journal of Economic Perspectives*, 16:4, pp. 91-114. doi: 10.1257/089533002320950993