2011–2012 Growth Model for Educator Evaluation Technical Report: FINAL

Update of the 2010-2011 Technical Report

AIR®
American Institutes for Research®

Prepared for The New York State Education Department

November 2012

Table of Contents

List of Appendices

List of Figures

List of Tables

Executive Summary

This document describes New York State's growth model for teacher and principal evaluation in grades 4 through 8, English Language Arts (ELA) and Mathematics for the 2011–2012 school year. For each student in grades 4-8, a student growth percentile (SGP) is calculated based on his or her ELA and Math State Assessment results. The calculated SGP compares each student's 2011-12 results with his/her 2010-11 results to determine how much growth has occurred. Each student's growth is then compared to the growth of students with similar academic testing history, which includes up to three years of assessment data, with adjustments made to account for measurement error. Before determining teacher or principal evaluation ratings based on the SGPs, the results are adjusted for whether a student lives in poverty, is an English Language Learner or has a disability. Students will then be attributed to teachers and schools based on linkage data. Educators are then assigned the average of the SGPs of the students they serve; this average is called the Mean Growth Percentile (MGP). The educator growth score this model yields is <u>one of the multiple measures</u> of effectiveness that result in a composite evaluation score for each educator.

The applied growth model accounts for measurement variance in the tests. Accounting for this variance removes the bias present in the conventional statistics used to estimate the effect of prior achievement on contemporary test scores, allowing for more accurate calculation of SGPs. The model also constructs and provides a 95% confidence range with each educator's MGP.

The idea of a growth model is to "level the playing field" so that educators can achieve a wide range of ratings regardless of the students they teach. Impact analysis shows that teachers MGPs have little relationship to the percentage of English language learners, and students with disabilities. Despite the model conditioning on prior year test scores, schools and teachers with students who had higher prior year test scores, on average, had higher MGPs. Teachers of classes with higher percentages of economically disadvantaged students had lower MGPs.

Results presented in this report are based on 2011–2012 and prior school years' data. The 2010-2011 "beta growth model" technical report, published in August 2012 and available here: http://usny.nysed.gov/rttt/docs/nysed-2011-beta-growth-tech-report.pdf, describes the **initial models** that **were estimated with 2010–2011 and prior school years' data** to determine the most accurate and fair way of designing a final model, with stakeholder input. 2010-2011 results were not used for evaluation purposes.

There were two notable changes between the two years:

- A growth model's quality and validity are improved when there are as many valid links between educators and their students as possible. **This year, data were provided for many more students and teachers than in the 2010-2011 beta model. Every teacher and principal had an opportunity to verify the students who were linked to them, and a link between a student and teacher was defined as 195 days (ELA) or 203 days (Mathematics) of enrollment between the student and teacher. In comparison, in the 2010-2011 school year beta model there was no minimum number of days of enrollment required. <u>Although a more stringent requirement was used, the fraction</u>**

<u>of students who were attributed to a school who could be linked to a teacher increased noticeably</u> to 83 percent, up from 60 percent in the beta model year, 2010–2011. The increase was due to the data submission being mandatory for districts in 2011-2012. Most of those students not linked to a teacher in 2011-2012 were accounted for in the data received from districts, but they were not enrolled with a teacher for the required duration.

- In addition, the quality of fit statistics (pseudo R-squared, as explained in the body of the report) for the adjusted model increased in all but one grade and subject, almost doubling in one grade and subject. **This represents a substantial improvement in predictions**. This is most likely due to a decrease in the fraction of students who scored the highest or second highest possible scale score on the grade 3 English Language Arts test from the 2010-2011 school year to the 2011-2012 school year.

INTRODUCTION

New York Education Law §3012-c requires a new performance evaluation system for classroom teachers and building principals in New York State. Under the new law, New York State will differentiate t**eacher and principal effectiveness using four rating cat**ego**R**ies: Highly Effective, Effective, Developing, and Ineffective (HEDI). Education Law §3012-c(2)(a) requires annual professional performance reviews (APPRs) to result in a single composite teacher or principal effectiveness score, which incorporates multiple measures of effectiveness. The results of the evaluations will be a significant factor in employment decisions, including but not limited to promotion, retention, tenure determinations, termination, and supplemental compensation. It will also affect decisions regarding teacher and principal professional development (including coaching, induction support, and differentiated professional development).

For 2011–2012, 20 percent of a teacher's or principal's evaluation will be based on student growth on State assessments (for teachers in grades 4–8 English Language Arts [ELA] an/or Mathematics and principals employed in schools with these grades/subjects), with another 20 percent being based on other locally selected measures of student achievement, identified through collective bargaining. The remaining 60 percent will be based on "other" measures of educator effectiveness including among measures, classroom observations of teachers.

For the 2011–2012 school year, Education Law §3012-c requires the use of student growth data for the first 20 percent, and the statute contemplates a "value-added" model for use in the 2012–2013 school year and beyond (increasing the first 20 to 25 percent upon approval of a value-added model). The statute requires that the value-added model be approved by the Board of Regents after consultation with an advisory committee. For the 2011–2012 school year, New York committed to the use of a student growth model that would generate student growth percentile (SGP) scores based on data from the New York State testing program in ELA and Mathematics in grades 3 through 8. These SGP scores are aggregated to the teacher and school levels and used for educator evaluation. The student growth model used in 2011–2012 takes into consideration up to three years of prior test history and three student characteristics: poverty, disability status, and status as an English Language Learner (ELL). As such, students are compared with similar students i.e., their growth is measured relative to that of other students with similar prior test scores and other characteristics.

For teachers whose students do not take State assessments in grades 4 through 8 in ELA or Mathematics, alternative measures of student learning growth will be developed for the growth subcomponent following State guidance around a student learning objectives process. Results from the growth model will also be incorporated into the State's metrics used for school accountability as part of New York's Elementary and Secondary Education Act (ESEA) waiver.

Because New York's law and regulations envision transitioning to a value-added model, the specific form of the model implemented to develop SGP scores was selected to enable a smooth transition to such a model. Specifically, SGPs are generated from a statistical model known as a mixed random effects model, which is described below. Three prior years of test scores are used as predictor variables, along with a set of measured characteristics for students. The model is an error-in-variables (EiV) regression that accounts for measurement variance in the prior test score variables and in the dependent variable.

Results presented in this report are based on 2011–2012 and prior school years' data. Initial models were estimated with 2010–2011 and prior school years' data to determine the most accurate and fair way of designing a final model, with stakeholder input. 2010-2011 results were not used for evaluation purposes.

The 2011–2012 data show some improvement over the 2010–2011 data:

- The single most important improvement is that the fraction of students linked to a teacher increased between the two years. In 2010–2011, 40 percent of students with valid data were not linked to a teacher. In 2011-2012, the fraction of students not linked to a teacher was 17 percent. At the same time, a stricter rule of association (at least 195 days for ELA and 203 days for Mathematics) was put into place. This rule was intended to require that teachers and students have a substantial amount of time together in order to attribute a student's growth to an individual teacher. Additionally, districts were required to submit teacher-student linkage data to the New York State Education Department (NYSED)for the 2011-2012 school year, which led to an overall increase in the data used for these analyses. Most of the students not linked to a teacher in 2011-2012 were accounted for in the data received from districts, but they were not enrolled with a teacher for the required duration.
- In addition, the quality of fit statistics (pseudo R-squared, as explained below) for the adjusted model increased in all but one grade and subject and almost doubled in grade 4 ELA, from 0.315 in 2010–2011 to 0.61 in 2011–2012. This improvement is most likely due to a decrease in the fraction of students who scored the highest or second highest possible scale score on the grade 3 ELA test from the 2010-2011 school year to the 2011-2012 school year.

The models described in this report were selected and developed based on technical and data considerations and on the recommendations of NYSED as well as members of a statewide advisory committee, (known as the Regents Task Force), which included teachers, principals, union representatives, school boards, school district and BOCES officials and other interested parties.. A list of Task Force members is provided in Appendix A. In addition, a technical advisory committee reviewed the student growth model for technical accuracy and utility. Technical advisory committee members are also listed in Appendix A.

This technical report contains four main sections:

- **Data:** Description of the data used to implement the student growth model, including data processing rules and relevant issues that arose during processing.
- **Model:** Statistical description of the model.
- **Reporting:** Description of reporting metrics and computation of effectiveness scores.
- **Results:** Overview of key model results aimed at providing information on model quality and characteristics.

DATA

To measure student growth and attribute that growth to educators, at least two sources of data are required: student test scores that can be observed over time and information describing how students are linked to schools, teachers, and courses (i.e., identifying which teachers teach which students for which tested subjects and which school(s) those students attended). In addition, models may also use other information about students and schools, such as student background characteristics, as predictors or controls in the models.

Notably, in 2011-2012, the linkage rule for teachers, schools, and districts was updated. Most significantly, a student must be enrolled for at least 195 days (ELA) or 203 days (Mathematics) to be attributed to a teacher. In 2010-2011, there was no enrollment minimum. Despite this more stringent requirement, substantially more students were linked to teachers in 2011-2012 than in 2010-2011. Of students with two consecutive test scores, the fraction linked to teachers increased from 60% to 83% between the two years.[1]

In 2011-2012, 1,948,911 students had valid data (as described in Appendix B).[2] Of those, 1,620,850 (about 83%) were linked to a teacher. 1,852,185 of students with valid data (about 95%) were linked to a school. Slightly more students (1,860,388) were linked to a district. More detail follows.

There are about 50,000 teachers in the incoming data files – these are teachers with at least one record in the files New York State delivered– as well as about 4,000 principals, and over 800 districts.[3] New York State Boards of Cooperative Education Services [BOCES] and charter schools with a distinct district code that submitted data to NYSED for linkage are included in the count of districts.

The following section describes the data used for model estimation in New York, including some of the issues and challenges that arose and how they were handled.

Teacher-Student-Course Linkages

A critical element of growth analyses is the accurate identification of the courses students are taking in which they learn the content and skills covered on the tests used to measure teacher contributions to their learning. Another critical element is identifying who is teaching those courses.

A first step is to identify which courses are considered "relevant"—that is, courses in which instruction is provided that is aligned to the test being used to measure student growth. New York has developed a common set of course codes across the State, and we used the courses

[1] What is required for a student to have valid data is described in Appendix B.
[2] From the last block of Tables B-2, there are 978,861 Math students included in the reference file (meaning they have valid data), and from the last block of Table B-3, there are 970,050 ELA students in the reference file. Summing these two numbers gives 1,948,911 unique students with valid data.
[3] The number of districts reported is higher than the number of districts in the State because schools that were linked to two districts were linked instead to a "pseudo-district" for reporting purposes.

identified as "relevant" by the State for analysis. Appendix C provides a list of the item descriptions used.

New York also provided data files showing student enrollment in courses and teacher assignment to those courses. Students enrolled in relevant courses were attributed to the teacher who was scheduled for that course. The State is transitioning to the collection of more detailed linkage data, which may provide opportunities for more fine-grained linkages in the future.

Although course section data were available, scores are provided only at the "course-" or "subject-" level, meaning that teachers' scores may reflect multiple classrooms of students in the same content area.

Recall that, this year, the definition of a link between a student and teacher was 195 days (ELA) or 203 days (Mathematics) of enrollment between the student and teacher. Some students were not enrolled for a sufficient duration in any courses identified as relevant but had test scores in relevant subjects. In these cases, students were attributed to schools (principals) but not individual teachers (through assignment to a "pseudo" teacher variable). Overall, only 17 percent of test-takers with sufficient data for inclusion in the model could not be linked to a teacher. This is a sharp improvement from 40 percent in 2010–2011 although there was no minimum number of days of enrollment required in 2010–2011. It is important to note that districts were not required to submit teacher-student linkage data to NYSED in 2010-2011, but were required to submit these data in 2011-2012, resulting in an overall increase in the linkages that could be made.

Note that the linkage rate is not expected to be 100 percent. Not all students receive a full year of instruction within a single classroom and school, with a link to that classroom teacher and school in administrative data. A number of situations may arise. Students may take the standardized test but not meet the linkage rules and minimum enrollment time described above. Students may switch classrooms and schools within a school year, sometimes back and forth and multiple times. Teachers may leave a school mid-year for various reasons.

Table 1 shows the linkage between students with valid data and teachers.[4] School/principal and district linkage rates are shown in the next subsection.

[4] What is required for a student to have valid data is described in Appendix B.

Table 1: Teacher-Student Linkage Rates

Grade	Students With Valid Test Data	Students With Valid Data Who Are Linked To Teachers For the Required Duration Of Time	Gross Linkage Rate
4	385,206	332,831	86.4%
5	390,933	337,920	86.4%
6	393,846	322,221	81.8%
7	388,162	311,893	80.3%
8	390,764	315,985	80.8%
all	1,948,911	1,620,850	83.1%

In 2011–2012, NYSED has set a minimum number of days that students must be enrolled in courses for teachers or principals to be attributed to them. For 2012–2013 and beyond, districts will provide the State with a more fine-grained "instructional linkage" time, and growth or value-added scores are expected to be weighted by the amount of instructional time provided by each teacher or in each school. As NYSED and AIR analyze that more fine-grained data, the "continuous enrollment" rule used in 2011-2012 may be modified to account for these more nuanced data. Appendix D shows all current attribution rules.

School and District Linkages

Students are linked to schools (principals) and districts based solely on their enrollment linkage indicator. Thus some students are linked to a district or a school but not a teacher.[5]

The linkage rates at these levels are higher than at the teacher level, with a 95% student-school linkage rate and 95.5% student-district linkage rate.

[5] It is also possible (but rare) for a student to be linked to a teacher but not to a school or district.

Table 2: School-Student Linkage Rates

Grade	Students With Two Consecutive Test Scores	Students With Two Consecutive Test Scores Who Are Linked To Schools	Gross Linkage Rate
4	385,206	363,906	94.5%
5	390,933	371,172	95.0%
6	393,846	373,649	94.9%
7	388,162	370,028	95.3%
8	390,764	373,430	95.6%
all	1,948,911	1,852,185	95.0%

Table 3: District-Student Linkage Rates

Grade	Students With Two Consecutive Test Scores	Students With Two Consecutive Test Scores Who Are Linked To Districts	Gross Linkage Rate
4	385,206	365,974	95.0%
5	390,933	372,702	95.3%
6	393,846	375,242	95.3%
7	388,162	371,634	95.7%
8	390,764	374,836	95.9%
all	1,948,911	1,860,388	95.5%

In addition, not all educators/districts had links to students. The numbers of incoming educators/districts and the number linked to a student are shown in Table 4.

Table 4: Linkage Rates for Educators and Districts

Educator/Aggregation Level	Number In Incoming Files	Number Linked
Teacher	50,009	37,450
Principal	4,300	3,668
District	848	957

A little over half of the sharp drop in the number of teachers when comparing the number in incoming files to the number linked is due to implementing the continuous enrollment criterion, and the rest is due to teachers of students without reported test scores (6,747 and 5,812 respectively).

Test Scores

New York's student growth model drew on test score data from the statewide testing program in grades 3 through 8 in ELA and Mathematics. These tests are given in the spring. This report describes construction and properties of the 2011–2012 growth model using student test results from the 2011–2012 school year and prior years.

Models were run separately for each grade and subject. Determining which prior achievement scores to use in predicting performance in a particular subject is one key decision to be made in implementing a growth model. The New York State tests at the elementary grade levels include a variety of content aimed at measuring a range of knowledge and skills in Mathematics and ELA. Although the specific content or skills covered may change from year to year, we use test scores in each subject area as the predictor for that subject area (e.g., Mathematics scores are used to predict Mathematics scores).

In addition to determining which prior test scores to use as predictors, we also need to determine how many years of prior achievement ought to be included as a predictor. In some cases, this may be driven by the availability of data—for example, fourth grade students typically have only one prior year of scores based on State assessment data. The benefit of including additional years of data is that it may improve the precision of the prediction.

New York's growth models include three prior test scores in the same subject area and missing data indicators for prior test scores except the immediate prior test score. For the 2011–2012 analyses, data from 2011–12 were used as outcomes, with prior achievement predictors coming from the three years before (going back to 2008–2009). Specific tests used vary by grade and subject and are as follows:

- Grade 4 ELA and Mathematics models use scores from grade 3 in ELA and Mathematics.
- Grade 5 ELA and Mathematics models use scores from grades 3 and 4 in ELA and Mathematics.
- Grades 6–8 ELA and Mathematics models use scores from grades 3–7 in ELA and Mathematics.

To implement the EiV approach, AIR used conditional standard errors published in the technical reports for the assessments' prior year test scores and a table of the same for 2011-2012 test scores.

Students who repeated or skipped a grade other than the immediate prior grade were included. For example, a grade 5 student with an immediate prior grade 5 score would be excluded, but a grade 5 student with an immediate prior grade 4 score and a second-year prior score also from grade 4 would be included (without using the test score from two years ago and instead setting a missing flag for that year). This approach was taken in order to include as many students as possible.

Students with no valid immediate prior or outcome scores were dropped from analysis—for example, students without a prior year test score or with a prior year test score for the same grade as the current year test score—and a growth measure cannot be computed for those students. Students whose scores could not be merged over time were dropped from analysis as well. More detail on exclusion rules and results of applying those rules (along with other specifications) is included in Appendix B.

Student and School Characteristics

The results of growth models are used to measure the effects of educators on student learning gains, taking into account a student's prior achievement. It is possible however, that some factors may impact student learning gains above and beyond educators' contributions, or these factors may be confounded or entangled with the educators' contributions. For example, learning gains for a student could be influenced by his or her disabilities beyond what would be expected based only on the student's prior achievement. Including a measure of disability status in the model makes an adjustment in what the current year score is expected to be, given disability status, relative to other students.

For 2011–2012, NYSED used models with and without covariates, so that both "unadjusted" and "adjusted" scores could be provided for teachers and principals. Unadjusted scores will be utilized for school accountability and adjusted scores for educator evaluation.

NYSED's regulations permit three specific control variables at the individual student level for inclusion in the model designed to produce adjusted scores through 2011–2012, without additional classroom- or school-level variables. These variables, which are listed below, were selected after consultation with the Regents Task Force.. Additional variables may be included in a value-added model in future years, including classroom- or school-level variables that may reflect the context in which learning occurs. These are the student-level predictor variables:

- **ELL status:** A Y/N variable was provided to indicate ELL status.
- **Students with disabilities (SWD) status:** A Y/N variable was provided to indicate SWD status.
- **Poverty or economic disadvantage (ED):** A Y/N variable was provided reflecting New York State's rules related to family income levels and participation in economic support programs. A description is provided in Appendix E.

MODEL

In this section we describe the implementation of the statistical model to measure student growth in New York. We begin with a description of the statistical model and follow with a description of how SGPs are derived from the estimates produced by the model. In addition, we describe how mean growth percentiles (MGPs) and all variance estimates are produced.

At the core of the New York growth model is the production of an SGP. This is a statistic that characterizes the student's current year score relative to other students with similar prior test score histories. For instance, an SGP equal to 75 denotes that the student's current year score is better than 75 percent of the students in the data with prior test score histories and other measured characteristics that are similar. It does *not* mean that the student's growth is better than that of 75 percent of all other students in the population.

One common approach to estimating SGPs is to use a quantile regression model, as was used in Colorado (Betebenner, 2009). This approach models the current year score as a function of prior test scores and finds the SGP by comparing the current year score to the predicted values at various quantiles of the conditional distribution. Although we use SGPs as a summary of student performance, relative to similar students, we do not use the Colorado growth model to calculate the SGPs: our approach is described below.

The methods described here do not rely on the quantile regression method for two reasons. First, the typical implementation of the quantile regression models the observed scores and makes no correction for measurement variance in the predictor variables or in the outcome variable. It is known that ignoring the measurement variance in the predictor variables yields bias in the model coefficients (e.g., Wei & Carroll, 2009). Further complicating the issue, the measurement variance in the outcome variable also adds to the bias in a quantile regression (Hausman, 2001), an issue that does not occur with linear regression.

Second, over time New York may wish to transition from the growth percentile reporting metric used in 2011-2012 to other metrics commonly used for reporting value-added results, and the current model can enable such a transition without change to its underlying statistical form.

The model described in this section is designed to account for measurement variance in the predictor variables, as well as in the outcome variable, to yield unbiased estimates of the model coefficients. Subsequently, these model coefficients are used to form a predicted score, which is ultimately the basis for the SGP. Because the prediction is based on the observed score, it is necessary to account for measurement variance in the prediction as well. Hence, the model accounts for measurement variance in two steps: first in the model estimation and second in forming the prediction.

Covariate Adjustment Model

The statistical model implemented for the State of New York is typically referred to as a covariate adjustment model (McCaffrey, Lockwood, Koretz, & Hamilton, 2003), as the current year observed score is conditioned on prior levels of student achievement as well as other possible covariates.

In its most general form, the model can be represented as:

$$y_{ti}=\mathbf{X}_i\boldsymbol{\beta}+\sum_{r=1}^{L}y_{t-r,i}\gamma_{t-r}+\sum_{q=1}^{Q}\mathbf{Z}_{qi}\boldsymbol{\theta}_q+e_i$$

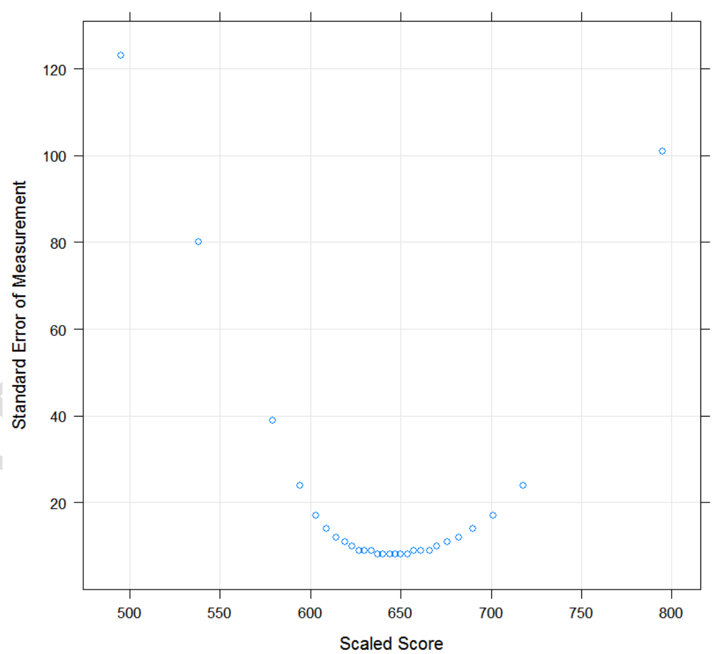
where y_{ti} is the observed score at time t for student i, $\mathbf{X}_i$ is the model matrix for the student and school-level demographic variables, $\boldsymbol{\beta}$ is a vector of coefficients capturing the effect of any demographics included in the model, $y_{t-r,i}$ is the observed lag score at time $t–r$ ($r \in 1,2,...,L$), $\boldsymbol{\gamma}$ is the coefficient vector capturing the effects of lagged scores, and $\mathbf{Z}_{qi}$ is a design matrix with one column for each unit in q ($q \in 1,2,...,Q$) and one row for each student record in the database. The entries in the matrix indicate the association between the test represented in the row and the unit (e.g., school, teacher) represented in the column. We often concatenate the sub-matrices such that $\mathbf{Z}=\{\mathbf{Z}_1,\ldots,\mathbf{Z}_Q\}$. $\boldsymbol{\theta}_q$ is the vector of effects for the units within a level. In New York, it represents the vector of school or teacher random effects for which we assume $\boldsymbol{\theta}_q \sim N(0, \sigma^2_{\boldsymbol{\theta}_q})$ for each level of q.

