

## A – F Model Review

### Accountability Advisory Group

The Accountability Advisory Group examined each of the proposed models to determine the impact on schools. We were provided with data from ADE by school. We used the point totals calculated by ADE for all the analyses.

For each model we described pros and cons, the characteristics of schools broken out by total points (schools are grouped into fifths – top 20%, next 20%, etc.), a graphic representation of the relationship between Free and Reduced Lunch (FRL) and total points and growth points. We then recommended several models as best meeting the committee’s criteria.

The models all used the same weighted proficiency and ELL points. The following is what varied across models.

- K-8 Models
  - Growth: we varied the weighting of SGP/SGT from 0-1 to 0-2/0-4. We also looked at the original Florida proposal and the Florida scale score bands proposal (dependent on getting the school level data).
  - Menu items: One menu item calculation used Grade 3 minimally proficient, chronic absenteeism and end of course math collapsed into one measure. The second menu calculation used the same grade 3 and absenteeism measure and then used the EOC math calculation for each grade/subject separately.
  - FAY status: The originally proposed models all used one year FAY requirements. The last two models add credit for students who are continuously enrolled for two and three years also.
- 9-12 Models
  - Growth: we varied the weighting of SGP/SGT from 0-1 to 0-2/0-4 as well as examining an ELA SGP – Algebra II proficient approach. We also looked at the original Florida proposal and the scale score bands proposal.
  - CCRI: We examined using student level data or student level data aggregated at the school level.
  - All were capped
  - FAY status: The originally proposed models all used one year FAY requirements. The last two models add credit for students who are continuously enrolled for two and three years also.

The committee set four criteria to judge schools. We examined the models with these criteria in mind.

- *Is it fair and equitable?* This was measured by examining the relationship between socio-economic status (FRL) and points. While no model will be free of a relationship between achievement and income because proficiency is highly correlated, as is SGT to a lesser extent, some models will be better than others because of the growth measure used.
- *Does it differentiate?* Do all schools or no schools get the points? Or are schools spread out across the total point values? The graphic depiction of total points and growth points highlights how much the model differentiates across schools.

- *Is it transparent?* The model must be clear in how one gets points and how those points are calculated. The model also should enable schools to anticipate how one's efforts will result in a change in points in the future.
- *Does it create the right incentives?* Schools will look at the models and do what is needed to maximize points, especially with easier to manipulate measures. We did not address this criteria as much as the others.

The AAG examined the float weight model, but since it did not notably improve the correlation with poverty and the committee had already rejected that approach, we did not present that.

## K-8 Models

The Accountability Advisory Group examined each of the six proposed models and two AAG created models to determine the impact on schools.

All models used the same weighted proficiency measure (.6, 1, 1.3). ELL proficiency and growth are also the same across models. What differed was the weight given to growth and the menu item calculation.

The committee had four criteria to judge schools. We examined the models with these criteria in mind:

- *Is it fair and equitable?* All models had some relationship between income and achievement because proficiency was correlated -0.794 with income. However, the range was from -0.427 to -0.620 so some were better than others. Model 2 had the lowest correlation with income (-.561) because the growth measure had almost no correlation to income (-.095). Model 3 had the highest correlation between growth and FRL (-.468). Models 5 (-.570) and 3 (-.584) also had moderately strong correlations between total points and FRL.

In addition, schools will feel it is fair if they have had the opportunity to instruct students. All models use current year Full Academic Year (FAY) students. Two models also give credit for the performance of multiple year FAY students.

- *Does it differentiate?* Do all schools or no schools get the points? Or are schools spread out across the total point values? The graphic depiction of total points and growth points highlights how much the model differentiates across schools. For proficiency and overall points, there is less variation among high income schools, but it does differentiate among schools. Model 3 had all schools with 40% or less Free Lunch get all or almost all the growth points, which does not distinguish across schools. Models two and five had growth points that distinguished across schools in a similar way no matter what the Free Lunch rate was, which indicates it is equitable and differentiates.
- *Is it transparent?* The model must be clear in how one gets points. The model also should enable schools to anticipate how one's efforts will result in a change in points in the future. Proficiency and ELL points are fairly transparent. Both the SGP/SGT and the Florida models are transparent in how the points are calculated but are somewhat opaque in what one needs to do to improve. SGP is a relative measures and depends on one's comparison group to calculate, so what would be needed to increase SGP is not specifically known. Scale scores band change has a more familiar measure (scale scores),

but there is no publicly known connection between scale scores and score points or items correct, so what performance would be needed to grow on this measure is not known either.