Corresponding to $\mathbf{Z}=\{\mathbf{Z}_1,\ldots,\mathbf{Z}_Q\}$, we define $\boldsymbol{\theta}'=(\boldsymbol{\theta}_1',\ldots,\boldsymbol{\theta}_Q')$. In the subsequent sections, we use the notation $\boldsymbol{\delta}'=\{\boldsymbol{\beta}', \boldsymbol{\gamma}'\}$, and $\mathbf{W}=\{\mathbf{X}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2},\ldots, \mathbf{y}_{t-L}\}$ to simplify computation and explanation.

Accounting for Measurement Variance in the Predictor Variables

All test scores are measured with variance, and the magnitude of the variance varies over the range of test scores. The standard errors (variances) of measurement are referred to as conditional standard errors of measurement (CSEMs) since the variance of a score is heteroscedastic and depends on the score itself. Figure 1 shows a sample from the grade 8 ELA test in New York. More information about CSEMs in New York State assessments can be found here: http://www.p12.nysed.gov/apda/reports/.

Figure 1: Conditional Standard Error of Measurement Plot, Grade 8 ELA



Treating the observed scores as if they were the true scores introduces a bias in the regression, and this bias cannot be ignored within the context of a high-stakes accountability system (Greene, 2001). In test theory, the observed score is described as the sum of a true score plus an independent variance component, $\boldsymbol{X}=\boldsymbol{X}_*+\boldsymbol{U}$ where $\boldsymbol{U}$ is a matrix of unobserved disturbances with the same dimensions as $\boldsymbol{X}$.

To describe how the model accounts for measurement variance, we first re-express the true score regression as:

$$\mathbf{y}_{t*}=\mathbf{X}\boldsymbol{\beta}+\sum_{r=1}^{L}\mathbf{y}_{t-r*}\boldsymbol{\gamma}_{t-r}+\sum_{q=1}^{Q}\mathbf{Z}_q\boldsymbol{\theta}_q+\mathbf{e}$$

We use * to denote the variables without measurement variance. For convenience, define the matrices $\mathbf{W}=\{\mathbf{X},\mathbf{y}_{t-1},\mathbf{y}_{t-2},\ldots,\mathbf{y}_{t-L}\}$, $\mathbf{W}_*=\{\mathbf{X},\mathbf{y}_{t-1*},\mathbf{y}_{t-2*},\ldots,\mathbf{y}_{t-L*}\}$, and $\boldsymbol{\delta}'=\{\boldsymbol{\beta}',\boldsymbol{\gamma}'\}$. Label the matrix of measurement variance disturbances $\boldsymbol{U}$ for disturbances associated with $\mathbf{y}_{t-1},\mathbf{y}_{t-2},\ldots,\mathbf{y}_{t-L}$, and label the vector of measurement disturbances with the dependent variable, $\mathbf{y}_t$, $\mathbf{v}$, hence $\mathbf{y}_t=\mathbf{y}_{t*}+\mathbf{v}$. Let $\boldsymbol{U}$ have the same dimension as $\mathbf{W}$, but only the final L columns of $\boldsymbol{U}$ are non-zero, so $\mathbf{W}=\mathbf{W}_*+\mathbf{U}$. If those disturbances were observed, the parameters $\{\boldsymbol{\delta}',\boldsymbol{\theta}'\}$ can be estimated using Henderson's methods (1950) by solving the following mixed model equations:

$$\begin{bmatrix}\mathbf{W}_*'\boldsymbol{\Omega}^{-1}\mathbf{W}_* & \mathbf{W}_*'\boldsymbol{\Omega}^{-1}\mathbf{Z}\\ \mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{W}_* & \mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{Z}+\mathbf{D}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}\\ \boldsymbol{\theta}\end{bmatrix}=\begin{bmatrix}\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{y}_*\\ \mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{y}_*\end{bmatrix}$$

The matrix $\mathbf{D}$ is made up of Q diagonal blocks, one for each level in the hierarchy. Each diagonal is constructed as $\sigma_q^2\mathbf{I}_q$ where $\mathbf{I}_q$ is an identity matrix with dimension equal to the number of units at level q, and σ_q^2 is the estimated variance of the random effects among units at level q. When concatenated diagonally, the square matrix $\mathbf{D}$ has dimension $m=\sum_{q=1}^{Q}J_q$.

Two complications intervene. First, we cannot observe $\mathbf{U}$, and second, the unobservable nature of this term, along with the heterogeneous measurement variance in the dependent variable, renders this estimator inefficient.

Addressing the first issue, upon expansion we see that:

$$\mathbf{W}_*'\boldsymbol{\Omega}^{-1}\mathbf{W}_*=(\mathbf{W}-\mathbf{U})'\boldsymbol{\Omega}^{-1}(\mathbf{W}-\mathbf{U})=\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{W}-\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{W}-\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{U}+\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{U}$$

Since $\mathbf{W}=\mathbf{W}_*+\mathbf{U}$, we have $E(\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{U})=E(\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{U}), E(\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{W})=E(\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{U})$, hence $\mathbf{W}_*'\boldsymbol{\Omega}^{-1}\mathbf{W}_*=\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{W}-\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{U}$. Furthermore, we have $\mathbf{W}_*'\boldsymbol{\Omega}^{-1}\mathbf{Z}=E(\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{Z})$, $\mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{W}_*=E(\mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{W})$, and $\begin{bmatrix}\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{y}_*\\ \mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{y}_*\end{bmatrix}=E\begin{bmatrix}\mathbf{W}'\boldsymbol{\Omega}^{-1}\mathbf{y}\\ \mathbf{Z}'\boldsymbol{\Omega}^{-1}\mathbf{y}\end{bmatrix}$.

Addressing the second issue, both the right-side and left-side variables in the model equation measured with variance contribute to the heteroscedasticity. While the correction $\mathbf{U}'\boldsymbol{\Omega}^{-1}\mathbf{U}$ eliminates the bias due to measurement variance, we still do not have a variance-free measure of $\mathbf{y}$ for any time period. Therefore, the residual is made up of:

$$\tilde{\mathbf{y}}-\mathbf{W}'\boldsymbol{\delta}=-\mathbf{U}'\boldsymbol{\delta}+\mathbf{v}+\mathbf{e},$$

where $\tilde{\mathbf{y}}=\mathbf{y}-\mathbf{Z}\boldsymbol{\theta}$, $\boldsymbol{\theta}$ is the conditional mean of the random effects. The residual variance of any given observation is

$$\sigma_{ti}^2=\sigma_e^2+\sigma_{v(ti)}^2+\sum_{r=1}^{L}\delta_{t-r}^2\sigma_{u,t-r(i)}^2,$$

where $\sigma_{v(ti)}^2$ is the known measurement variance of the dependent variable for student i at time t. Similarly, $\sigma_{u,t-r(i)}^2$ are the known measurement variance of r prior test scores. Now, let $\mathbf{\Omega}$ be a diagonal matrix of dimension N with diagonal elements σ_{ti}^2.

With the above, we can define the mixed model equations as:

$$\begin{bmatrix} \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{W}-\mathbf{U}'\mathbf{\Omega}^{-1}\mathbf{U} & \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{W} & \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{Z}+\mathbf{D}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\theta} \end{bmatrix}=\begin{bmatrix} \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{y} \end{bmatrix}$$

Replacing $U'\Omega^{-1}U$ with Its Expectations

As indicated, **U** is unobserved, and so solving the mixed model equation cannot be computed unless **U** is replaced with some observed values. First, the mixed model equations are redefined as:

$$\begin{bmatrix} \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{W}-\mathbf{S} & \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{W} & \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{Z}+\mathbf{D}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\theta} \end{bmatrix}=\begin{bmatrix} \mathbf{W}'\mathbf{\Omega}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{\Omega}^{-1}\mathbf{y} \end{bmatrix}$$

where **S** is a diagonal "correction" matrix with dimensions $p \times p$ accounting for measurement variance in the predictor variables, $p=p_{\mathbf{X}}+L$, and $p_{\mathbf{X}}$ is the column dimension of **X**.

The matrix **S** is used in lieu of $\mathbf{U}'\mathbf{\Omega}^{-1}\mathbf{U}$ based on the following justification. Recall that we previously defined $\mathbf{\Omega}$ as diag$(\sigma_{t1}^2, \sigma_{t2}^2, \ldots, \sigma_{tN}^2)$ and the matrix of unobserved disturbances is:

$$\mathbf{U}=\begin{bmatrix} \mathbf{0}_{p\mathbf{X}} & \mathbf{U}_L \end{bmatrix}$$

where $\mathbf{0}_{p\mathbf{X}}$ is a matrix of dimension of $p_{\mathbf{X}}$ with elements of 0, and:

$$\mathbf{U}_L=\begin{bmatrix} u_{11} & u_{12} & \ldots & u_{1L} \\ u_{21} & u_{22} & \ldots & u_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & \ldots & u_{NL} \end{bmatrix}$$

The theoretical result of the matrix operation yields the following symmetric matrix:

$$\mathbf{U}_L'\mathbf{\Omega}^{-1}\mathbf{U}_L=\begin{bmatrix} \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i1}^2 & \ldots & \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i1}u_{iL} \\ \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i1}u_{i2} & \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i2}^2 & \ldots \\ \vdots & \vdots & \ddots \\ \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i1}u_{iL} & \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{i2}u_{iL} \ldots & \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}u_{iL}^2 \end{bmatrix}$$

The theoretical result is limited only because we do not observe u_{ip} since it is latent. However, $E(u_{ip}u_{ip})=\sigma_{ip}^2$ where σ_{ip}^2 is taken as the conditional standard error of measurement for student i. The theoretical result also simplifies because variances of measurement on different variables are by expectation uncorrelated, $E(u_{ip}u_{ip'})=0$ when $p \neq p'$.

Because the conditional standard error of measurement varies for each student i and the off-diagonals can be ignored, let **S** be:

$$\mathbf{S}=\text{diag}\left(0,\ldots,0,\sum_{i=1}^N \frac{1}{\sigma_{ti}^2}\sigma_{u,t-1(i)}^2, \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}\sigma_{u,t-2(i)}^2, \ldots, \sum_{i=1}^N \frac{1}{\sigma_{ti}^2}\sigma_{u,t-L(i)}^2\right)$$

where $\sigma^2_{u,j(i)}$ denotes the measurement variance for the jth, $j = (1, 2, \ldots L)$, variable measured with variance.

Specification for New York Growth Model

The preceding section provides details on the general modeling approach and specifically how measurement variance is accounted for in the model. The exact specification for the New York model is described as:

$$y_{gi}=\mu+\sum_{l=1}^{K}\beta_l y_{g-r,i}+\sum_{s=1}^{M}\tau_s m_{si}+\sum_{q=1}^{J}\gamma_q x_{qi}+\theta_m+\theta_{m(j)}+\varepsilon_i$$

where y_{gi} is the current year test scale score for student i in grade g; μ is the intercept; β_l is the set of coefficients associated with the three prior test scores; τ_s is the set of coefficients associated with the missing variable indicators; γ_q is the set of coefficients associated with the student-level measured characteristics (which include binary variables for poverty, SWD, and ELL status); and θ_m, $\theta_{m(j)}$, and ε_i are the school, teacher, and student random effects.

There are only two missing variable indicators for the prior scores. This occurs because students are required to have the immediate prior score, $y_{g-1,i}$, in order to be included in the model. However, the values for $y_{g-2,i}$ or $y_{g-3,i}$ can be missing. If so, they are accounted for via the missing flag as:

$$\tau=\begin{cases}1 & \text{if missing}\\ 0 & \text{otherwise}\end{cases}$$

As noted earlier, the model is implemented separately for each grade and subject. When the model includes the covariates for poverty, SWD, and ELL status, it is referred to as the "adjusted" model. The "unadjusted" model is simply a special case of this model that does not contain those fixed effects, but everything else is the same.

Student Growth Percentiles

The model described in the section above yields unbiased estimates of the fixed effects. For purposes of the growth model, a predicted value and its variance for each student are required to compute the SGPs as:

$$\mathrm{SGP}_i=\Phi\left(\frac{y_i-\hat{y}_i}{\sqrt{\sigma^2_{\hat{y}f,i}}}\right)$$

where y_i is the observed value of the outcome variable and $\hat{y}_i=\boldsymbol{w}'\boldsymbol{\delta}$ where $\boldsymbol{w}'$ is the ith row of the model matrix $\boldsymbol{W}$ and the notation $\sigma^2_{\hat{y}fi}$ is used to mean the variance of the predicted value of y for the ith student.

Here the regression is of the form

$$\boldsymbol{y}=\boldsymbol{W}\boldsymbol{\delta}+\epsilon$$

where

$\epsilon \sim N(0, \sigma_e^2)$

The classic variance of a predictor is, for this case,

$$\sigma_{yf,i}^2 = (1 + \boldsymbol{w_i}'(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{w_i})\sigma_e^2$$

Where σ_e^2 is the standard error of the regression. However, in this case, we wish to make two refinements. The first is to use the actual variance on y_i, called $\sigma_{y_i}^2$, rather than the population variance on y_i, called $\sigma_{y_i}^2$, which is already included in σ_e^2. This is done by subtracting the population variance and adding back the individual variance. Thus, the variance on the predictor becomes

$$\sigma_{yf,i}^2 = (1 + \boldsymbol{w_i}'(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{w_i})\sigma_e^2 + \sigma_{y_i}^2 - \sigma_{y_i}^2$$

The second refinement is to replace the population variance in $\boldsymbol{w_i}$, called $\boldsymbol{\Sigma}$, with the individual variance in $\boldsymbol{w_i}$, called $\boldsymbol{\Sigma_i}$. This is done in the same way as the variance in y_i, so the variance estimate is now

$$\sigma_{yf,i}^2 = (1 + \boldsymbol{w_i}'(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{w_i})\sigma_e^2 + \sigma_{y_i}^2 - \sigma_{y_i}^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$$

There is then a predicted value for each student that is used to compute the SGP. However, that prediction is based on the estimates of the fixed effects that were corrected for measurement variance but based on the observed score in the vector $\boldsymbol{w}$.

Figure 2 below provides an illustration of how the SGPs are found from the previously described approach. The illustration considers only a single predictor variable although the concept can be generalized to multiple predictor variables, as presented above.

For each student, we find a predicted value conditional on his or her observed prior scores and the model coefficients. To illustrate the concept, assume we find the prediction and its variance but do not account for the measurement variance in the observed scores used to form that prediction. We would form a conditional distribution around the predicted value and find the portion of the normal distribution that falls below the student's observed score. This is equivalent to:

$$\mathrm{SGP}_i = \int_{-\infty}^{y_i} f(x)dx$$

with $f(x) \sim N(\hat{y}_i, \sigma_{yf_i}^2)$ although this is readily accomplished via the cumulative normal distribution function, Φ..

Figure 2: Sample Growth Percentile from Mixed Model

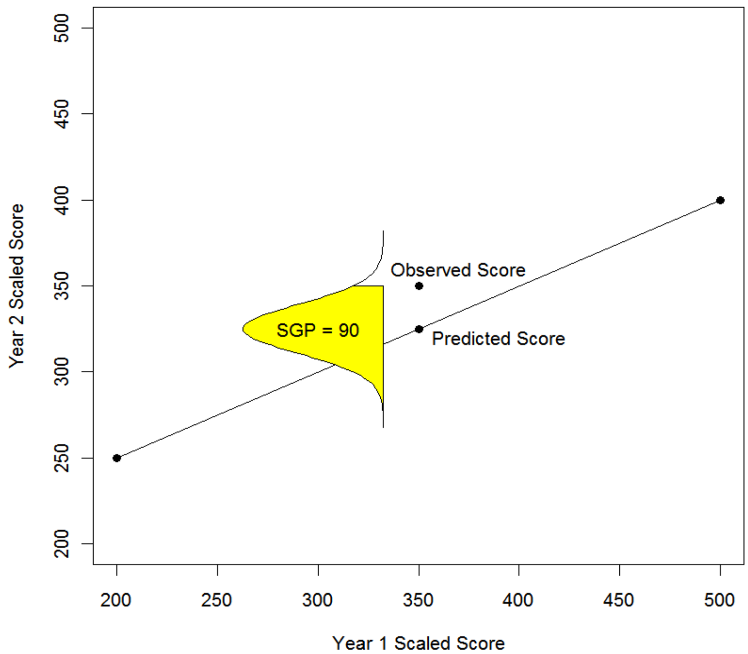
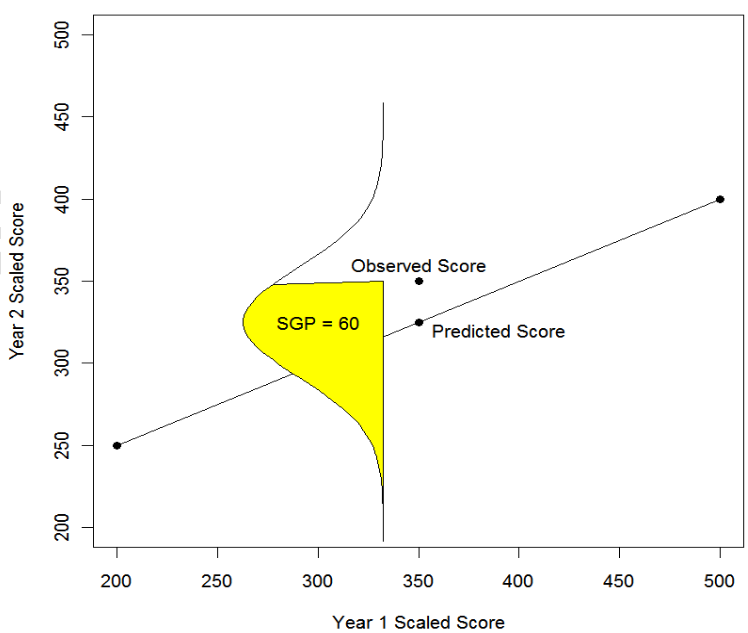


Figure 3 below illustrates the same hypothetical student as above. Note that the observed score and predicted value are exactly the same. However, the prediction variance is larger than in Figure 2 above. As a result, when we integrate over the normal from $-\infty$ to yi , the SGP is 60 and not 90 as in the example above. This occurs because the conditional density curve has become more "spread out," reflecting less precision in the prediction.

Figure 3: Sample Growth Percentile from Mixed Model



Interpolating Standard Errors of Measurement at the Lowest and Highest Obtainable Scale Scores (LOSS and HOSS)

The linear model used to produce student-level predictions $\boldsymbol{yi}$ can cause these predictions to fall outside the boundaries of the defined score scale. Let the floor or ceiling in the data be denoted as $\boldsymbol{\eta f}$ and $\boldsymbol{\eta c}$, respectively. It is therefore possible that $\boldsymbol{yi}< \boldsymbol{\eta f}$ or $\boldsymbol{\eta c}<\boldsymbol{yi}$. However, the observed score can never fall outside these bounds.

When a prediction falls outside the boundaries of the score scale, it can cause bias in the statistics used to characterize a student, teacher, or school. This phenomenon seems to occur as a result of the large conditional standard errors of measurement at the extreme scores, $\boldsymbol{csem(\theta i)}$. The procedure below is implemented to deal with these large standard errors.

Interpolation Procedure for Conditional Standard Errors of LOSS and HOSS

Interpolate new conditional standard errors of measurement from the following *k*th degree polynomial regression:

$$\boldsymbol{yi=\mu+k=1M\beta k\theta ik+ei}$$

where $\boldsymbol{yi}$ is $\boldsymbol{csem(\theta i)2}$ and $\boldsymbol{\theta}$ is the observed score for the *i*th student. The square root of the fitted values will then be used in lieu of the CSEM:

$$\boldsymbol{yi=\mu+k=1M\beta k\theta ik}$$

Implementation

Implement the linear regression and subsequently the growth model using the following steps:

1. Subset the data and include all scores except the values at $\boldsymbol{\eta f}$ and $\boldsymbol{\eta c}$.
2. Set M = 2 and run linear regression.
3. Using the coefficients from the regression, interpolate new values of $\boldsymbol{csem(\theta i)2}$ for the scores at the LOSS and HOSS.
4. Use $\boldsymbol{csem(\theta i)}$ from step 3 in lieu of the standard error of the LOSS or HOSS in the test score data.
5. Run the growth model.
6. Verify that $\boldsymbol{\eta f<yi<\eta c}$ for all *i*.
7. If the inequality in step 6 is not true, return to step 2 and increase to M = M + 1.
8. Repeat until the inequality in step 6 is true or M is large.

If the process above ended and it was not true that $\eta f<yi<\eta c$ for all i, then for all yi where $\eta f>yi$, set $yi=\eta f$.Similarly, for all yi where $\eta c<yi$, set $yi=\eta c$.

Mean Growth Percentiles

Once the SGPs are estimated for each student, group-level (e.g., teacher-level) statistics can be formed that characterize the typical performance of students within a group. The NYSED growth model technical advisory committee recommended using the mean SGP when providing educator scores. Hence, group-level statistics are expressed as the mean SGP within a group. This is referred to as the MGP.

For each aggregate unit j ($j \in 1,2,\ldots,J$), such as a class, the interest is a summary measure of growth for students within this group. Within group j we have $\{SGPj1,\ SGPj2,\ \ldots,\ SGPjN\}$. That is, there is an observed SGP for each student within group j.

Then the MGP for unit j is produced as:

$\theta j = \text{mean}(SGPj(i))$

Like all statistics, the MGP is an estimate, and it has a variance term. For New York, AIR provides the following measures of variance for the MGP.

The analytic standard error of the MGP is computed within unit j as:

$\text{se}(\theta j) \approx \text{sd}(\text{SGP}ij)Nj$

where $\text{sd}(\text{SGP}ij)$ is the sample standard deviation of the SGPs in group j and N is the number of students in group j.

The analytic standard error has two particular limitations. First, MGPs are bounded between 1 and 99; hence, the standard errors cannot be used to form confidence limits around the MGP because the confidence limits must be asymmetric. Second, the standard errors do not account for potential non-normality of the distribution of the SGPs. To account for these issues, a bootstrap procedure was implemented.

For each student, we compute an SGP based on

$\text{SGP} = \Phi yi - y\ i\sigma x\ f,i2$

The following bootstrap procedure is used:

1. Produce $\text{SGP} = \Phi yi - yi\sigma y\ f,i2$
2. For each teacher j, sample nj students with replacement where nj is the total number of students linked to teacher j.
3. Compute the MGP for teacher j as
   a. $\text{MGPj} = 1nji=1nj\text{SGPi}$

    b. Store the value computed in step 3(a).
4. Repeat steps 2 and 3 five hundred times.
5. Find the bootstrap confidence intervals and standard errors from the vector $\{\text{MGP}_j^1, \text{MGP}_j^2, \ldots, \text{MGP}_j^{500}\}$ where the superscript denotes the qth iteration of the bootstrap. The confidence interval is a range from the 2.5th percentile to the 97.5th percentile of the replicated MGP values.

Aggregation of Growth Percentiles

Teachers in some cases and principals will have students from different grades and with results from different tested subjects. For evaluation purposes, there is a need to aggregate these SGPs and form summary measures.

Because the SGPs are expressed as percentiles, they are free from scale-specific inferences and can be easily combined since percentiles are all on the same scale. For any aggregate-level statistics to be provided (in this case, MGPs), we simply pool all SGPs of relevant students and find the mean of the pooled SGPs. Variances of these SGPs are found using the same methods described above. More detail on what scores will be provided follows.

REPORTING

These are the main metrics generated based on the student growth model:

- **Unadjusted MGPs:** Unadjusted MGPs are created by finding the mean of the SGPs at the relevant level produced by a growth model that does *not* include any contextual variables.
- **Adjusted MGPs (and upper and lower limits based on confidence intervals for these adjusted MGPs):** Adjusted MGPs are created by finding the mean of the SGPs at the relevant level produced by a growth model that *does* include three contextual variables at the individual student level (ELL status, SWD status, and economic disadvantage).
- **Percent of students above the median:** Percent of students above the median is found as the percent of students above the State median in the subject and grade using the nominal student SGP from the adjusted model. Note that a teacher-specific growth score computed by NYSED based on the percentage of students earning a State-determined level of growth was required by 2012 New York State legislation as an optional, locally selected measure. It will be provided to all districts but will be of consequence for evaluation purposes only where districts have locally agreed to it in collective bargaining.
- **Number of student scores included in estimates at each reported level:** This element gives the count of student scores used to compute the adjusted or unadjusted MGP at any given level.
- **Growth rating (HEDI classification for growth subcomponent):** Based on an overall adjusted MGP for a teacher or principal across grades and subjects, the growth rating will describe the educator's performance category (Highly Effective, Effective, Developing, or Ineffective). Information on how these classifications are derived is available from NYSED.

- **Growth score:** Based on an overall MGP for a teacher, a growth score of 0–20 will give a finer-grained score within each State growth performance category. Note that this range may change to 0–25 in subsequent years if a value-added model is implemented.

MGP results for individual educators will be broken down, on descriptive reports provided to districts and educators, by subject and grade in addition to overall results. Overall MGPs (that is, MGPs based on all students who meet the criteria for a given teacher or school) will also be broken down by several different subgroups (ELL, SWD, economically disadvantaged, and low- and high-achieving). ELL, SWD, and economically disadvantaged subgroups will be created using the same rules used for the model (i.e., yes/no for all variables). Subgroups of low- and high-achieving students will be based on performance levels. Specifically, a student who is at level 1 in either subject will be considered low-achieving; a student who is at level 4 in either subject will be considered high-achieving; students at levels 2 in one subject and 3 in the other are not included in either category. Students at levels 2 or 3 in both subjects are not included in either category as well. Growth subcomponent ratings (HEDI categories and associated evaluation points) are not broken down by any of these subcategories.

Minimum Sample Sizes for Reporting

One question to be addressed when providing teacher or principal results is whether some of them should be suppressed because the number of students used to create the MGP would jeopardize either personal student information or the quality or stability of the estimate. Setting no (or a low) minimum sample size will result in the greatest number of teachers and principals receiving information; on the other hand, the quality of the information they receive may be poor.

When a larger number of student tests are aggregated together to estimate an MGP, the estimate is more accurate in terms of its estimated standard error. Thus, smaller classes have less accurate estimates. To maintain the quality of the estimates, it is therefore desirable to specify a minimum number of students below which the MGP is not reported. For 2011-2012, a threshold of 16 student scores, which could have been eight in each subject or all 16 in one subject, was chosen to minimize the number of teachers dropped from the analyses, while maintaining the statistical accuracy of estimates. Analytics were performed to determine the minimum number of student scores required that would ensure any uncertainty in MGPs was minimized to the extent possible. Accordingly, educator scores based on fewer than 16 student scores are not reported.

After applying these rules, the number of teachers, principals, and districts with reported results is shown in table 5.

Table 5: Reporting Rates for Educators and Districts

Educator/Aggregation Level	Number Linked	Number Meeting the Minimum Number of

		Student Scores Requirement
Teacher	37,450	33,801
Principal	3,668	3,556
District	957	932

Performance Categories

Teacher and principal performance (effectiveness) scores are based on their **<u>overall</u>** MGP scores (that is, the results for all the students for whom a teacher or principal is responsible in Mathematics and/or ELA across grades) and confidence ranges around their overall MGP scores. As noted, to create these overall MGP scores, we simply find the mean of SGPs for all students who are linked to a particular teacher or school.

For 2011–2012, teachers and principals are to be assigned to four performance categories and earn points within each category, as shown in Table 6.

Table 6: Growth Score Ranges and Categories

HEDI Classification	Definition	HEDI Score Range
Ineffective	Results are well below State average for similar students (or district goals if no State test).	0–2
Developing	Results are below State average for similar students (or district goals if no State test).	3–8
Effective	Results meet State average for similar students (or district goals if no State test).	9–17
Highly Effective	Results are well above State average for similar students (or district goals if no State test).	18–20

A description of the method used to assign teachers and principals to these categories is available from NYSED.

RESULTS

This section provides an overview of the results of model estimation using 2011-2012 as the most recent test year available and the available information about teacher/student data linkages. A pseudo R-squared statistic and summary statistics characterizing the SGPs, MGPs, and their precision provide an overview of model fit. Note that this section focuses on teacher-level results, although additional information on principal/school-level results is available in the appendices.

Appendices F and G provide model parameters, including model coefficients and variance components by grade and subject.

Model Fit Statistics

The R-square is a statistic commonly used to describe the goodness-of-fit for a regression model. Because the model implemented here is a mixed model and not a least squares regression, we refer to this as a pseudo R-square. Table 7 presents the pseudo R-square values for each grade and subject, computed as the squared correlation between the fitted values and the outcome variable.

Table 7: Pseudo R-Square Values by Grade and Subject

Subject	Grade	Unadjusted Model	Adjusted Model
ELA	4	0.61	0.61
	5	0.63	0.64
	6	0.66	0.67
	7	0.64	0.65
	8	0.63	0.64
Math	4	0.60	0.61
	5	0.65	0.66
	6	0.62	0.62
	7	0.70	0.70
	8	0.66	0.66

Student Growth Percentiles

The SGPs describe a student's current year score relative to other students in the data with similar prior academic histories and measured characteristics. Table 8 provides the correlation between the prior-year scale score and the SGP for each grade and subject. These small negative correlations coefficients are a result of using the EiV approach to account for measurement variance in the prior year scale score; the correlation need not be zero.

Table 8: Correlation between SGP and Prior Year Scale Score

Grade	ELA	Math
4	–0.13	–0.12
5	–0.15	–0.03
6	–0.15	–0.03
7	–0.12	–0.05
8	–0.14	–0.13

Mean Growth Percentiles

As described earlier in this report, teachers' MGPs are aggregate educator-level statistics, computed as the mean SGP for all students associated with a roster (teacher) or school (principal). In this section, we provide descriptive statistics on the MGP, the average of SGPs within the school or roster (with roster referring to all the students associated with a teacher in the growth analyses, which can cover one or more classrooms).

At the teacher level, about half of the teachers have results for students in only one subject (either Mathematics or ELA), and in those instances their MGP is subject-specific and combined across all sections of the subject. For teachers with results for students in both subjects, the MGP is combined across subjects.

Figure 4 below provides a histogram of the combined teacher MGPs, which is combined across subjects for teachers of ELA and Math, or across sections for single-subject teachers for the adjusted model (includes ELL, SWD, and ED). The results are normally distributed.

Figure 4: Distribution of Teacher MGP by Grade, Adjusted Model

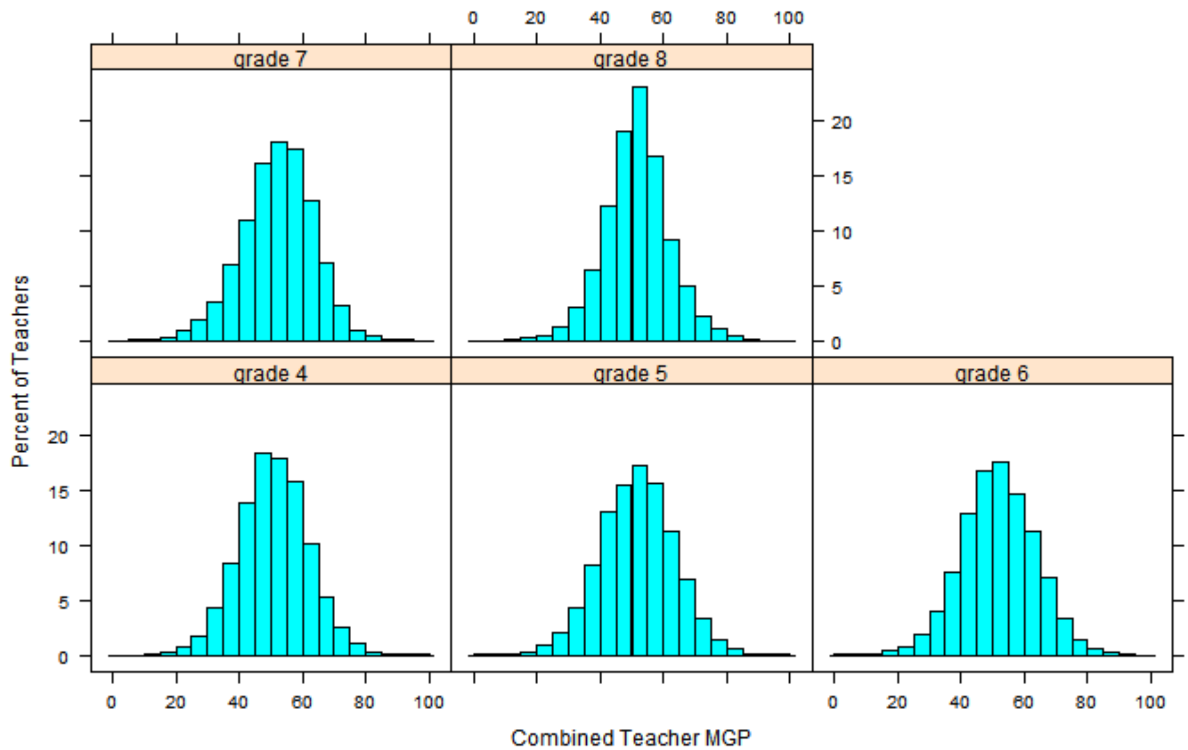
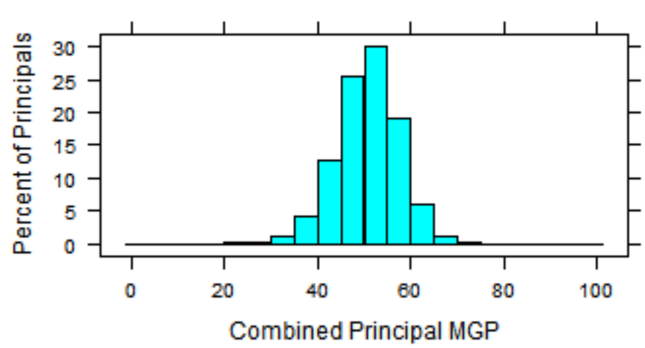


Figure 5 shows that for principals, the results are more lumped together in the center than for teachers, with a slight skew creating a longer "tail" on the left side – note that the sample sizes are different, with fewer principal MGPs than teacher MGPs.

Figure 5: Distribution of Principal MGP, Adjusted Model



Precision of the Mean Growth Percentiles

The caterpillar plot in Figure 6 is a random sample of 100 teacher MGPs taken from the New York data. The MGPs are sorted from lowest to highest with the corresponding 95 percent confidence range showing the lower and upper limits of the MGP. Figure 7 shows the same type of plot for principals where larger underlying samples mean that there is substantially less variation in the MGP and the error bars are narrower. Generally, these plots give an idea of the distribution of teacher MPGs and a typical confidence range.

Figure 6: Overall MGP with 95% Confidence Interval Based on Random Sample of 100 Teachers

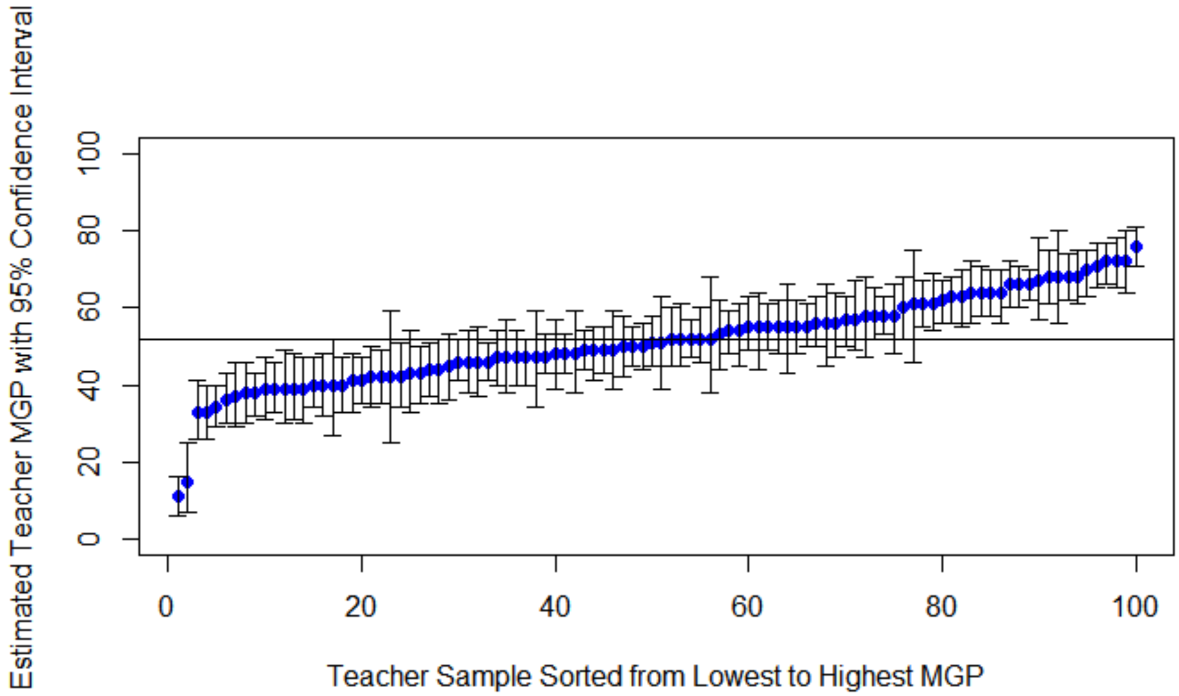


Figure 7: Overall MGP with 95% Confidence Interval Based on Random Sample of 100 Principals

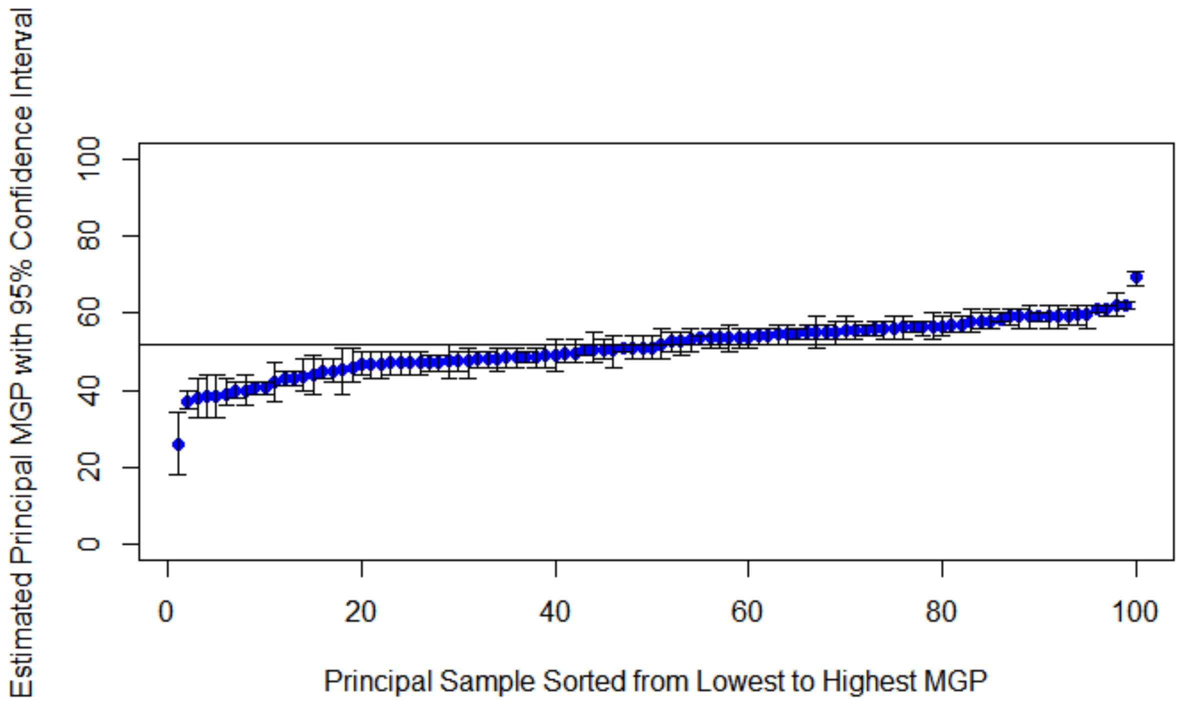


Table 9 provides the share of educators whose MGPs are significantly above or below the state mean for that educator type, using the 95 percent confidence intervals from the bootstrap. In all cases, the percent exceeding the mean is larger than what would be expected by chance alone (i.e., 5 percent expected by chance, or 2.5 percent for each table entry).

Table 9: Percent of Educator MGPs Above or Below Mean at the 95% Confidence Level

Educator Type	Below Mean	Above Mean
Teacher	23.7%	25.5%
Principal	32.2%	34.5%

These figures provide a means to visually gauge the precision of the MGPs. However, it may also be useful to examine a statistic to assess the precision of the teacher-level MGPs. The statistic, which we specify as ρ, is the ratio of the mean standard error to the standard deviation between teacher MGPs:

$$\rho = \sigma sd(\theta j)$$

where σ is the mean standard error of the MGP and $sd(\theta j)$ is the standard deviation between teacher MGPs. In theory, the highest possible value is 1, and it represents a complete lack of precision in the measure—when the ratio is 1, the variation in MGPs is explained entirely by sampling variation. Smaller values of ρ are associated with more precisely measured MGPs.

Table 10 below provides the mean standard errors, the standard deviation, and the value of ρ for the adjusted model by grade (again, for combined-subject MGPs). The values of the ratio (ρ) quantify imprecision in the estimates.

Table 10: Mean Standard Errors, Standard Deviation, and Value of ρ for Adjusted Model by Grade for Teachers and for Principals in All Grades

Grade (Teachers)	Adjusted Mean SE	Adjusted Standard Deviation	Adjusted ρ
4	4.0	10.8	0.37
5	4.0	11.6	0.35
6	4.2	11.5	0.36
7	3.8	11.1	0.34
8	3.5	10	0.35
Principals	1.93	6.6	0.26

Impact Data Results

Table 11 below provides the correlations of the combined-subject MGP (or for teachers with only one subject, their single-subject MGP) with five classroom characteristics: the three control variables at the individual student level NYSED's regulations permit for inclusion in the model and that were selected after consultation with Regents Task Force - percent ELL, percent SWD, percent poverty (or economically disadvantaged (ED)) –and the mean prior ELA or Mathematics score of the students. Correlations are presented for adjusted MGPs (the adjusted model includes demographic variables for individual students).[6] The correlations show that including these demographic variables in the adjusted model reduces but does not eliminate correlations between MGPs and classroom aggregates of model variables as compared to the unadjusted model.

The scatter plots shown in Figures 8 through 12 provide visual representations of the data underlying the correlations for teachers, and Figures 13 to 17 provide similar pictures of the data underlying school-level (principal MGP) correlations.[7] Ideally, the adjusted MGPs of teachers and principals will be uncorrelated with classroom or school characteristics, given the fact that student-level variables controlling for these characteristics are included in the model. MGPs, however, could be correlated with student characteristics for several reasons. One is that classroom factors may affect the learning environment—for example, a teacher in a highly concentrated ELL class may offer different instruction to a student with a given test history than

[6] The impact of these demographic characteristics on the expected value of students' current test scores used to compute SGPs can be seen through the model coefficients presented in Appendix G. The inclusion of these variables serves to make SGPs for students with different demographic characteristics comparable, given the prior test scores included in the model.

[7] These figures are all shown for teachers, broken down by grade and subject, in Appendix H. The results in this section are combined over grades and subjects.

another teacher with few ELL students. The correlation may also reflect differences in teacher characteristics and performance of students that vary by characteristics of classrooms and schools that are not fully captured by individual students' prior test scores. It is important to emphasize that while a correlation does exist between the aggregate demographics of a teacher's classroom and his/her MGP, it is still possible for all teachers to receive a wide range of MGPs, regardless of the composition of their class as can be seen in each of the scatter plots.

Figure 8 shows the relationship between MGPs and the classroom share of ELL students. The correlation is low and indicates that the MGP is unrelated to percent ELL in a class. This suggests that, in general, there is no significant advantage or disadvantage in terms of MGP if teachers have a higher share of ELL students in their classes. The dispersion of points around the line does indicate that there is a great deal of variability in MGPs across shares of ELL students in a classroom, meaning that teachers with varying concentrations of ELL students in their classrooms can achieve a wide range of MGPs.

Table 11 indicates that there is a moderate correlation with the percent SWD and percent economically disadvantaged in a class. This correlation is negative, indicating that as the percent SWD or percent economically disadvantaged in a class increases, the MGP tends to decrease. However, teachers with high and low concentrations of SWD and economically disadvantaged students are still able to achieve MGPs across the distribution.

The correlation with mean prior scores is also moderate in size and positive. This suggests that the MGP tends to be higher on average for classes serving students with higher mean levels of prior ability in both ELA and in Mathematics, even though SGPs for individual students have a low or moderate negative correlation with prior year SGPs (Table 8). The finding that MGPs for teachers are higher in classrooms where mean prior achievement is higher suggests that there are factors other than prior test scores (and the demographic variables included in the adjusted model) that may be related to the observed growth in test scores, but one can see in the scatter plots that teachers of low and high achieving students receive a range of MGPs.