- *Does it create the right incentives?* The weighted proficiency measure incentivizes moving all students across categories instead of a binary proficient/non-proficient count. The multiple year FAY model uses the binary proficient status, but that is because we are looking at different levels of FAY. This approach focuses schools on getting the most improvement over time for their students. In the area of growth, we want to incentivize improvement for all schools, high income and low income. So, we want to minimize the use of models that all high income schools do well on without much effort and those where many poor schools cannot get many growth points.

K – 8: We recommend the consideration of the following models:

- *Model 7: Growth at 50% and more credit for stable students.* This model had a minimal relationship between poverty and growth (-0.083) and the lowest relationship between total points and poverty (-0.427). Also, it would be perceived as fairer because schools would get more points for the students that were at the school for a longer period of time. Growth was set at 50%, but it is anticipated that this could be put on a more equal footing with proficiency when the proficiency rates increase.
- *Model 2: SGP/SGT growth at 0-1 and menu items with combined math.* This had a relatively lower correlation between total points and poverty (-0.561) overall and the growth measure had no correlation with FRL (-0.095).
- We want to note that the *Florida V2 model* did provide a lower correlation to FRL overall than all models. However, AG is not supporting this version because it provided .75 points to minimally proficient students who made no movement, providing no incentive to improve. Another Florida version has been drafted and provided to ADE for consideration, but we did not receive all the data needed to analyze that model.

## 9 - 12 Models

The Accountability Advisory Group examined each of the ten proposed models and two additional AAG created models to determine the impact on schools.

All models used the same weighted proficiency measure (.6, 1, 1.3). ELL proficiency and growth are also the same across models. What differed was the weight given to growth, the way growth was defined, and the menu item calculation using student data aggregated at the school level or individual student level data.

The committee had four criteria to judge schools. We examined the models with these criteria in mind:

- *Is it fair and equitable?* All models had some relationship between income and achievement because proficiency was correlated -0.794 with income. The relationship between total points and FRL ranged from -0.385 to -0.647. These were mostly less correlated with lunch than the K-8 models. Model 6 (-0.385) had by far the lowest correlation with poverty.

In addition, schools will feel it is fair if they have had the opportunity to instruct students over time. All models use current year Full Academic Year (FAY) students. Two models also give credit for the performance of multiple year FAY students.

- *Does it differentiate?* Do all schools or no schools get the points? Or are schools spread out across the total point values? The graphic depiction of total points and growth points highlights how much each model differentiates across schools overall. The CCRI calculation at the school level results in most schools getting all the points, while the CCRI calculated at the student level results in few schools getting all the points. This may change as the CCRI measures are developed over time.
- *Is it transparent?* Proficiency and ELL points are fairly transparent. Both the SGP/SGT, SGP/Alg II change, and the Florida models are transparent in how the points are calculated but are somewhat opaque in what one needs to do to improve. SGP is a relative measure and depends on one's comparison group to calculate, so what would be needed to increase SGP is not specifically known. Scale scores band change has a more familiar measure (scale scores), but there is no publicly known connection between scale scores and score points or items correct, so what performance would be needed to grow on this measure is not known either. Algebra II change is transparent in what one must do to improve. The CCRI measure is transparent in what one must do to get points and improve. In fact, this may be the area where schools make the most efforts to increase their points because it is clear what one must do (e.g., get more students to start a FAFSA).
- *Does it create the right incentives?* The weighted proficiency measure incentivizes moving all students across categories instead of a binary proficient/non-proficient count. The multiple year FAY model uses the binary proficient status, but that is because we are looking at different levels of FAY. This approach focuses schools on getting the most improvement over time for their students. The CCRI option of starting a FAFSA could be gamed by schools if there is not a requirement that the FAFSA be completed.