Table 11: Teacher MGP Correlated with Class Characteristics

Percent	Adjusted Model
ELL in Class	0.00
SWD in Class	–0.06
ED in Class	–0.10
Mean Prior ELA	0.10
Mean Prior Math	0.13

Figure 8: Relationship of Teacher MGP Scores to Percent of ELL Students

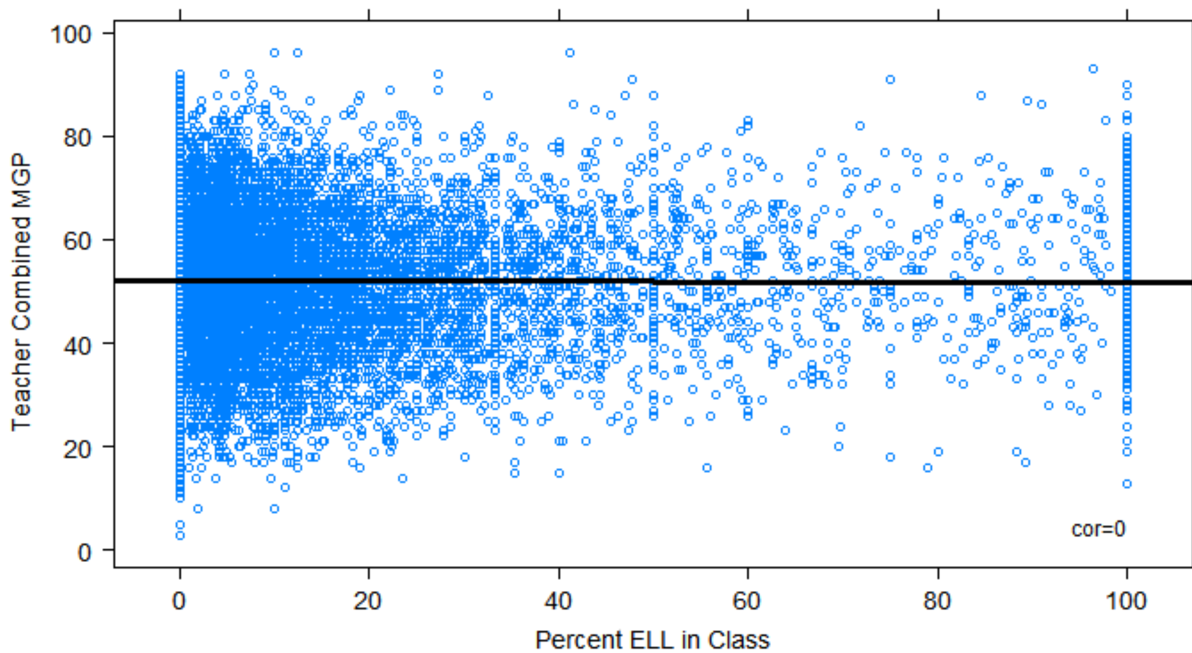


Figure 9: Relationship of Teacher MGP Scores to Percent SWD in Class

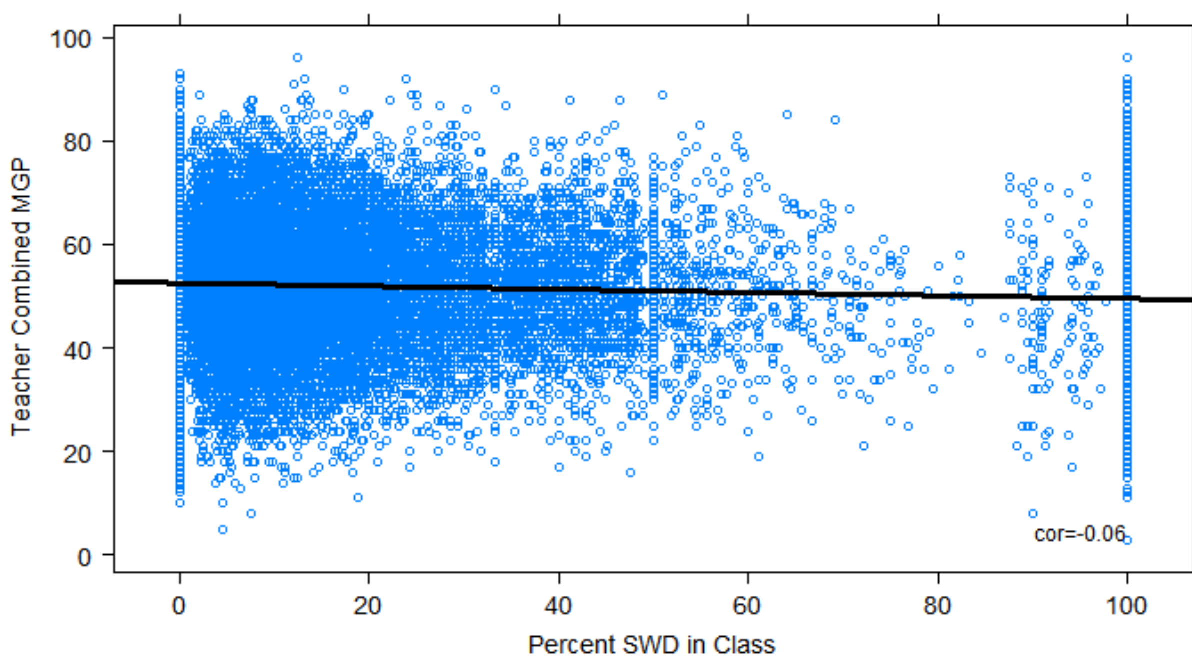


Figure 10: Relationship of Teacher MGP Scores to Percent of Economically Disadvantaged Students

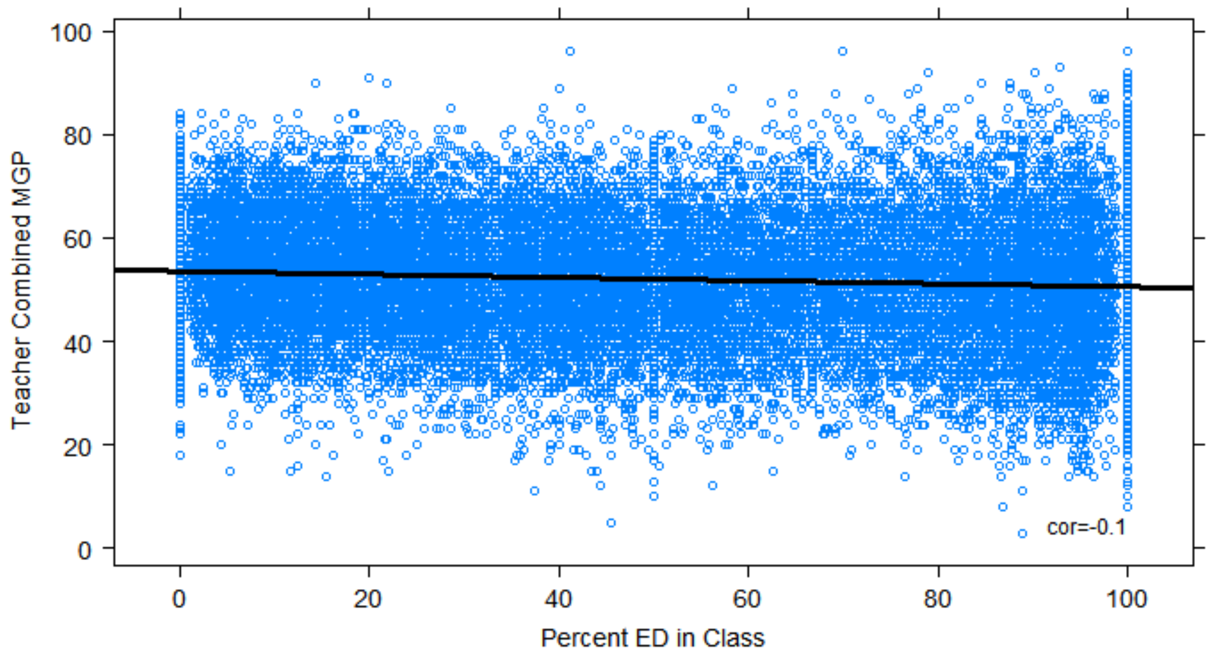


Figure 11: Relationship of Teacher MGP Scores to Mean Prior ELA Scores

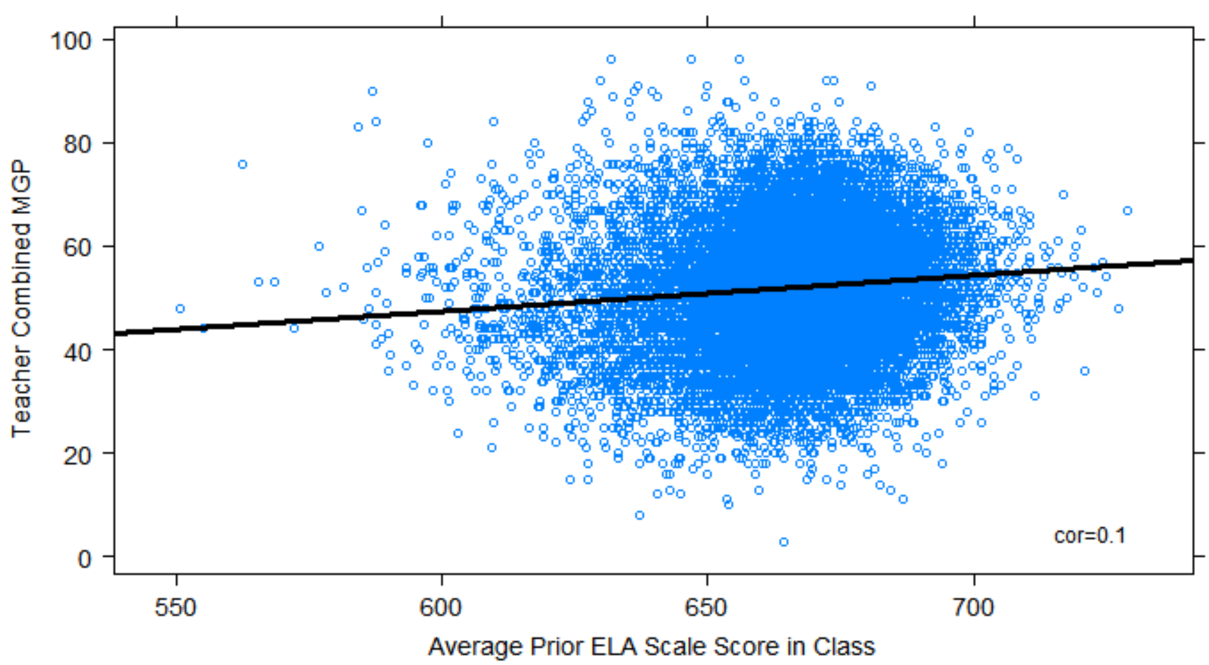


Figure 12: Relationship of Teacher MGP Scores to Mean Prior Mathematics Scores

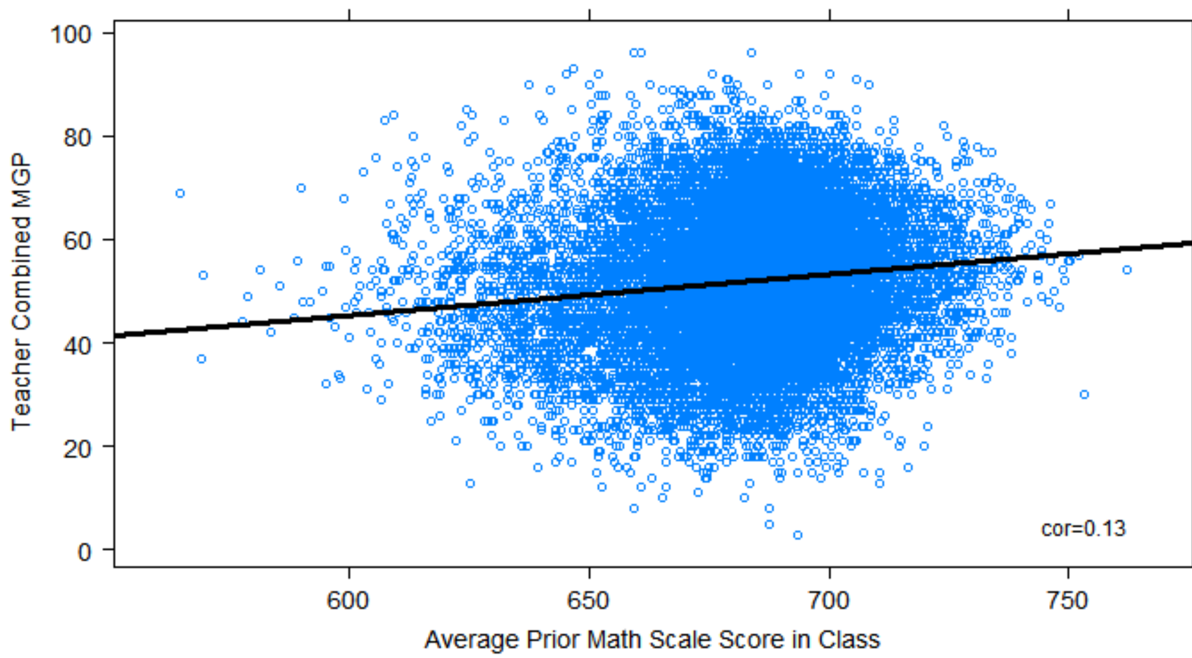


Table 12 below provides the observed correlations of principal MGPs with the five school characteristics matching those presented for teachers: percent ELL in school, percent SWD in school, percent poverty in school, and the mean prior ELA or Mathematics score of the students. Because these are school-level characteristics, they are combined-subject, combined-grade results. The scatter plots shown in Figures 14 through 19 provide visual representations of these data. Results are presented for adjusted MGPs, and, as was the case at the teacher level, the inclusion of student-level covariates reduces correlations when the unadjusted MGPs are compared with the adjusted MGPs.

The low correlation of the aggregate ELL to MGPs indicates that the MGP is unrelated to percent ELL in a school. This suggests that, in general, there is no significant advantage or disadvantage to principals conditional on these school characteristics, although there may be some relative advantage or disadvantage to some principals, given the variability in the data.

However, there is a moderate correlation with the percent SWD and percent economically disadvantaged in a school. This correlation is negative, indicating that as the percent SWD or percent economically disadvantaged in a school increases, the MGP tends to decrease.

The correlation with mean prior scores is also moderate in size and positive. This suggests that the MGP tends to be higher in schools with students with higher mean levels of ability in both ELA and in Mathematics. The finding that MGPs for principals are larger in schools in which mean prior achievement is higher suggests that students with higher SGPs (given their prior test scores) may attend schools in which there are larger shares of students who have higher test scores.

Table 12: Principal MGP Correlated with Demographic Characteristics

Percent	Adjusted Model
ELL in School	0.05
SWD in School	–0.23
ED in School	–0.11
Mean Prior ELA	0.35
Mean Prior Math	0.40

Figure 13: Relationship of Principal MGP Scores to Percent of ELL Students

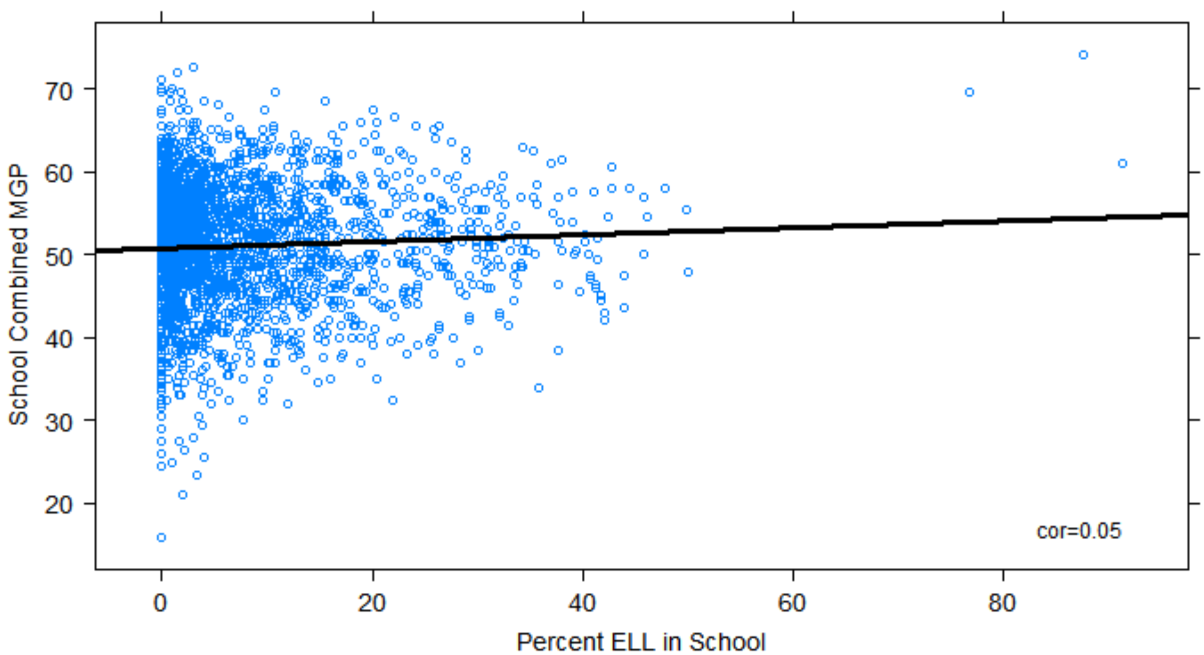


Figure 14: Relationship of Principal MGP Scores to Percent SWD in School

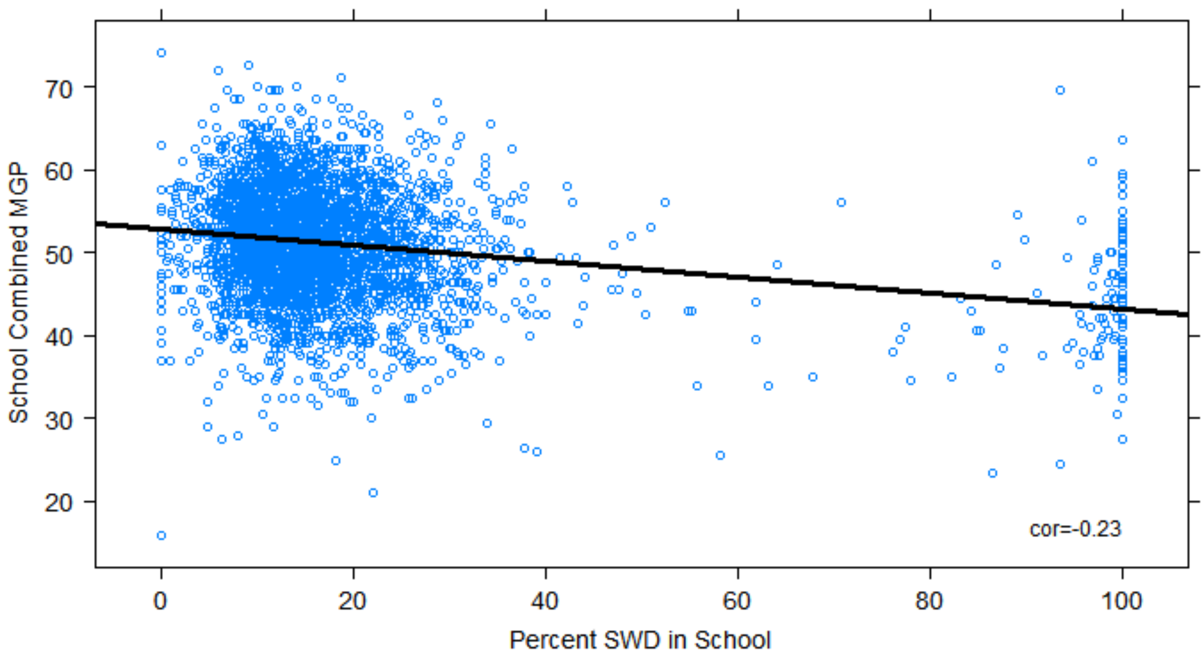


Figure 15: Relationship of Principal MGP Scores to Percent of Economically Disadvantaged Students

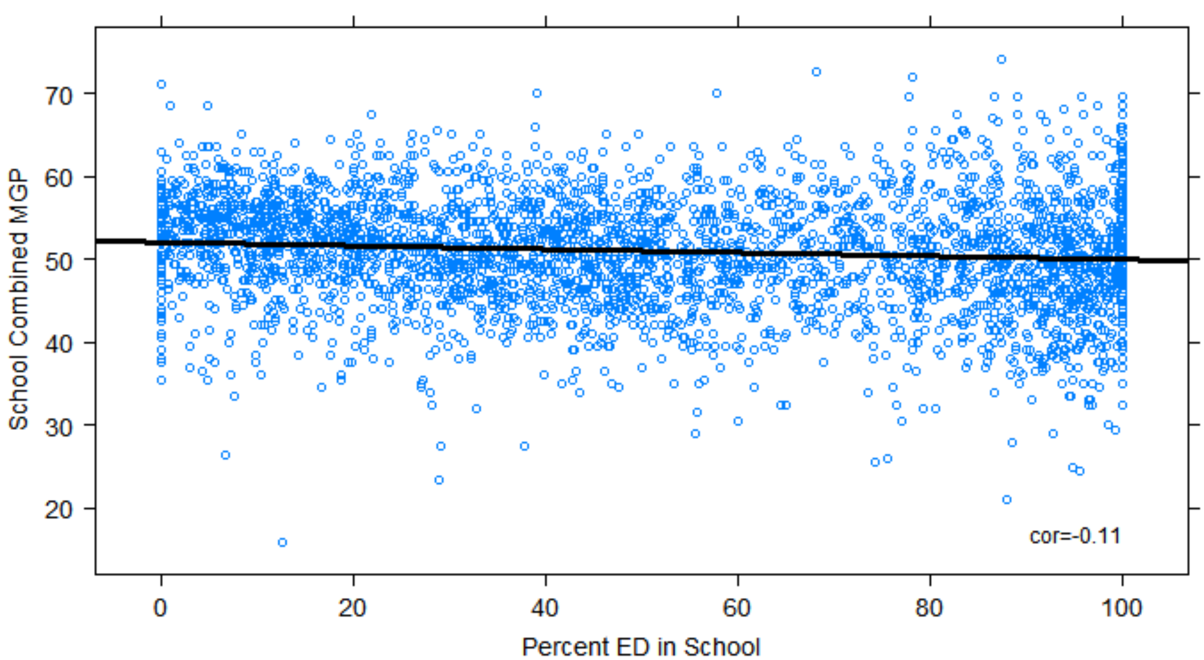


Figure 16: Relationship of Principal MGP Scores to Average Prior ELA Scores

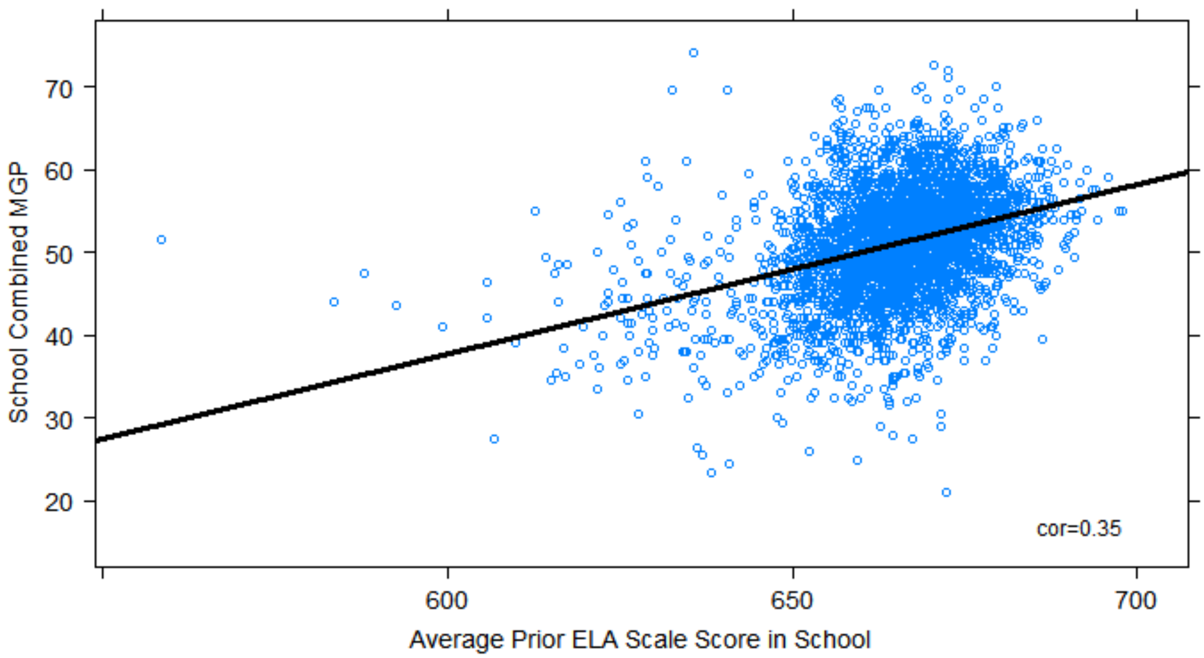
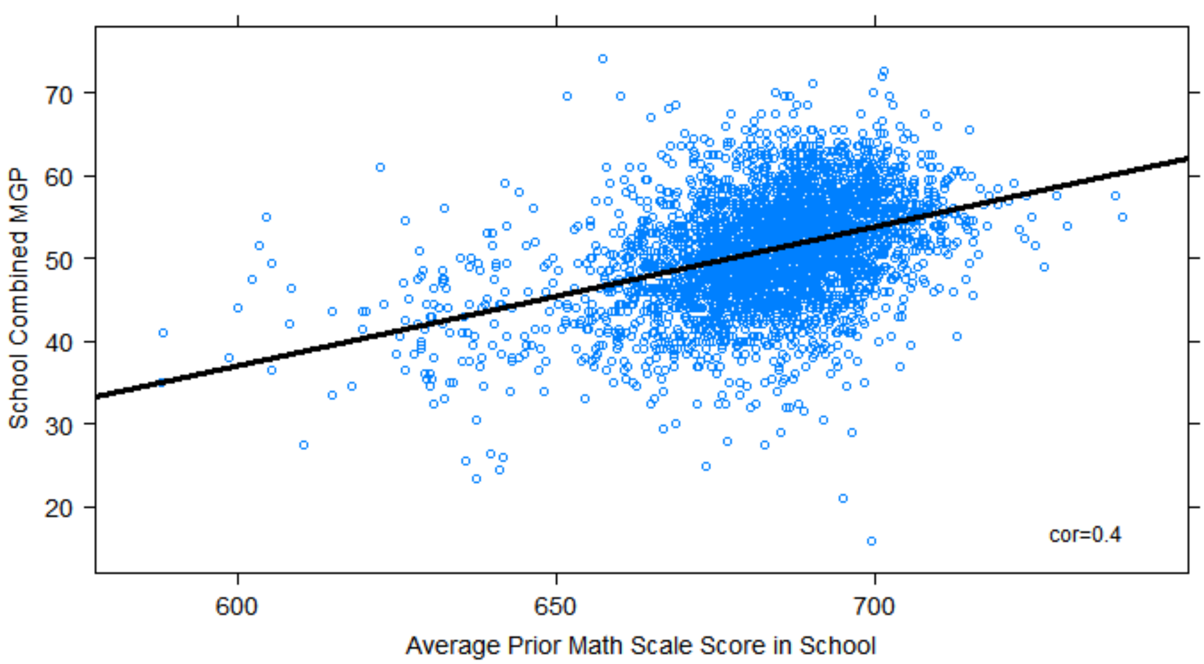


Figure 17: Relationship of Principal MGP Scores to Average Prior Mathematics Scores



CONCLUSION

The model selected to estimate growth scores for New York State provides a fair and accurate method for estimating individual teacher and principal effectiveness based on specific regulatory requirements for a "growth model" in the 2011-2012 school year. The model was selected and developed based on technical and data considerations and on the recommendations of a variety of stakeholders.

New York State plans to further refine the administrative data it collects from districts about student-teacher linkages to account for more students and teachers. It also plans to develop and seek Board of Regents approval for a "value-added" model which will allow some additional covariates to be included in the analyses and may include some other technical refinements. This may involve including additional variables at the classroom and school levels to help adjust for differences in teacher and principal outcomes not captured by student-level variables.

REFERENCES

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practices 28*(4), 42–51.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*(1), 1–26.

Greene, W. (2001). Estimating econometric models with fixed effects. Leonard N. Stern School, Finance Department, Working Paper Series 01-10. New York: New York University, Leonard N. Stern School of Business.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives 15*(4), 57–67.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics 9*, 226–252.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

Wei, Y., & Carroll, R. J. (2009). Quantile regression with measurement error. *Journal of the American Statistical Association 104*, 1129–1143.

Appendix A. Task Force and Technical Advisory Committee Members

Participant	Affiliation
Technical Advisory Committee	
Dan Goldhaber	University of Washington
Hamilton Lankford	State University of New York at Albany
Daniel F. McCaffrey	RAND
Jonah Rockoff	Columbia University
Tim R. Sass	Georgia State University
Douglas Staiger	Dartmouth College
Marty West	Harvard University
James A. Wyckoff	University of Virginia

Regents Task Force for Teacher and Principal Effectiveness			
First Name	**Last Name**	**Title**	**Office**
Milton	Cofield	Co-Chair of Task Force and Vice Chancellor	NYS Board of Regents
Lester	Young	Co-Chair of Task Force	NYS Board of Regents
Heather	Adams	Assistant in Research and Educational Services	New York State United Teachers (NYSUT)
Michele	Ashley	Principal	Jesse Owens School, PS26 in Brooklyn
Jackie	Bennett	Special Representative	United Federation of Teachers (UFT)
Vicki	Bernstein	Executive Director of Teacher Recruitment & Quality	New York City Department of Education (NYC DOE)
Jack	Bierwirth	Superintendent	Herricks Public Schools
Crystal	Boling-Barton	Vice President	New York State Federation of School Administrators
Jane	Bullowa	Assistant Superintendent for Instructional Services	Ulster BOCES
Kirsten	Busch Johnson	Executive Director, Teacher Effectiveness	New York City Department of Education (NYC DOE)
Joanna	Cannon	Executive Director of Talent Research and Data	New York City Department of Education (NYC DOE)
Amy	Carpenter	Teacher	The Captain Manuel Rivera Jr. School, PS/MS 279
Stacey	Caruso-Sharpe	Mathematics Teacher	Lynch Literacy Academy, Amsterdam City School District
Nicole	Catapano	Coordinator for Data Analysis	Washington-Saratoga-Warren-Hamilton-Essex BOCES
Florence	Chapin	Executive Director, Performance Management Strategies	New York City Department of Education (NYC DOE)
Cynthea	Cuthbert	Special Education Teacher	Monroe 2 -Orleans BOCES
Jim	Dexter	District Superintendent	Washington-Saratoga-Warren-Hamilton-Essex BOCES
Jane	Doherty	Teacher	WHEELS (Washington Heights Expeditionary Learning Outward Bound School)
Johanna	Duncan-Poitier	Chancellor's Deputy for the Education Pipeline and Vice Chancellor for Community Colleges	State University of New York (SUNY)
Mark	Fish	Principal	South Glens Falls CSD, Oliver W. Winch Middle School
Kerry	Flynn	Principal	Loudonville School
Paul	Gasparini	Principal	Jamesville-DeWitt High School

Regents Task Force for Teacher and Principal Effectiveness			
First Name	**Last Name**	**Title**	**Office**
Aminda	Gentile	Vice President	United Federation of Teachers (UFT)
Michael	Glover	District Superintendent	Genesee Valley Educational Partnership (BOCES)
Patricia	Gould	Associate Counsel	New York State School Boards Association (NYSSBA)
Janella	Hinds	Special Representative	United Federation of Teachers
Jere	Hochman	Superintendent	Bedford Central School District
William	Hughes	Music Teacher	South Orangetown Central School District
Jon	Hunter	Superintendent	Fairport Central School District
Dafny	Irizarry	ESL Teacher	Central Islip School District
Lloyd	Jaeger	Superintendent	Millbrook Central School District
Demetra	Keane	Assistant Professor	Mercy College, School of Education
Richard	Kessler	Executive Director	The Center for Arts Education
Joan	Lucariello	University Dean of Education	The City Unversity of New York (CUNY)
John	Mahony	Career and Technical Teachers	Questar III BOCES, Aviation Technology
Robert	McClure	Superintendent-in-Residence, Professional Development Team	New York State Council of School Superintendents (NYSCOSS)
Michael	McDermott	Principal	Scarsdale Middle School
Peter	McNally	Executive Vice President	Council of Supervisors and Administrators
Michael	Mendel	Secretary	United Federation of Teachers (UFT)
Paul	Mooney	Principal	Jefferson-Lewis BOCES, Howard G. Sackett Technical Center
Maria	Neira	Vice President	New York State United Teachers
Donald	Nickson	Deputy Executive Director	School Administrators Ass'n of New York State

Appendix B. Data Processing Rules and Results

Table B-1: Data Processing Rules

Description		Rationale
D.1	Exclude from analysis students with missing, invalid, or duplicate ID numbers after matching procedures have been implemented	Not possible to merge student records over time
D.2	Exclude from analysis all students with no current achievement scores or standard error of measurement (SEM) for those scores in grades and subjects used in growth analysis	Not possible to generate growth scores without current achievement
D.3	Exclude from analysis all students with no prior achievement scores in grades and subjects used in growth analysis	Not possible to generate growth scores without any prior achievement
D.3b	Exclude from analysis all students with no immediate prior achievement scores or standard error of measurement (SEM) for those scores in grades and subjects used in growth analysis	NYSED decision
D.4	Include in analysis students with missing demographic data on variables included in models	Possible to estimate models with missing data employing missing flags although too many missing flags will bias coefficients
D.5	Exclude from analysis students with an atypical grade progression, including students with multiple records in current or prior year(s) with contradictory grades (e.g., enrolled in one school as grade 5 and in another as grade 6; tested in one year as grade 5 and in the following year as grade 4).[8] Exception: retained students, see D.6b	Not possible to validate which grade is correct. Also addresses rules regarding students who repeat or skip a grade
D.5b	Retained students—regardless of year retained—are assumed to have an atypical grade progression. Students retained in the most recent year are excluded according to D.6. However, include other students who have a test history with *some* of the appropriate grade progression. For example, if a grade 8 student has prior scores for grades 6 three	

[8] Note: If multiple years of prior scores used, exclude years with contradictory grades, e.g., if two years of prior achievement used and contradictory grades in only one year, can still include the other prior score/grade and keep student in analysis.

Description		Rationale
	years ago, 6 two years ago, and 7 last year, keep the second grade 6 score and the grade 7 score (setting the first grade 6 score to 0 and the missing indicator for this variable to 0). Similar rules apply to students who skip a grade. Make sure that the right score is kept for the proper year (i.e., grade 6 score must be kept from when they should have been in grade 6 to ensure that variables are added to the model with the appropriate time lag).	
D.6	Exclude from analysis all students with multiple records in current or prior year(s) with contradictory test scores[9]	Not possible to validate which score is correct
D.7	Exclude from analysis all students with only **invalidated**, or "did not attempt," or out-of-range test scores in current or prior years[10]	Valid test scores needed for value-added model

[9] Note: If multiple years of prior scores used, exclude years with contradictory grades, e.g., if two years of prior achievement used and contradictory grades in only one year, can still include the other prior score/grade and keep student in analysis.

[10] Note: If multiple years of prior scores used, exclude years with invalidated or "did not attempt" or out-of-range scores, e.g., if two years of prior achievement used and these scores exist in only one year, can still include the other prior score and keep student in analysis.

Table B-2: Math Data Processing

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
2008-09 Score File Processing				
Number of Records in the input ScoreFile	All	2008-09	2,531,671	--
Number of Records in the Score File After keeping only the records for the current subject	All	2008-09	1,271,409	1,260,262
Number of Records in the Score File after deleting blank/Invalid SSID	All	2008-09	1,271,409	0
Number of Records in the Score File after deleting Out of Range Scores	All	2008-09	1,270,385	1,024
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2008-09	1,270,385	0
Number of Records in the Score File after deleting all duplicates by SSID	All	2008-09	1,269,921	464
Number of Records in the Score File after deleting records with missing SEM	All	2008-09	1,269,921	0
2009-10 Score File Processing				
Number of Records in the input ScoreFile	All	2009-10	2,540,947	--
Number of Records in the Score File after keeping only the records for the current subject	All	2009-10	1,277,724	1,263,223
Number of Records in the Score File after deleting blank/Invalid SSID	All	2009-10	1,277,724	0
Number of Records in the Score File after deleting Out of Range Scores	All	2009-10	1,276,339	1,385
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2009-10	1,276,337	2
Number of Records in the Score File after deleting all duplicates by SSID	All	2009-10	1,275,858	479
Number of Records in the Score File after deleting records with missing SEM	All	2009-10	1,275,858	0
2010-11 Score File Processing				
Number of Records in the input ScoreFile	All	2010-11	2,540,183	--
Number of Records in the Score File after keeping only the records for the current subject	All	2010-11	1,276,374	1,263,809
Number of Records in the Score File after deleting blank/Invalid SSID	All	2010-11	1,276,374	0
Number of Records in the Score File after deleting Out of Range Scores	All	2010-11	1,275,437	937

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2010-11	1,275,435	2
Number of Records in the Score File after deleting all duplicates by SSID	All	2010-11	1,274,949	486
Number of Records in the Score File after deleting records with missing SEM	All	2010-11	1,274,948	1
2011-12 Score File Processing				
Number of Records in the input ScoreFile	All	2011-12	2,529,590	--
Number of Records in the Score File after keeping only the records for the current subject	All	2011-12	1,270,328	1,259,262
Number of Records in the Score File after deleting blank/Invalid SSID	All	2011-12	1,270,328	0
Number of Records in the Score File after deleting Out of Range Scores	All	2011-12	1,269,199	1,129
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2011-12	1,269,199	0
Number of Records in the Score File after deleting all duplicates by SSID	All	2011-12	1,268,751	448
Number of Records in the Score File after deleting records with missing SEM	All	2011-12	1,268,751	0
Number of Records in the Score File after deleting grade3 records	All	2011-12	1,062,404	206,347
Number of Records in the data for a specific grade	4	2011-12	214,738	--
Number of Records in the data for a specific grade	5	2011-12	205,138	--
Number of Records in the data for a specific grade	6	2011-12	217,861	--
Number of Records in the data for a specific grade	7	2011-12	205,994	--
Number of Records in the data for a specific grade	8	2011-12	218,673	--
Merging 2011-12 Score File with Prior Year Score Files				
Number of Records in the Merged Score File after deleting records with prior grade greater than current grade	All	2011-12	1,062,404	--
Number of Records in the Merged Score File after deleting records without immediate prior score	All	2011-12	978,861	83,543
Number of Records for a specific grade after merging 3 years of data	4	2011-12	193,498	--
Number of Records for a specific grade after merging 3 years of data	5	2011-12	196,354	--
Number of Records for a specific grade after merging 3 years of data	6	2011-12	197,835	--

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
Number of Records for a specific grade after merging 3 years of data	7	2011-12	194,955	--
Number of Records for a specific grade after merging 3 years of data	8	2011-12	196,219	--
Teacher-Course File Processing				
Number of Records in input TeacherCourseFile	ALL	2011-12	22,792,100	--
Number of Records in TeacherCourseFile after subsetting for subject records	ALL	2011-12	3,790,457	19,001,643
Number of Records in TeacherCourseFile after keeping the record with latest reporting date	ALL	2011-12	2,776,793	1,013,664
Number of Records in TeacherCourseFile after keeping the record that meet continous enrollment	ALL	2011-12	2,121,073	655,720
Number of Records in TeacherCourseFile after keeping unique classes	ALL	2011-12	1,285,491	835,582
Merging Teacher-Course File to Merged Student Test Score Files*				
Number of Records in StudentTeacherCourseFile	ALL	2011-12	1,452,588	--
Number of Records in StudentTeacherCourseFile after deleting records with courses but no scores	ALL	2011-12	1,064,005	388,583
Number of Records in StudentTeacherCourseFile after keeping only 6 records per student	ALL	2011-12	1,064,001	4
Description of Final Reference File Used For Analysis*				
Number of Records in the Reference File	ALL	2011-12	978,861	--
Number of Records in the Reference File for a specific grade	4	2011-12	193,498	--
Number of Records in the Reference File for a specific grade	5	2011-12	196,354	--
Number of Records in the Reference File for a specific grade	6	2011-12	197,835	--
Number of Records in the Reference File for a specific grade	7	2011-12	194,955	--
Number of Records in the Reference File for a specific grade	8	2011-12	196,219	--

*Note that the number of observations shown in the merging teacher-course and student test score file represents the number of unique teacher-course-student links, while the number of observations in the final reference file description represents the number of students. That is, the difference is due to the fact that a student may be linked to multiple teachers/courses. The file is transformed for statistical analysis to include one record per student.

Table B-3: ELA Data Processing

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
2008-09 Score File Processing				
Number of Records in the input ScoreFile	All	2008-09	2,531,671	--
Number of Records in the Score File After keeping only the records for the current subject	All	2008-09	1,260,262	1,271,409
Number of Records in the Score File after deleting blank/Invalid SSID	All	2008-09	1,260,262	0
Number of Records in the Score File after deleting Out of Range Scores	All	2008-09	1,259,423	839
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2008-09	1,259,419	4
Number of Records in the Score File after deleting all duplicates by SSID	All	2008-09	1,258,913	506
Number of Records in the Score File after deleting records with missing SEM	All	2008-09	1,258,912	1
2009-10 Score File Processing				
Number of Records in the input ScoreFile	All	2009-10	2,540,947	--
Number of Records in the Score File After keeping only the records for the current subject	All	2009-10	1,263,223	1,277,724
Number of Records in the Score File after deleting blank/Invalid SSID	All	2009-10	1,263,223	0
Number of Records in the Score File after deleting Out of Range Scores	All	2009-10	1,262,133	1,090
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2009-10	1,262,132	1
Number of Records in the Score File after deleting all duplicates by SSID	All	2009-10	1,261,659	473
Number of Records in the Score File after deleting records with missing SEM	All	2009-10	1,261,659	0
2010-11 Score File Processing				
Number of Records in the input ScoreFile	All	2010-11	2,540,947	--
Number of Records in the Score File After keeping only the records for the current subject	All	2010-11	1,277,724	1,263,223
3:Number of Records in the Score File after deleting blank/Invalid SSID	All	2010-11	1,277,724	0
Number of Records in the Score File after deleting Out of Range Scores	All	2010-11	1,276,339	1,385

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2010-11	1,276,336	3
Number of Records in the Score File after deleting all duplicates by SSID	All	2010-11	1,275,807	529
Number of Records in the Score File after deleting records with missing SEM	All	2010-11	1,275,807	0
2011-12 Score File Processing				
Number of Records in the input ScoreFile	All	2011-12	2,540,183	-
Number of Records in the Score File After keeping only the records for the current subject	All	2011-12	1,263,809	1,276,374
Number of Records in the Score File after deleting blank/Invalid SSID	All	2011-12	1,263,809	0
Number of Records in the Score File after deleting Out of Range Scores	All	2011-12	1,262,568	1,241
Number of Records in the Score File after deleting one of the duplicates by SSID Item_Description ScaleScore School_ID District_ID S_EnrolledGrade S_Lastname S_FirstName	All	2011-12	1,262,566	2
Number of Records in the Score File after deleting all duplicates by SSID	All	2011-12	1,262,094	472
Number of Records in the Score File after deleting records with missing SEM	All	2011-12	1,262,093	1
Number of Records in the Score File after deleting grade3 records	All	2011-12	1,053,087	--
Number of Records in the data for a specific grade	4	2011-12	212,888	--
Number of Records in the data for a specific grade	5	2011-12	203,138	--
Number of Records in the data for a specific grade	6	2011-12	215,910	--
Number of Records in the data for a specific grade	7	2011-12	204,102	--
Number of Records in the data for a specific grade	8	2011-12	217,049	--
Merging 2011-12 Score File with Prior Year Score Files				
Number of Records in the Merged Score File after deleting records with prior grade greater than current grade	All	2011-12	1,053,087	--
Number of Records in the Merged Score File after deleting records without immediated prior score	All	2011-12	970,050	83,037
Number of Records in the for a specific grade after merging 3 years of data	4	2011-12	191,708	--
Number of Records in the for a specific grade after merging 3 years of data	5	2011-12	194,579	--
Number of Records in the for a specific grade after merging 3 years of data	6	2011-12	196,011	--

Data Processing Description	Grade	Year	Resulting # Obs After Exclusion	# Records excluded
Number of Records in the for a specific grade after merging 3 years of data	7	2011-12	193,207	--
Number of Records in the for a specific grade after merging 3 years of data	8	2011-12	194,545	--
Teacher-Course File Processing				
Number of Records in input TeacherCourseFile	ALL	2011-12	22,792,100	--
Number of Records in TeacherCourseFile after subsetting for subject records	ALL	2011-12	2,555,495	20,236,605
Number of Records in TeacherCourseFile after keeping the record with latest reporting date	ALL	2011-12	1,950,434	605,061
Number of Records in TeacherCourseFile after keeping the record that meet continous enrollment	ALL	2011-12	1,618,011	332,423
Number of Records in TeacherCourseFile after keeping unique classes	ALL	2011-12	1,285,491	835,582
Merging Teacher-Course File to Merged Student Test Score Files*				
Number of Records in StudentTeacherCourseFile	ALL	2011-12	1,144,399	--
Number of Records in StudentTeacherCourseFile after deleting records with courses but no scores	ALL	2011-12	1,097,101	47,298
Number of Records in StudentTeacherCourseFile after keeping only 6 records per student	ALL	2011-12	1,097,080	21
Description of Final Reference File Used For Analysis*				
Number of Records in the Reference File	ALL	2011-12	970,050	--
Number of Records in the Reference File for a specific grade	4	2011-12	191,708	--
Number of Records in the Reference File for a specific grade	5	2011-12	194,579	--
Number of Records in the Reference File for a specific grade	6	2011-12	196,011	--
Number of Records in the Reference File for a specific grade	7	2011-12	193,207	--
Number of Records in the Reference File for a specific grade	8	2011-12	194,545	--

*Note that the number of observations shown in the merging teacher-course and student test score file represents the number of unique teacher-course-student links, while the number of observations in the final reference file description represents the number of students. That is, the difference is due to the fact that a student may be linked to multiple teachers/courses. The file is transformed for statistical analysis to include one record per student.

Appendix C. Item Descriptions Used in Analysis

Table C-1: Relevant Item Descriptions

Item Description
Grade 3 ELA
Grade 3 Math
Grade 4 ELA
Grade 4 Math
Grade 5 ELA
Grade 5 Math
Grade 6 ELA
Grade 6 Math
Grade 7 ELA
Grade 7 Math
Grade 8 ELA
Grade 8 Math

Appendix D. Attribution Rules

Teacher attribution is is done when a student has "continuous enrollment linkage." Table D-1 describes the system by which continuous enrollment linkage is generated, using a teacher/student/course file.

Table D-1: Teacher Attribution Rules

Attribution Rule	
A.1	Enrollment must be in one of the classes listed in Appendix C.
A.2	Duplicate records (one sharing a description, District ID, School ID, teacher ID, SSID, section number, course number, start date, end date) keeping only the latest reporting date.
A.3	Because the data file may contain several entries for a district/school/student/teacher/course number, using all records that link a student and a district/school/teacher, generate a list of unique school year days that fall after the start of the school year and before the test date.
A.4	When the total number of days is at least as large as 195 (ELA) or 203 (Math), then there is a link.

Principal enrollment is handled entirely with the "School enrollment flag." When this is "yes" then a student is linked to the principal at the school on the testing file. When this is not yes, then the student is not linked to a principal.

Appendix E. Demographic Variable Definitions

New York has provided a Y/N variable to identify students as SWD, ELL, and ED.

The poverty indicator is a Y/N variable that was provided as is by NYSED. It is used to determine which cohort members should be included in the economically disadvantaged group for district and school accountability. An economically disadvantaged student is a student who participates in, or whose family participates in, economic assistance programs such as the following:

- The Free- or Reduced-price Lunch Programs (Note that the United States Department of Agriculture has authorized the use of enrollment in free- and reduced-price lunch programs to identify students from low-income families for Title I reporting purposes.) Please consult the NYSED's Office of Child Nutrition Program Administration for guidelines.
- Social Security Insurance (SSI)
- Food Stamps
- Foster Care
- Refugee Assistance (cash or medical assistance)
- Earned Income Tax Credit (EITC)
- Home Energy Assistance Program (HEAP)
- Safety Net Assistance (SNA)
- Bureau of Indian Affairs (BIA)
- Family Assistance: Temporary Assistance for Needy Families (TANF)

If one student in a family is identified as low income, all students from that household (economic unit) may be identified as low income.