9 – 12: The high school recommendation is a work in progress until we receive some additional data on the new model being considered.

- *Model 6: SGP/SGT weighted 0-2/0-4 and school level CCRI.* This approach had the lowest relationship between total points and poverty (-0.385).
- *Model 2: SGP ELA/Alg II Prof, weight B, school level CCRI.* This model had a relatively lower correlation with FRL (-0.452) overall with a low growth point – FRL correlation (-0.031).
- *Model with Growth at 50% and more credit for stable students.* This model has promise but we have not worked through all the issues with calculating it yet. Specifically, only eleventh graders have three year FAY, tenth graders could have two year FAY and ninth graders current year FAY. If this model is of interest we could work with ADE to get the appropriate data and calculate it.

## Recommended Model Details

### K - 8 Model 7: ES Stability 3 Year with 50% Growth

This model calculates proficiency and highly proficient as 1 and minimally proficient and partially proficient as 0. Then it evaluates the rate proficiency by the stability of the student: 3-year stable students' proficiency, 2-year stable students' proficiency and 1-year students' proficiency. SGP/SGT weighted on 1 and then assigned as 50% of the model equally divided by SGT and SGP. This results in the following eight for each indicator.

Component	Percentage
1 Year FAY Proficiency	5
2 Year FAY Proficiency	10
3 Year FAY Proficiency	15
Growth of SGP	25
Growth on SGT	25
ELL Proficiency	5
ELL Growth	5
Menu 2 Acceleration Readiness	10

SGP Growth				SGT Growth			
Prior Year HP	0	0.25	0.5	Prior Year HP (Stay Up)	0	0.25	
Prior Year P	0	0.375	0.625	Prior Year P (Keep Up)	0	0.5	
Prior Year PP	0	0.625	0.875	Prior Year PP (Catch Up)	0	0.75	
Prior Year MP	0	0.75	1	Prior Year MP (Catch Up)	0	1	
	0-33	34-66	67-99				
	Current Year Low Growth	Current Year Average Growth	Current Year High Growth		Current Year Did Not Meet Target	Current Year Met or Exceeded Target	

- New rigorous standards recognizes the importance of growth and it is weighted at 50%, this can be adjusted in the future as proficiency increases
- The relationship between proficiency and poverty is high as in other models at -0.7806
- The relationship between growth and poverty is low at -0.0829
- The relationship between the overall points earned and poverty for this model is -0.4268

#### Distributions:

In order to evaluate the impact of the model the schools were put into five groups and their characteristics are described below:

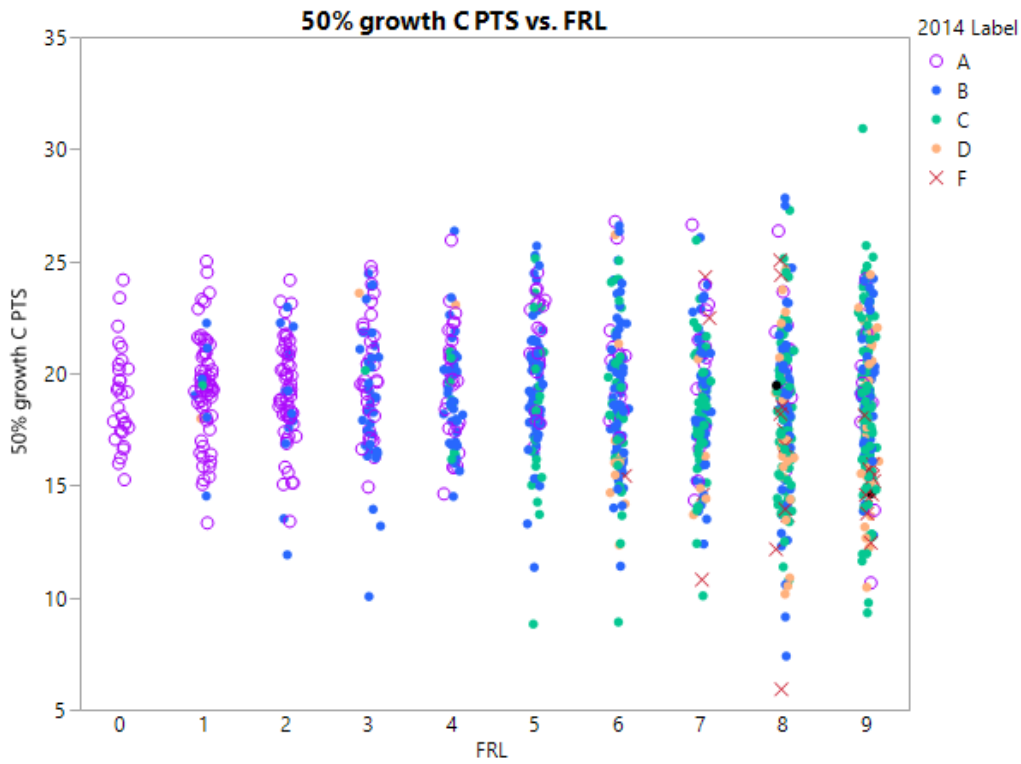
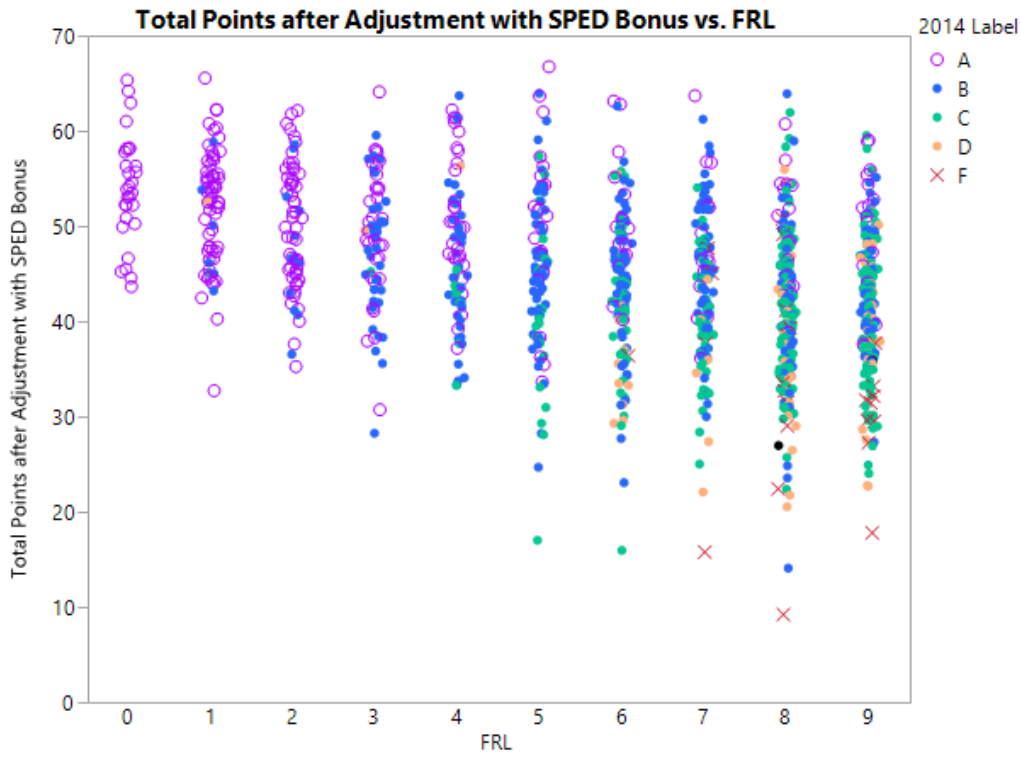
#### Total Score