Appendix F. Model Variance Components

For each grade and subject, the growth models were fit to the data with both teacher and school random effects. The model decomposes total variation in the outcome into three unique components: variance between teachers within a school, variance between schools, and variance between students within a class. Table F-1 provides the standard deviation of the components for all subjects, grades, and models. Figures F-1 through F-2 provide a graphical look at these same data, with bar charts corresponding to a subject and model combination, shown across grades 4 through 8.

Table F-1. Magnitude of Teacher and School Effect Standard Deviations

Subject	Grade	Root Mean Square Error	Teacher	School
ELA	4	12.8	3.7	4.0
	5	6.7	2.3	2.4
	6	4.1	1.6	2.0
	7	2.7	1.9	2.2
	8	7.6	1.4	1.9
Math	4	14.5	6.1	5.7
	5	11.4	6.6	5.9
	6	12.0	6.2	6.6
	7	6.1	4.7	4.2
	8	11.1	4.7	4.5

Figure F-1. Magnitude of Teacher and School Effects, ELA Adjusted Models by Grade

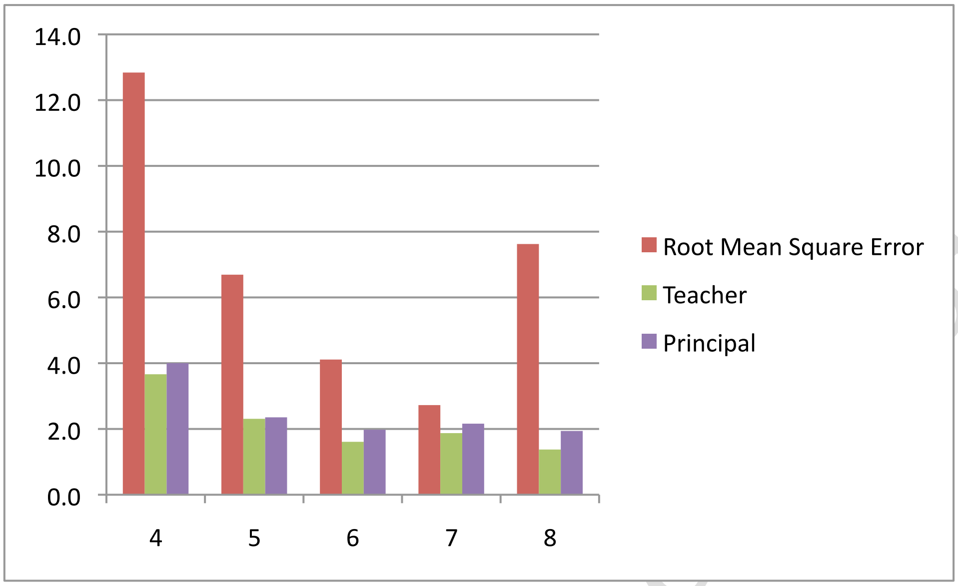
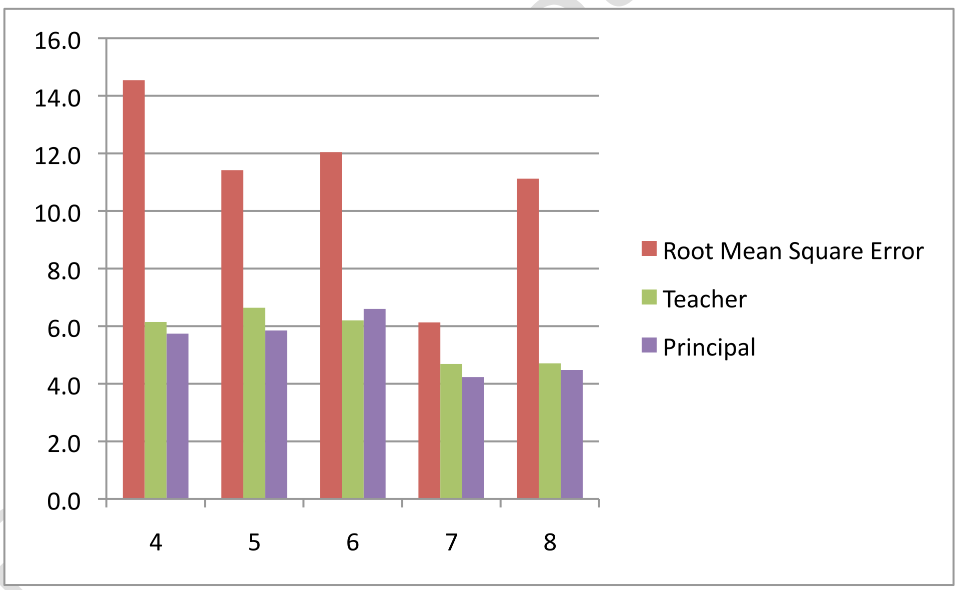


Figure F-2. Magnitude of Teacher and School Effects, Math Adjusted Models by Grade



Appendix G. Model Coefficients

Table G-1. Grade 4 ELA Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	-170.415	2.447	0.000
Student SWD status	-3.875	0.133	0.000
Student ELL status	-0.197	0.176	0.265
Student Poverty status	-1.768	0.106	0.000
2010-11 Assessment Score	1.274	0.004	0.000

Table G-2. Grade 5 ELA Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	263.166	1.046	0.000
Student SWD status	-2.116	0.073	0.000
Student ELL status	0.004	0.107	0.970
Student Poverty status	-0.850	0.059	0.000
Missing Indicator: 2009-10 Assessment Score	36.888	1.251	0.000
2009-10 Assessment Score	0.055	0.002	0.000
2010-11 Assessment Score	0.550	0.002	0.000

Table G-3. Grade 6 ELA Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	163.597	1.321	0.000
Student SWD status	-1.125	0.054	0.000
Student ELL status	-0.432	0.087	0.000
Student Poverty status	-0.802	0.043	0.000
Missing Indicator: 2008-09 Assessment Score	28.031	1.166	0.000
2008-09 Assessment Score	0.041	0.002	0.000
Missing Indicator: 2009-10 Assessment Score	74.383	1.764	0.000
2009-10 Assessment Score	0.113	0.003	0.000
2010-11 Assessment Score	0.593	0.004	0.000

Table G-4. Grade 7 ELA Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	92.228	1.354	0.000
Student SWD status	-1.818	0.050	0.000
Student ELL status	0.853	0.084	0.000
Student Poverty status	-0.408	0.040	0.000
Missing Indicator: 2008-09 Assessment Score	44.423	1.003	0.000
2008-09 Assessment Score	0.067	0.002	0.000
Missing Indicator: 2009-10 Assessment Score	26.069	0.956	0.000
2009-10 Assessment Score	0.040	0.001	0.000
2010-11 Assessment Score	0.756	0.003	0.000

Table G-5. Grade 8 ELA Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	4.389	1.795	0.014
Student SWD status	-2.719	0.080	0.000
Student ELL status	-0.428	0.140	0.002
Student Poverty status	-0.287	0.064	0.000
Missing Indicator: 2008-09 Assessment Score	26.861	1.701	0.000
2008-09 Assessment Score	0.040	0.003	0.000
Missing Indicator: 2009-10 Assessment Score	30.647	1.764	0.000
2009-10 Assessment Score	0.048	0.003	0.000
2010-11 Assessment Score	0.896	0.005	0.000

Table G-6. Grade 4 Mathematics Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	-288.532	2.672	0.000
Student SWD status	-5.426	0.143	0.000
Student ELL status	0.721	0.191	0.000
Student Poverty status	-1.379	0.119	0.000
2010-11 Assessment Score	1.425	0.004	0.000

Table G-7. Grade 5 Mathematics Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	82.289	1.452	0.000
Student SWD status	-2.837	0.120	0.000
Student ELL status	1.758	0.171	0.000
Student Poverty status	-1.070	0.098	0.000
Missing Indicator: 2009-10 Assessment Score	65.823	1.884	0.000
2009-10 Assessment Score	0.094	0.003	0.000
2010-11 Assessment Score	0.785	0.003	0.000

Table G-8. Grade 6 Mathematics Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	81.303	1.544	0.000
Student SWD status	-3.774	0.120	0.000
Student ELL status	2.229	0.191	0.000
Student Poverty status	-1.349	0.094	0.000
Missing Indicator: 2008-09 Assessment Score	30.588	2.112	0.000
2008-09 Assessment Score	0.042	0.003	0.000
Missing Indicator: 2009-10 Assessment Score	93.690	2.639	0.000
2009-10 Assessment Score	0.138	0.004	0.000
2010-11 Assessment Score	0.696	0.005	0.000

Table G-9. Grade 7 Mathematics Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	84.315	1.101	0.000
Student SWD status	-1.856	0.085	0.000
Student ELL status	1.470	0.140	0.000
Student Poverty status	-0.788	0.065	0.000
Missing Indicator: 2008-09 Assessment Score	35.235	1.545	0.000
2008-09 Assessment Score	0.051	0.002	0.000
Missing Indicator: 2009-10 Assessment Score	83.396	2.408	0.000
2009-10 Assessment Score	0.123	0.004	0.000
2010-11 Assessment Score	0.697	0.003	0.000

Table G-10. Grade 8 Mathematics Model Coefficients, Adjusted Model

Effect Name	Effect	Standard Error	p-value
Constant Term	55.340	1.531	0.000
Student SWD status	-0.971	0.115	0.000
Student ELL status	5.458	0.190	0.000
Student Poverty status	0.361	0.087	0.000
Missing Indicator: 2008-09 Assessment Score	10.130	2.786	0.000
2008-09 Assessment Score	0.013	0.004	0.002
Missing Indicator: 2009-10 Assessment Score	56.419	3.968	0.000
2009-10 Assessment Score	0.084	0.006	0.000
2010-11 Assessment Score	0.820	0.006	0.000