Group	# FRL > 70%	# FRL < 30	# Title / Non		# Charter / Non		2014 A-F				
							A	B	C	D	F
Top 20%	9%	41%	14%	34%	25%	18%	41%	14%	6%	10%	0%
80%	16%	26%	18%	24%	21%	20%	26%	23%	13%	9%	12%
60%	20%	21%	21%	18%	15%	21%	18%	23%	23%	13%	4%
40%	25%	9%	22%	15%	18%	20%	10%	25%	26%	19%	12%
Bottom 20%	30%	3%	25%	9%	21%	20%	5%	16%	32%	49%	73%
							A	B	C	D	F

#### Correlations

	FRL
50% growth C PTS	-0.0829
Total Proficiency	-0.7806
menu2_cap	0.0109
Total Points before Adjustment	-0.2666
Total Points after Adjustment with SPED	-0.4268
Bonus	

Scatterplot – Growth v. FRL; Total Pts. – FRL, correlations



## K-8 Model 2

All models use weighted proficiency (.6, 1, 1.3) and the same calculation for ELL proficiency and growth.

Growth uses a weight of 1, each, for SGP and SGT, and growth is capped at 40 points, and Menu Items (EOC math, Gd 3 min prof, chronic absences).

Pro:

- The model has the lowest relationship between high poverty and growth (-.095).
- ELL points were available to 46% of schools. Most schools received all the growth points (78% - 5, 91% - 5 or 4). Four in ten schools got all the proficiency points with an additional 18% getting 4 points and 21% getting 3 points.

Cons:

- The average Proficiency Points was 24.7/40.
- This lower weighting of growth effectively underweights growth in the overall model since no school earned the full 40 points.
- The maximum growth points were earned by 0% of schools, 1 school obtained about 30 (30.9) and only 3% received 20 or more points.
- Growth point distributions were similar across poverty categories, but the higher income schools had a more compressed range and the lower income schools showed more variance.
- On the menu points 22% received 10, 46% received 5 and 32% received no points.
- There is a high inverse correlation between proficiency points and poverty (-.794).
- There is a moderately high inverse correlation between total points and poverty (-.561).

### Distributions:

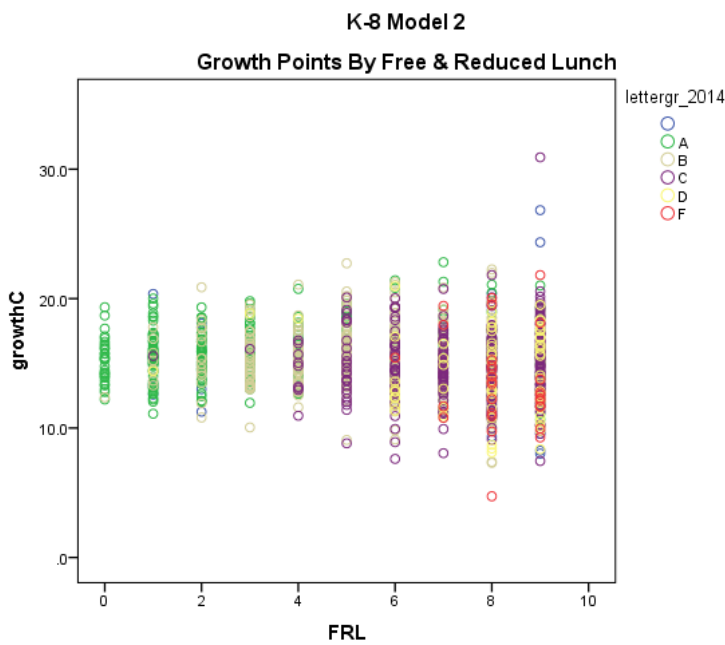
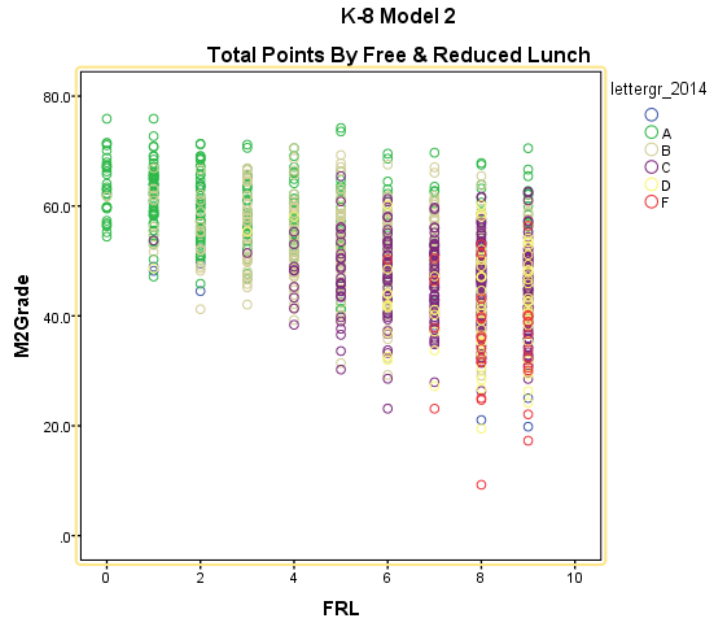
In order to evaluate the impact of the model the schools were put into five groups and their characteristics are described below:

Group	#	# FRL>70%	# FRL < 30	% Title	% Rural	% Charter	An 'A' Letter Grade in 2014
Top 20%	262	26%	37%	54%	11%	24%	63%
80%	262	41%	31%	58%	18%	22%	41%
60%	262	46%	19%	76%	20%	14%	24%
40%	262	54%	13%	77%	23%	20%	15%
Bottom 20%	262	73%	1%	88%	32%	22%	5%
# Schools	1310						

In the following graphs we have plotted the total number of points and the growth points by free and reduced lunch category (chunked in ten percent intervals). We colored each school by their 2014 letter grade so you can see where they fell last time by their total points.

This first graph demonstrates the relationship between overall points earned, the schools' level of poverty as measured by free and reduced lunch and their 2014 letter grade.

Total points is correlated with the free lunch rate (-.561), with the higher the points the lower the free lunch rate. The growth points had a low correlation with lunch rate (-.095).



This second graph shows the relationship between the growth points earned in the model and poverty. This data demonstrates that no schools, regardless of poverty are able to earn the maximum number of growth points.



## 9-12 Model 6

This model uses SGP/SGT. These are weighted on the 0-2 weighting scale for SGP and the 0-4 weighting scale for SGT. A CCRI is included and accounted for at the school level.

### Highlights:

- Weighted proficiency points ranged from 1.6 to 40 with an average of 19.89.
- The average growth points were 16.1. Approximately 20% of the schools with growth points achieved the full 20 points. And close to 65% earned at least 15 growth points.
- The correlation with the proxy we use for poverty, FRL, is at  $-.385$ .
- ELL points were available to approximately 20% of the schools (64 out of 295). Of these schools approximately half received their ELL growth points and half received their ELL proficiency points.

### Distributions:

Group	#	# FRL>70%	# FRL < 30	% Title	% Rural	% Charter	An 'A' Letter Grade in 2014
Top 20%	59	6	21	25	9	18	49
80%	59	7	7	20	16	22	28
60%	59	14	11	33	17	11	20
40%	59	19	4	39	15	12	10
Bottom 20%	59	32	3	40	20	26	6
# Schools	295	78	46	157	77	89	113

Group	#	% of FRL>70%	% of FRL < 30	% of Title	% of Rural	% of Charter	% of 'A' Letter Grade in 2014
Top 20%	59	7.69%	45.65%	15.92%	11.69%	20.22%	43.36%
80%	59	8.97%	15.22%	12.74%	20.78%	24.72%	24.78%
60%	59	17.95%	23.91%	21.02%	22.08%	12.36%	17.70%
40%	59	24.36%	8.70%	24.84%	19.48%	13.48%	8.85%
Bottom 20%	59	41.03%	6.52%	25.48%	25.97%	29.21%	5.31%

In the following graphs we have plotted the total number of points and the growth points by free and reduced lunch category (chunked in ten percent intervals). We colored each school by their 2014 letter grade so you can see where they fell last time by their total points.