Appendix H. Impact Charts by Grade and Subject

Figure H-1 Teacher MGP by Percent ELL-ELA Grade 4

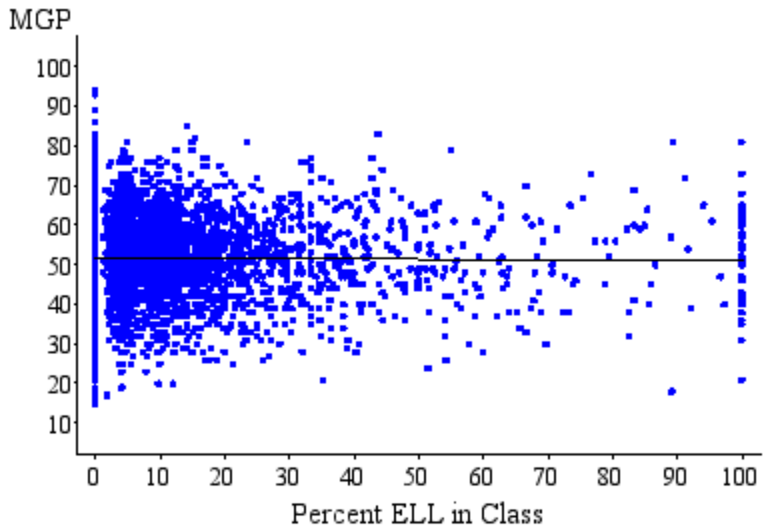


Figure H-2 Teacher MGP by Percent SWD-ELA Grade 4

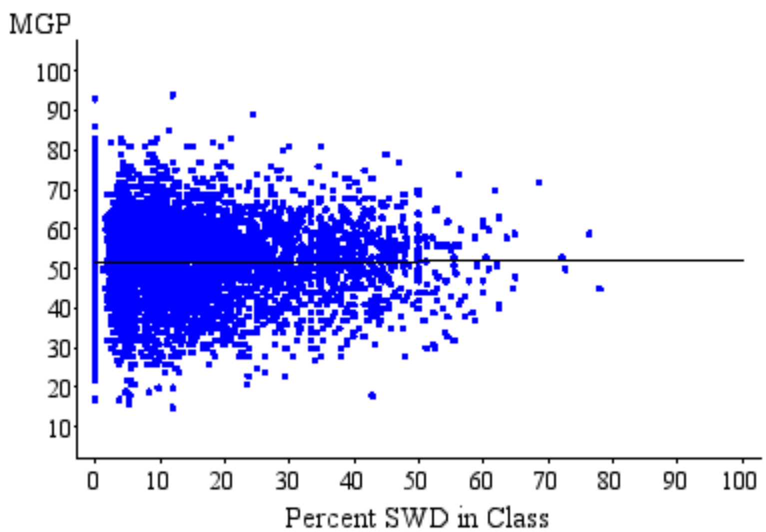


Figure H-3 Teacher MGP by Percent Economically Disadvantaged-ELA Grade 4

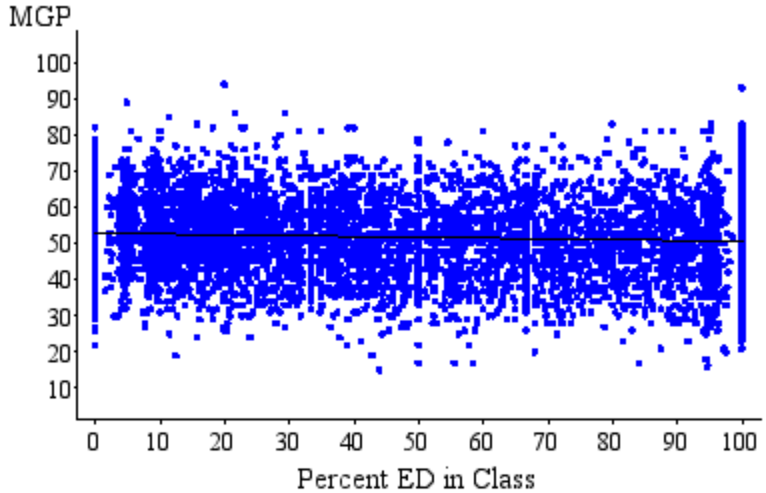


Figure H-4 Teacher MGP by Percent Nonwhite-ELA Grade 4

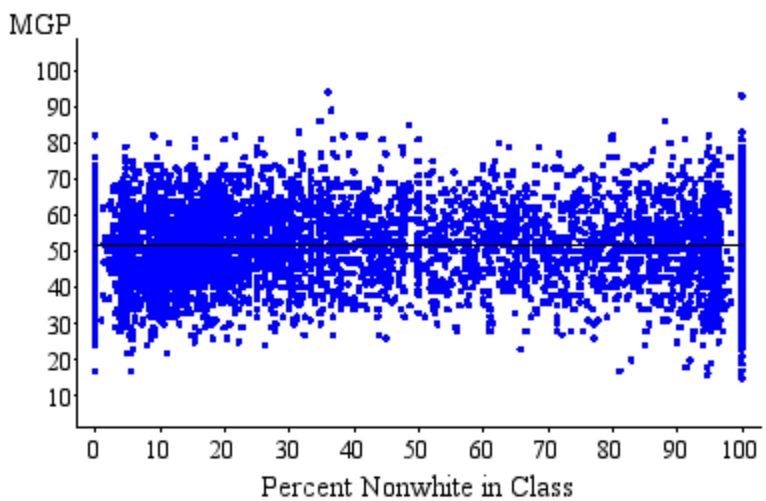


Figure H-5 Teacher MGP by Students' Mean Prior Scale Score-ELA Grade 4

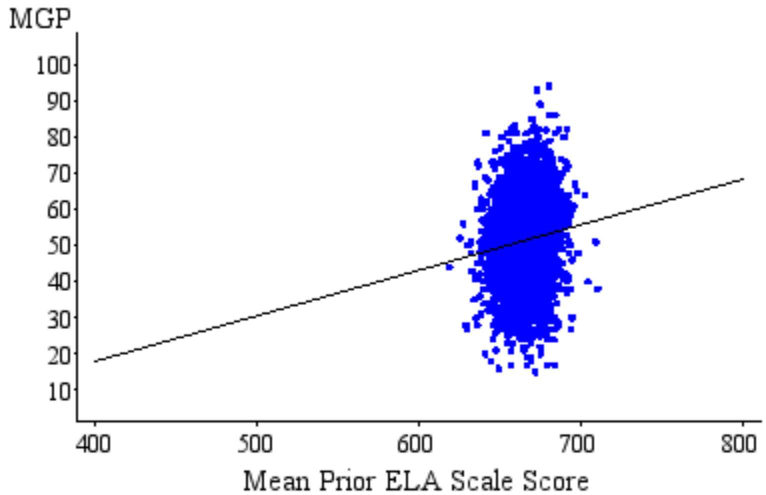


Figure H-6 Teacher MGP by Percent ELL-Math Grade 4

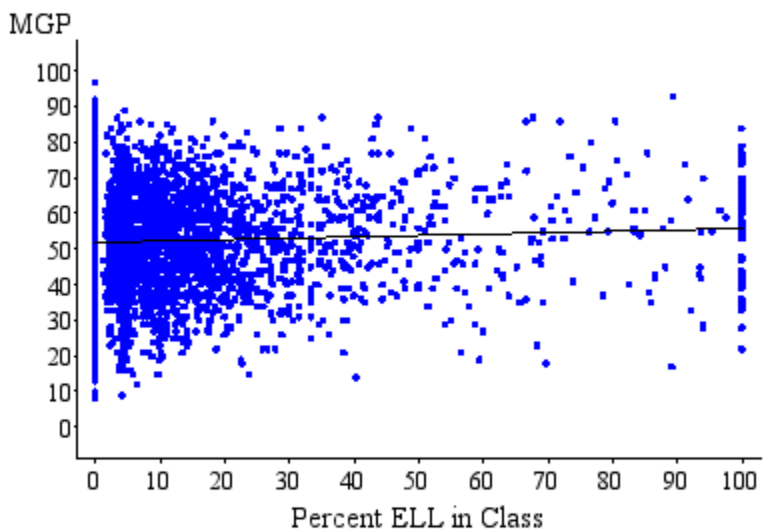


Figure H-7 Teacher MGP by Percent SWD-Math Grade 4

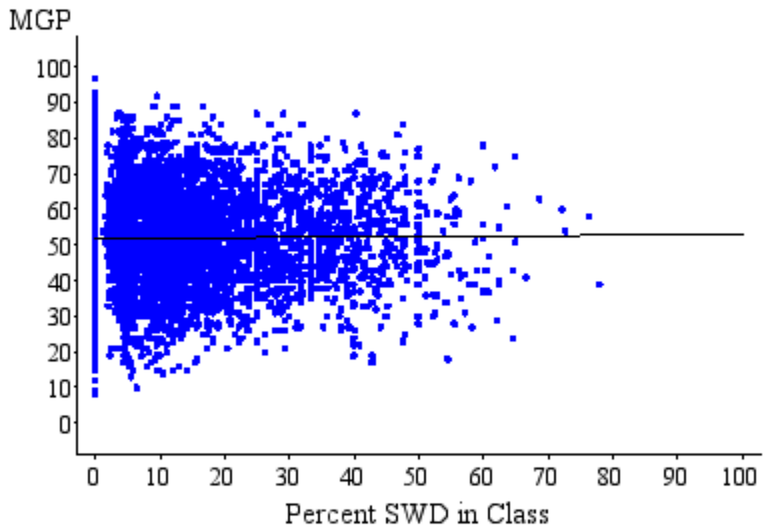


Figure H-8 Teacher MGP by Percent Economically Disadvantaged-Math Grade 4

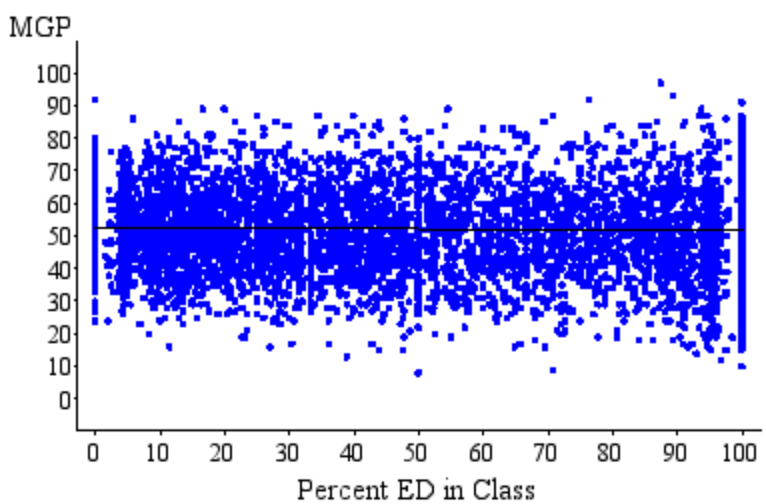


Figure H-9 Teacher MGP by Percent Nonwhite-Math Grade 4

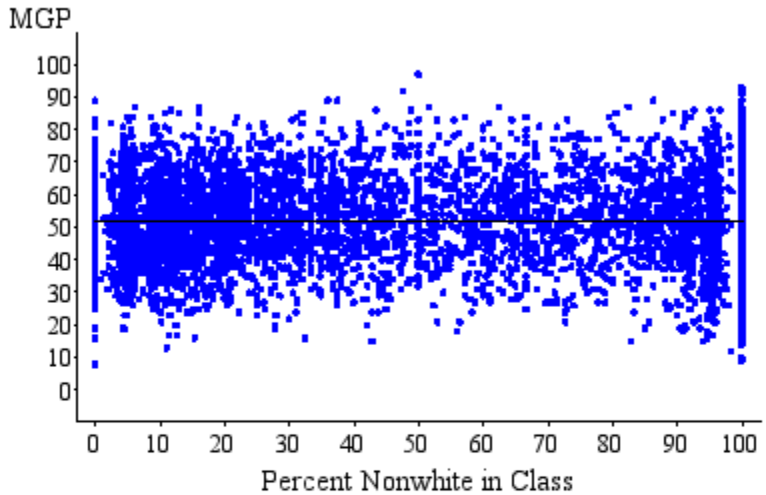


Figure H-10 Teacher MGP by Students' Mean Prior Scale Score-Math Grade 4

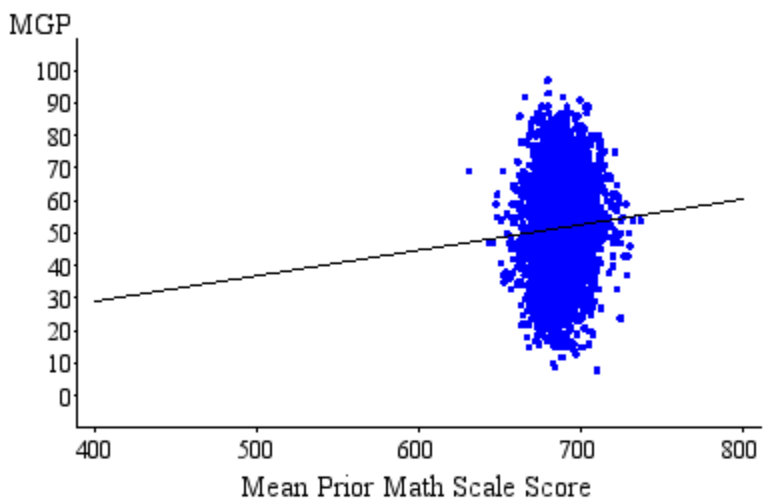


Figure H-11 Teacher MGP by Percent ELL-ELA Grade 5

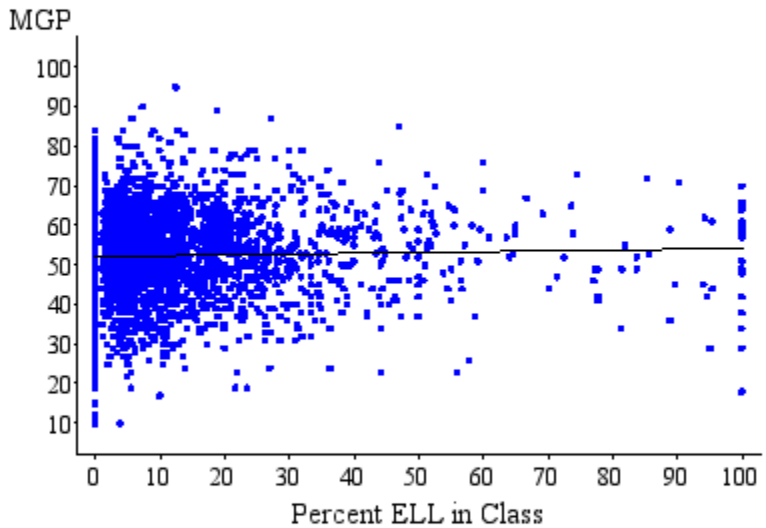


Figure H-12 Teacher MGP by Percent SWD-ELA Grade 5

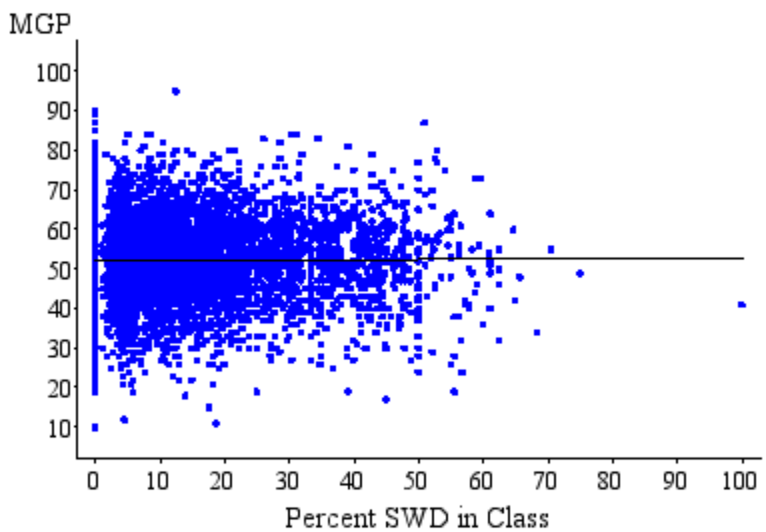


Figure H-13 Teacher MGP by Percent Economically Disadvantaged-ELA Grade 5

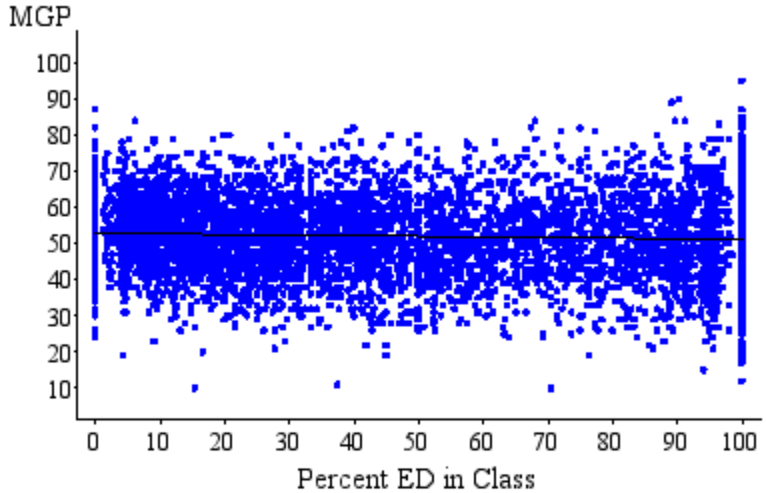


Figure H-14 Teacher MGP by Percent Nonwhite-ELA Grade 5

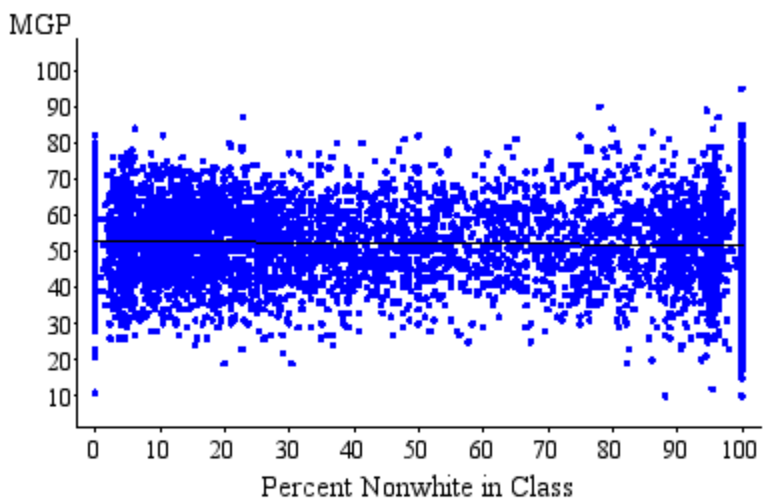


Figure H-15 Teacher MGP by Students' Mean Prior Scale Score-ELA Grade 5

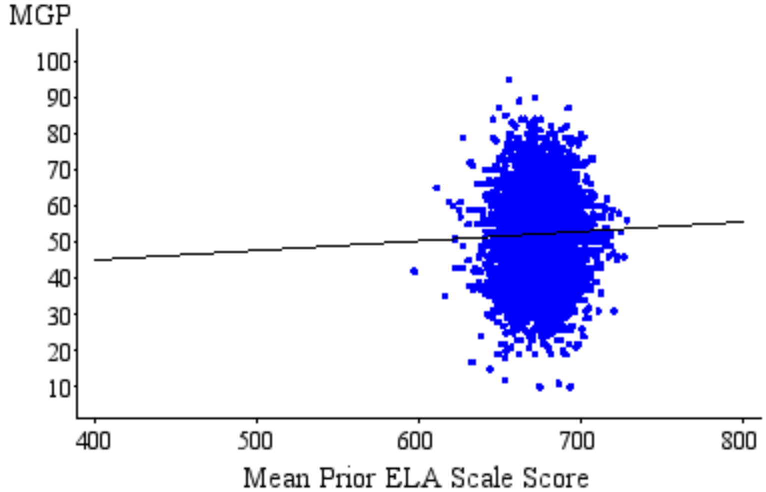


Figure H-16 Teacher MGP by Percent ELL-Math Grade 5

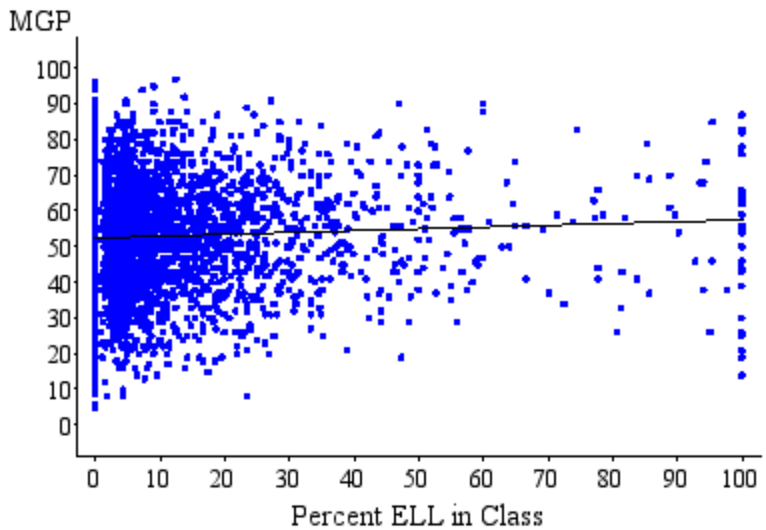


Figure H-17 Teacher MGP by Percent SWD-Math Grade 5

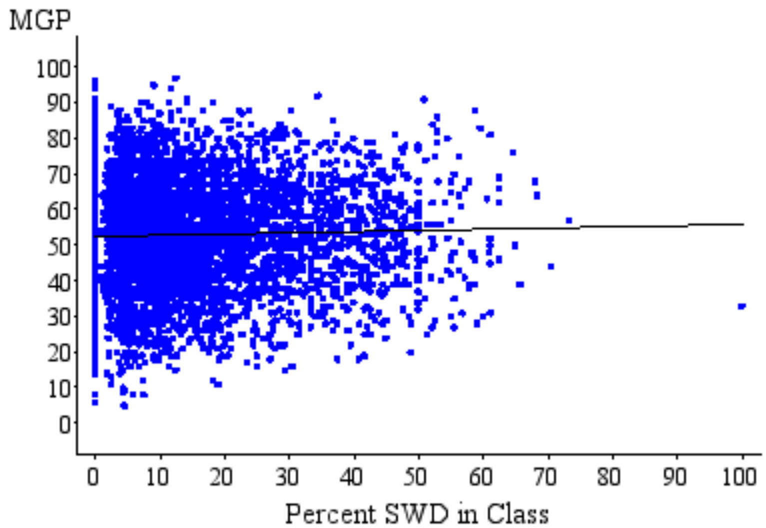


Figure H-18 Teacher MGP by Percent Economically Disadvantaged-Math Grade 5

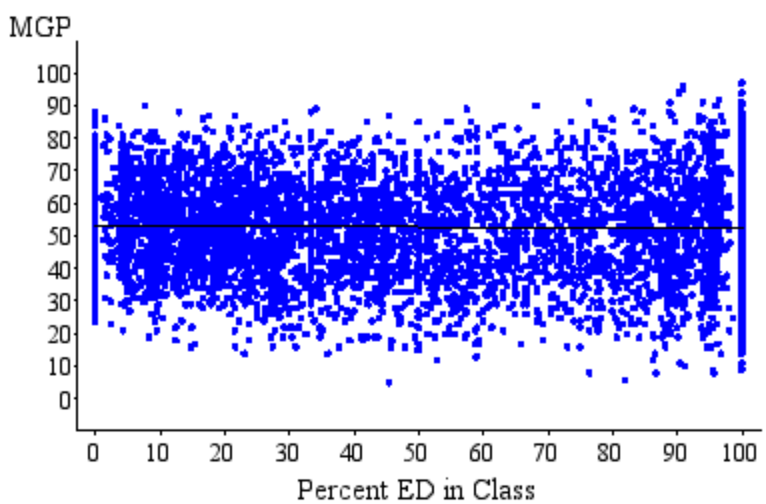


Figure H-19 Teacher MGP by Percent Nonwhite-Math Grade 5

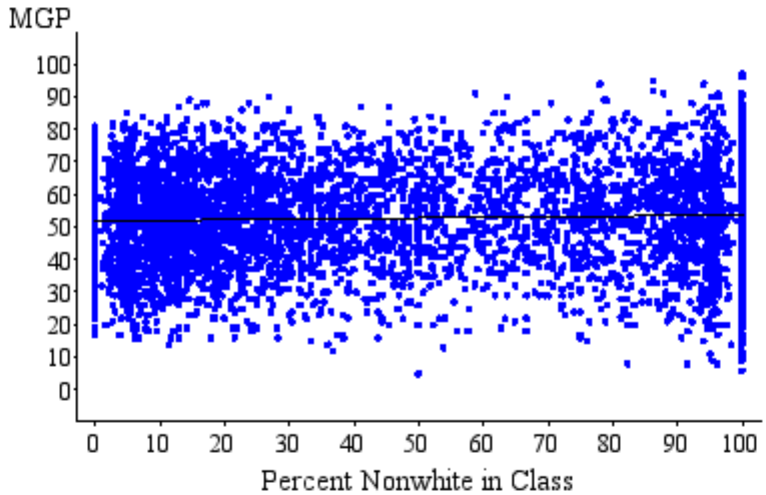


Figure H-20 Teacher MGP by Students' Mean Prior Scale Score-Math Grade 5

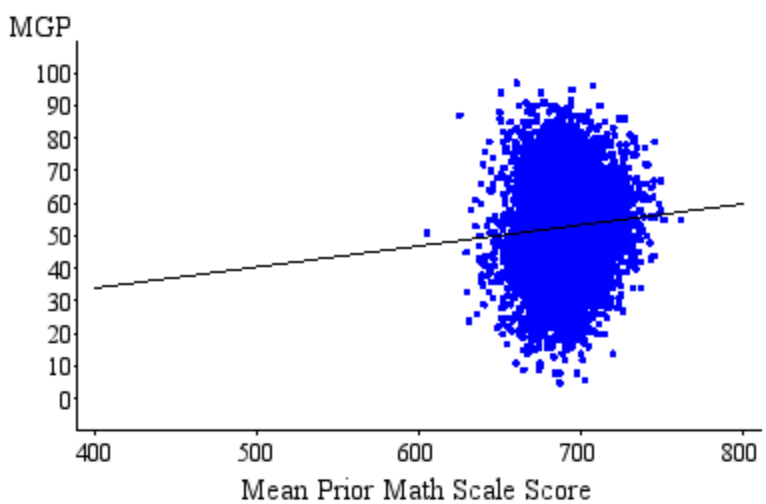


Figure H-21 Teacher MGP by Percent ELL-ELA Grade 6

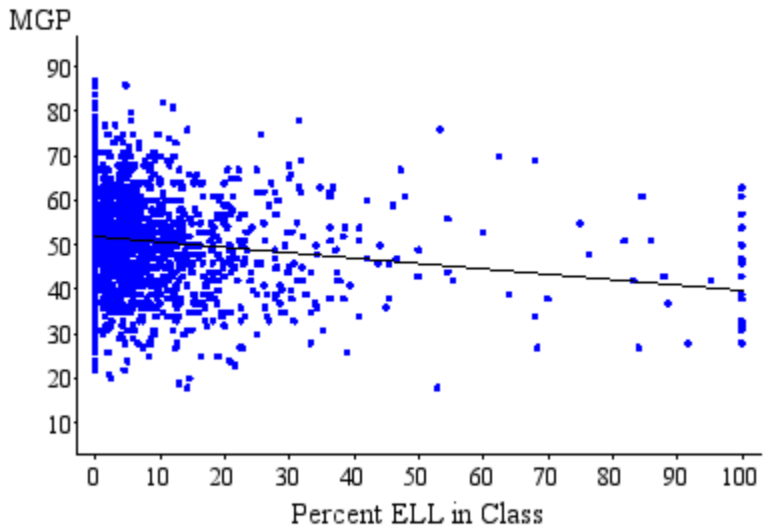


Figure H-22 Teacher MGP by Percent SWD-ELA Grade 6

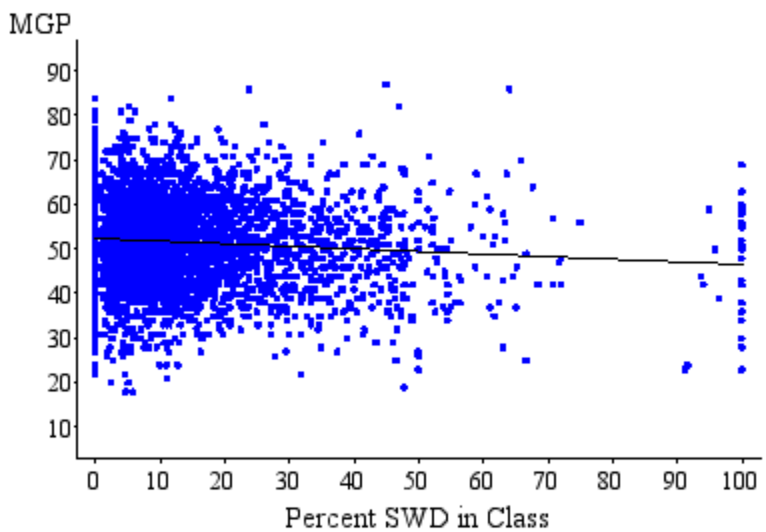


Figure H-23 Teacher MGP by Percent Economically Disadvantaged-ELA Grade 6

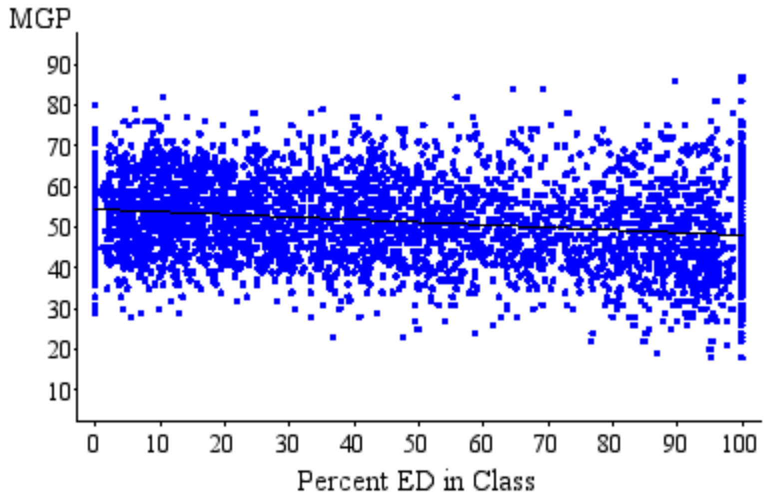


Figure H-24 Teacher MGP by Percent Nonwhite-ELA Grade 6