## 9-12 Model 2

Model formula: SGP ELA/% Weight B prof Alg 2 and school level CCR

**Pros:**

- The relationship between high poverty and growth in this model lower than other models (.031).
- ELL points are available to 20% of schools (N=64/295); of these schools 50% of these schools earned the full ELL proficiency and full ELL growth points.
- CCRI resulted in 7 schools not having data to receive points (#NULL!) and 288 schools receiving the full 15 points.

**Cons:**

- The average proficiency points is 19.8/40. The average for growth is 8.9/20
- The maximum growth points earned by 0.7% of HS. 10% received 15 or more points; 47% received 10 or more points.
- There was a high inverse correlation between total points and lunch rate (-.452); the higher the percentage of students in poverty, the lower the total points earned by the school.
- This model does not use SGP in Math.
- Only 7% of high poverty schools earn points to put them in the top 20% of schools.

**Distributions:**

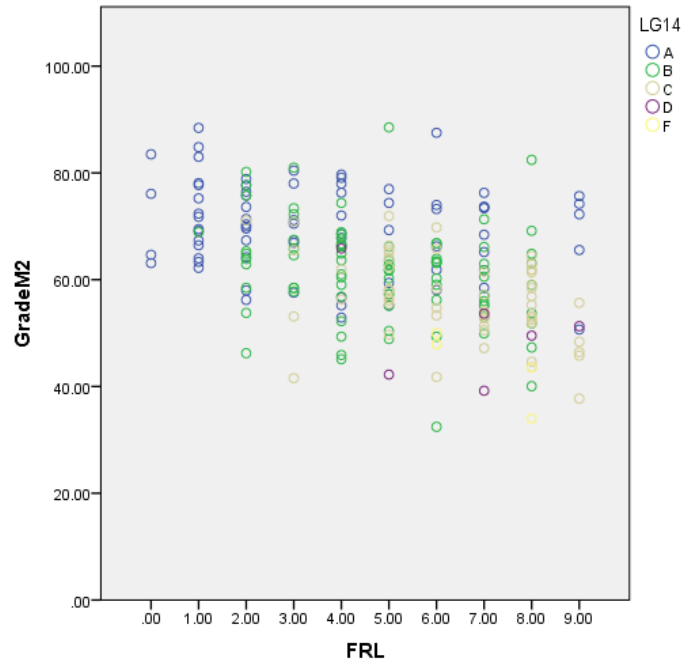
In order to evaluate the impact of the model the schools were put into five groups and their characteristics are described below:

Group	Total # of schools	% FRL>70%	% FRL < 30	% Title	%Rural	%Charter	An "A" Letter grade in 2014
Top 20%	59	7.0%	34.8%	16.6%	5.3%	28.1%	44.2%
80%	59	4.7%	28.3%	10.8%	15.8%	16.9%	29.2%
60%	59	20.9%	21.7%	20.4%	10.5%	14.6%	14.2%
40%	59	25.6%	8.7%	27.4%	21.1%	7.9%	8.8%
Bottom 20%	59	41.9%	6.5%	24.8%	47.4%	32.6%	3.5%
# of schools	295	43	46	157	19	89	113

This first graph demonstrates the relationship between overall points earned, the schools' level of poverty as measured by free and reduced lunch and their 2014 letter grade. In the following graphs we have plotted the total number of points and the growth points by free and reduced lunch category (chunked in ten percent intervals). We colored each school by their 2014 letter grade so you can see where they fell last time by their total points.

Total points are moderately correlated with the free lunch rate (-.449), with the higher the points the lower the free lunch rate. The growth points are not correlated with lunch rate, but less so than the total points (.031). Most schools with free and reduced lunch rates of less than 30% earn 80 points or greater.

### HS Model 2: Total Points by FRL



### HS Model 2: Total Growth Points by FRL