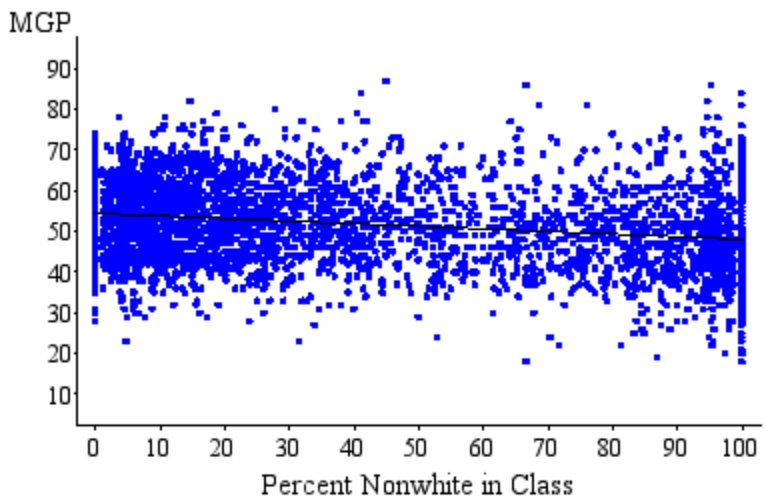


Figure H-25 Teacher MGP by Students' Mean Prior Scale Score-ELA Grade 6

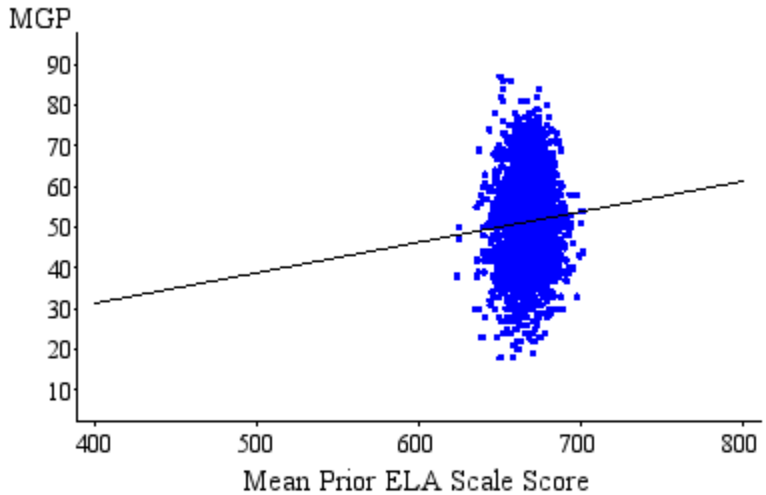


Figure H-26 Teacher MGP by Percent ELL-Math Grade 6

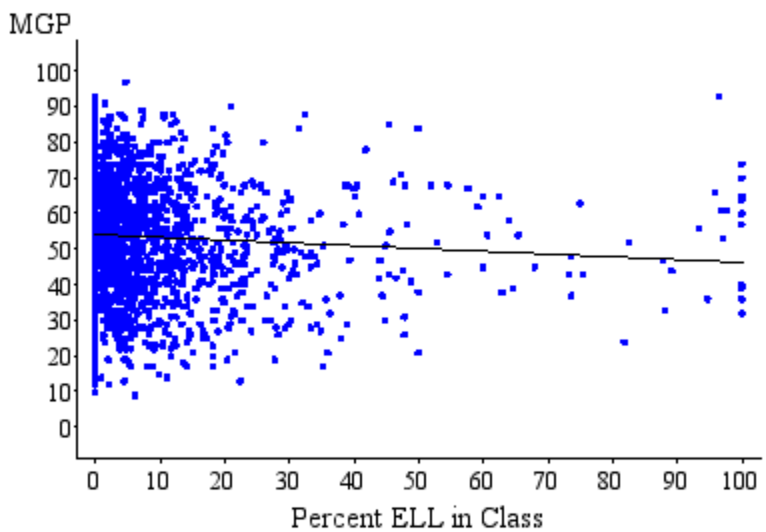


Figure H-27 Teacher MGP by Percent SWD-Math Grade 6

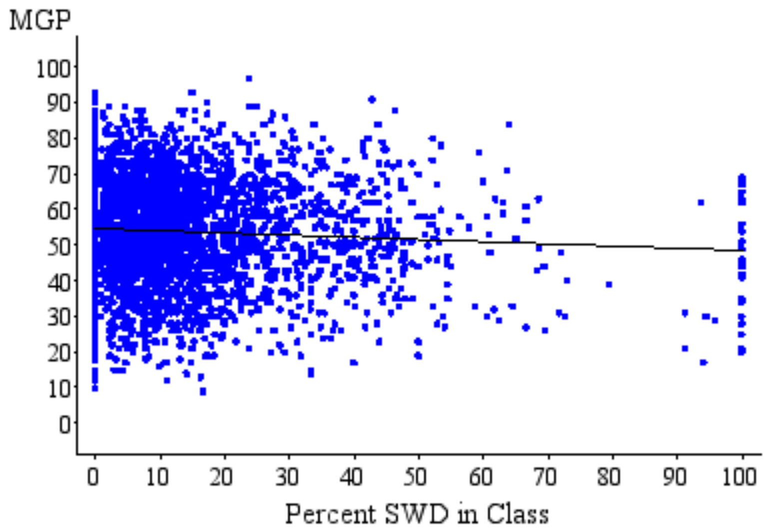


Figure H-28 Teacher MGP by Percent Economically Disadvantaged-Math Grade 6

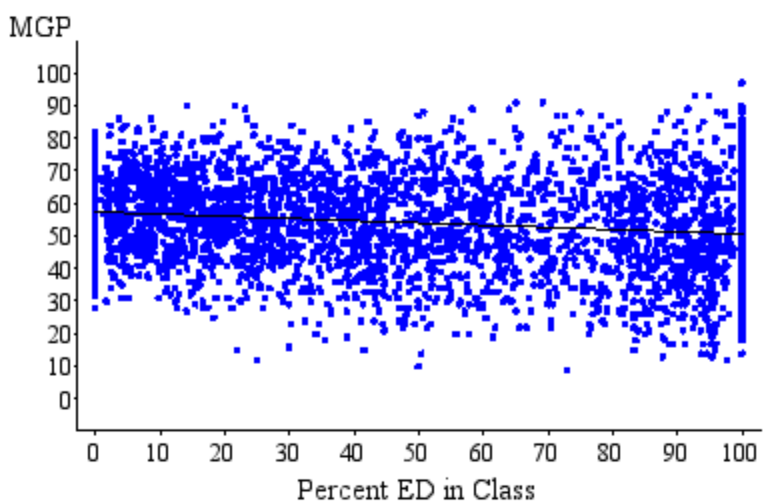


Figure H-29 Teacher MGP by Percent Nonwhite-Math Grade 6

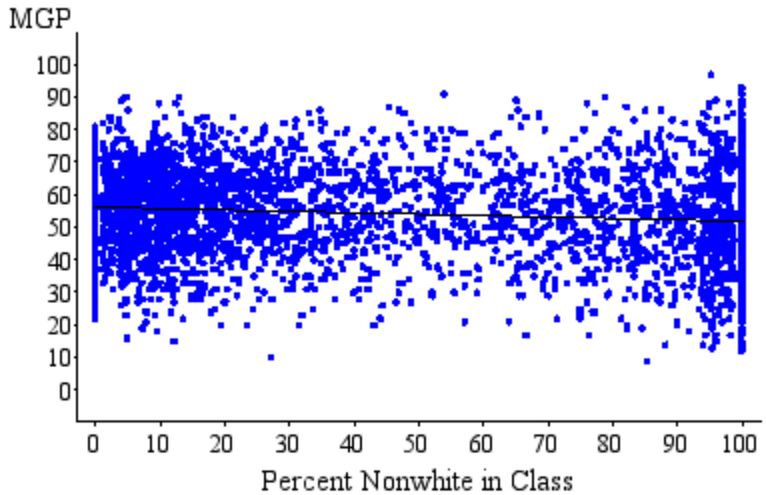


Figure H-30 Teacher MGP by Students' Mean Prior Scale Score-Math Grade 6

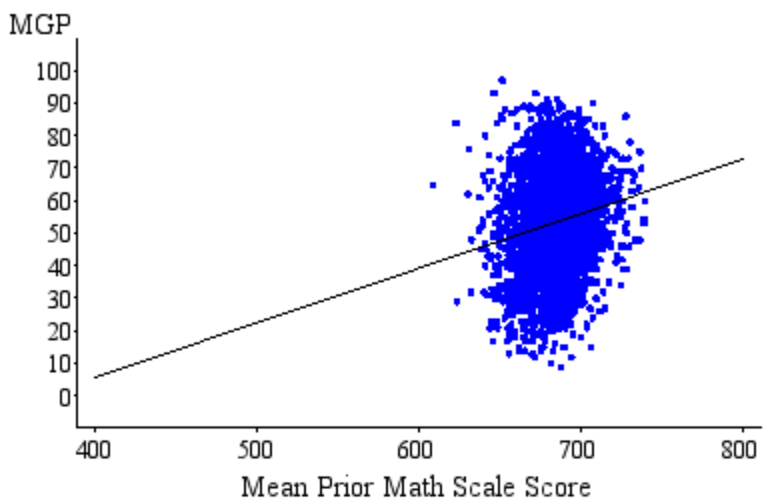


Figure H-31 Teacher MGP by Percent ELL-ELA Grade 7

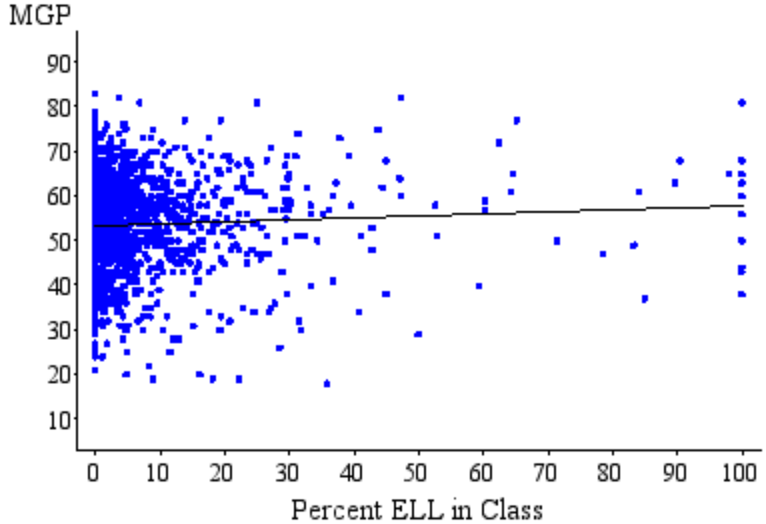


Figure H-32 Teacher MGP by Percent SWD-ELA Grade 7

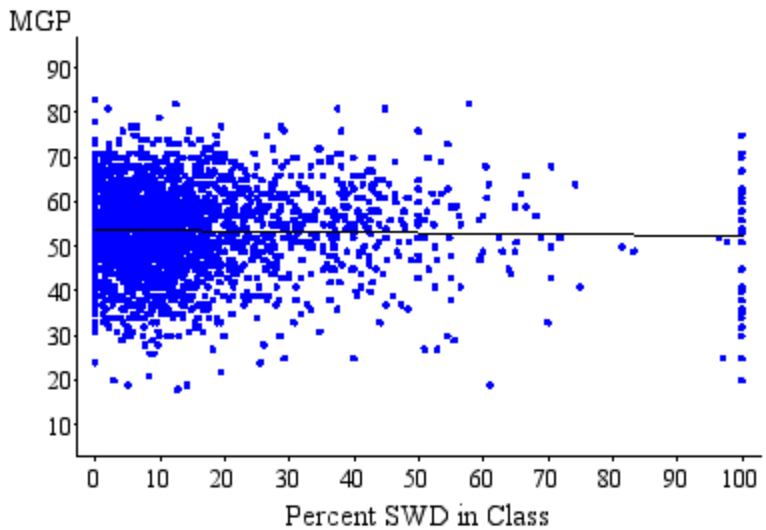


Figure H-33 Teacher MGP by Percent Economically Disadvantaged-ELA Grade 7

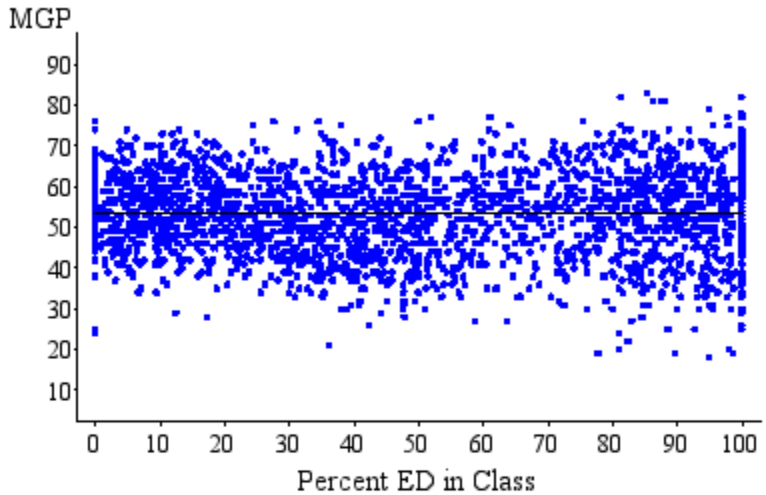


Figure H-34 Teacher MGP by Percent Nonwhite-ELA Grade 7

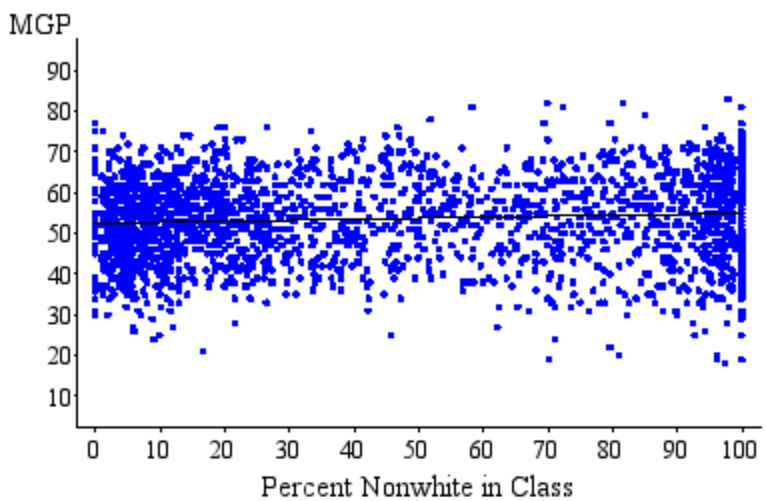


Figure H-35 Teacher MGP by Students' Mean Prior Scale Score-ELA Grade 7

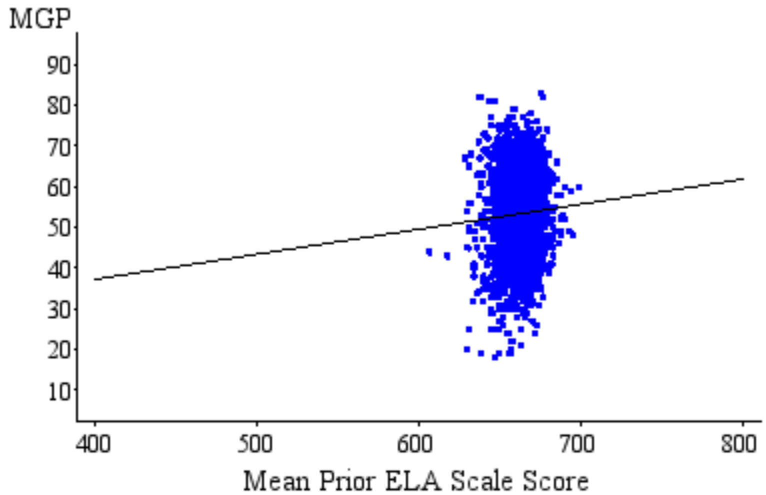


Figure H-36 Teacher MGP by Percent ELL-Math Grade 7

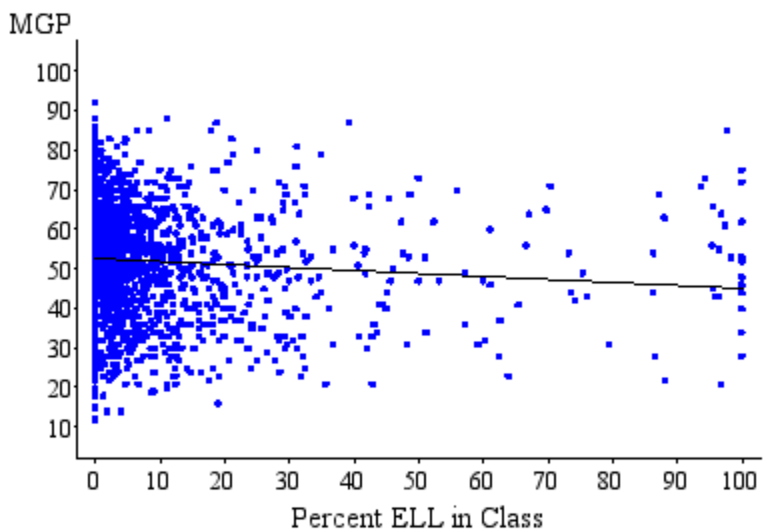


Figure H-37 Teacher MGP by Percent SWD-Math Grade 7

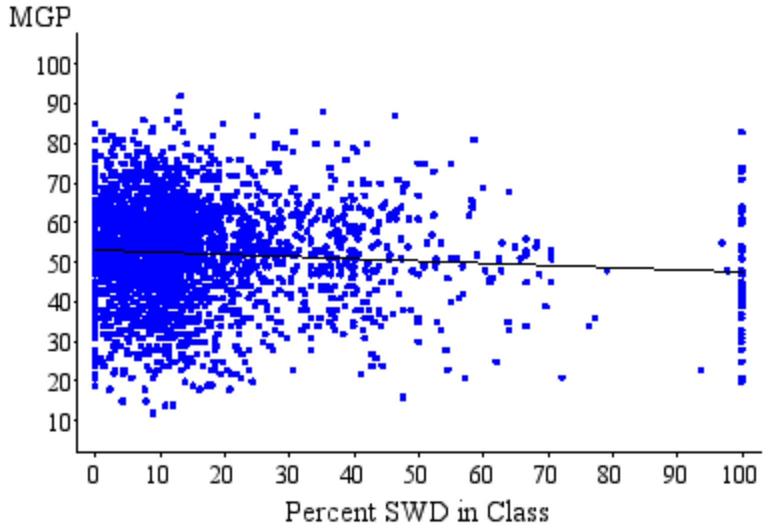
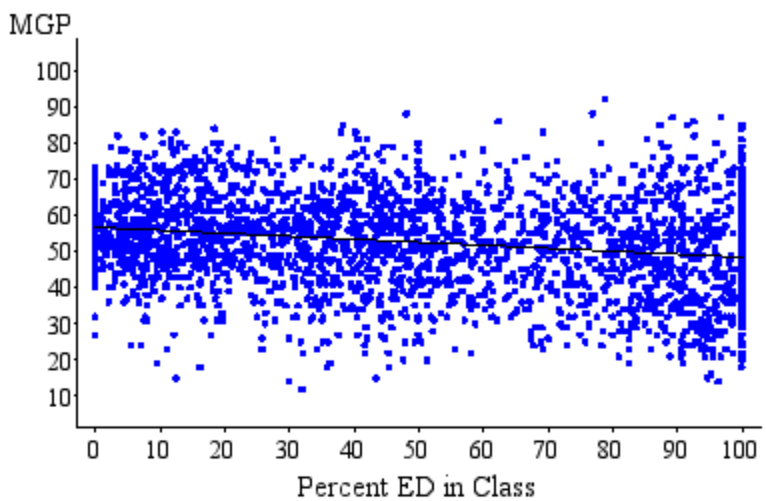


Figure H-38 Teacher MGP by Percent Economically Disadvantaged-Math Grade 7



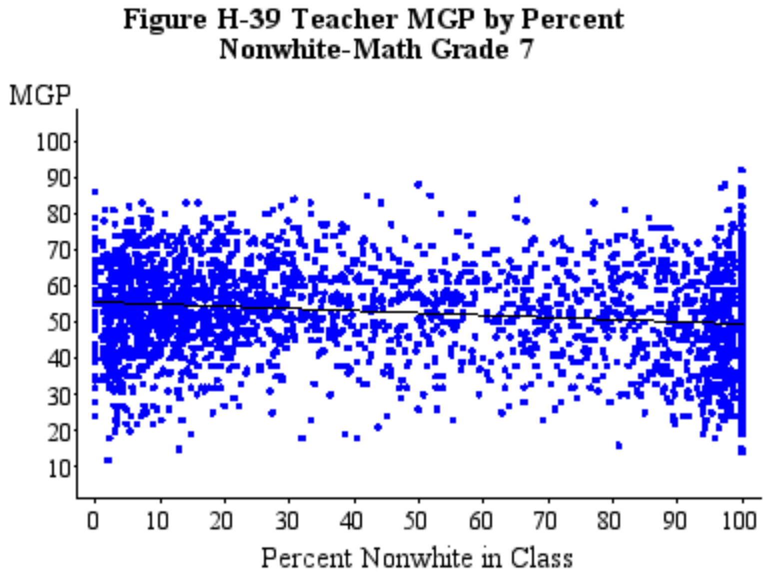


Figure H-39 Teacher MGP by Percent Nonwhite-Math Grade 7
MGP
100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
Percent Nonwhite in Class

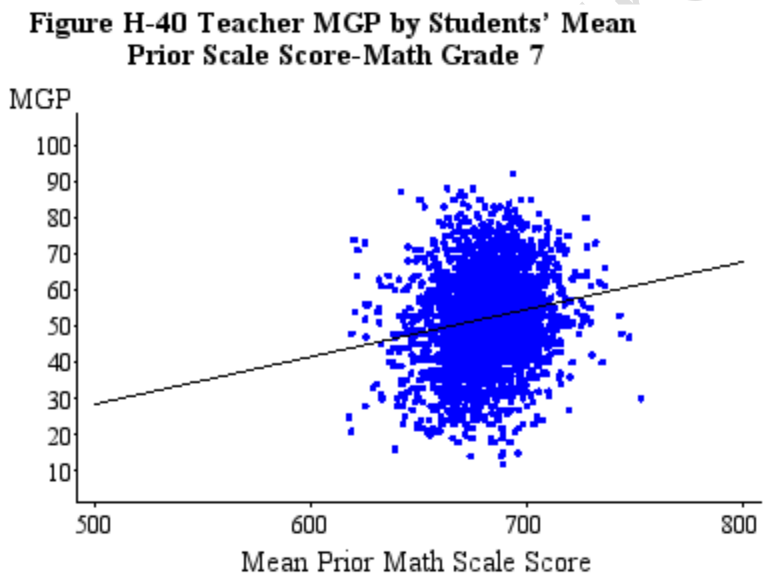


Figure H-40 Teacher MGP by Students' Mean Prior Scale Score-Math Grade 7
MGP
100
90
80
70
60
50
40
30
20
10
500
600
700
800
Mean Prior Math Scale Score

Figure H-41 Teacher MGP by Percent ELL-ELA Grade 8

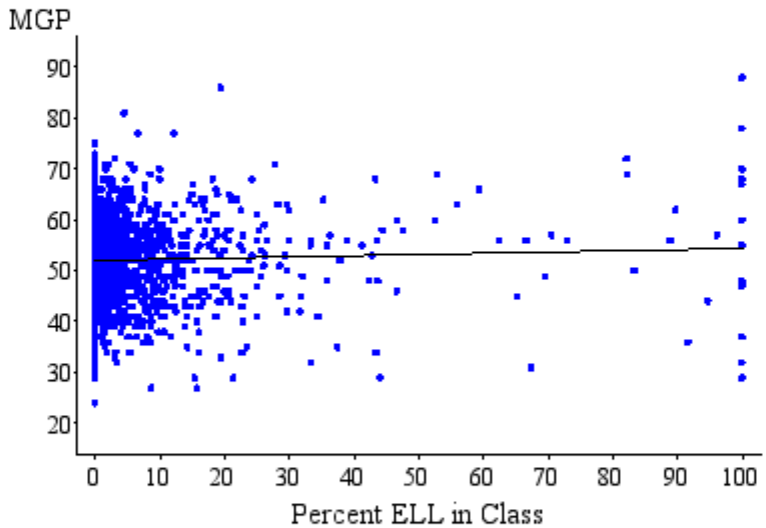


Figure H-42 Teacher MGP by Percent SWD-ELA Grade 8

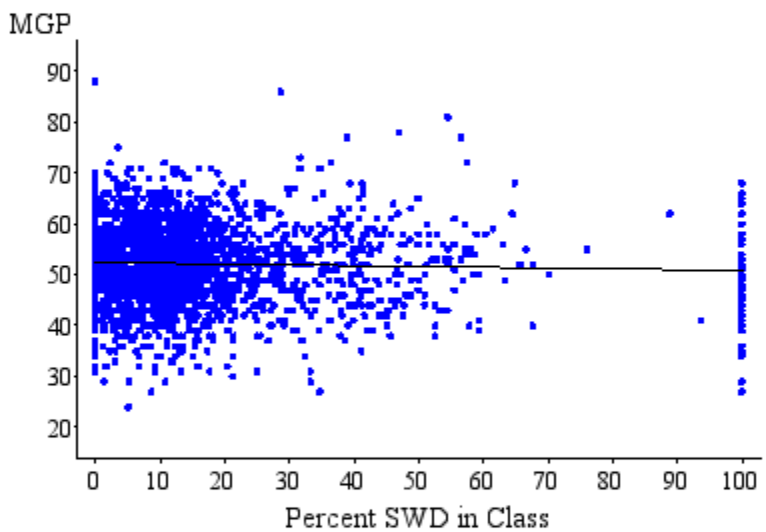


Figure H-43 Teacher MGP by Percent Economically Disadvantaged-ELA Grade 8

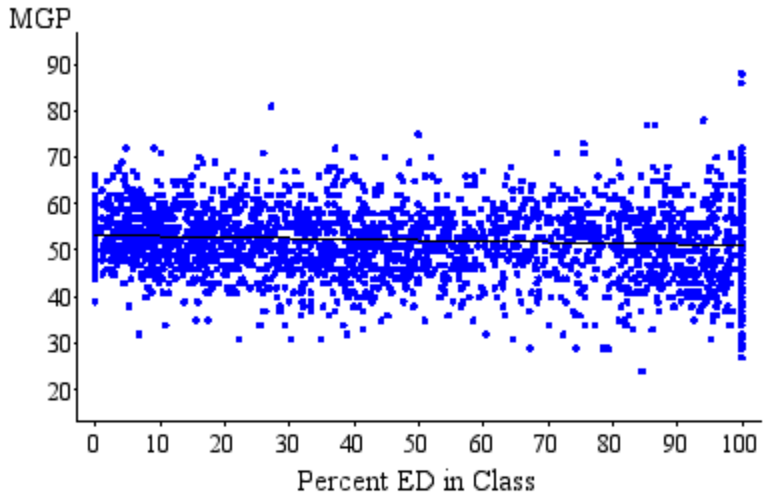


Figure H-44 Teacher MGP by Percent Nonwhite-ELA Grade 8

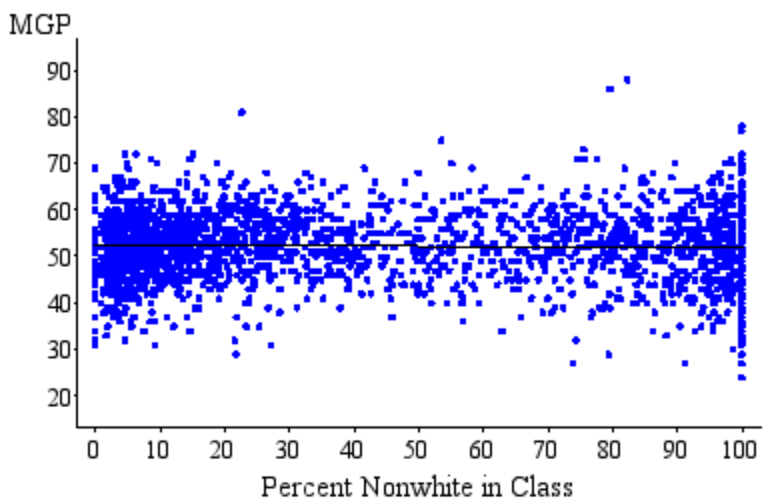


Figure H-45 Teacher MGP by Students' Mean Prior Scale Score-ELA Grade 8

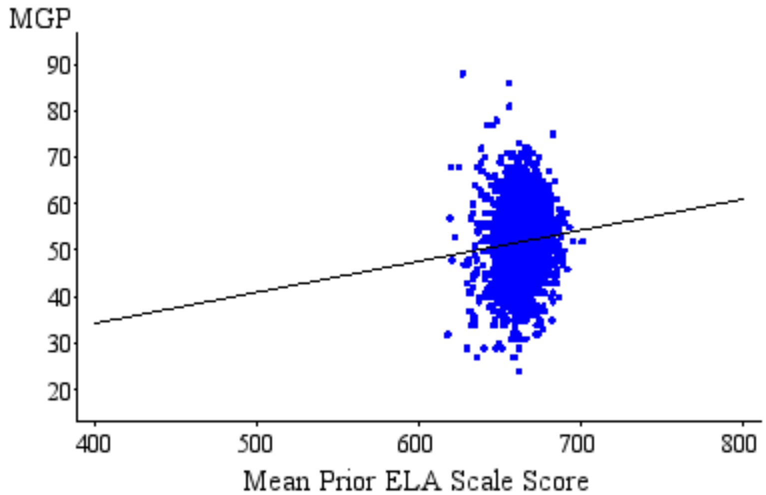


Figure H-46 Teacher MGP by Percent ELL-Math Grade 8

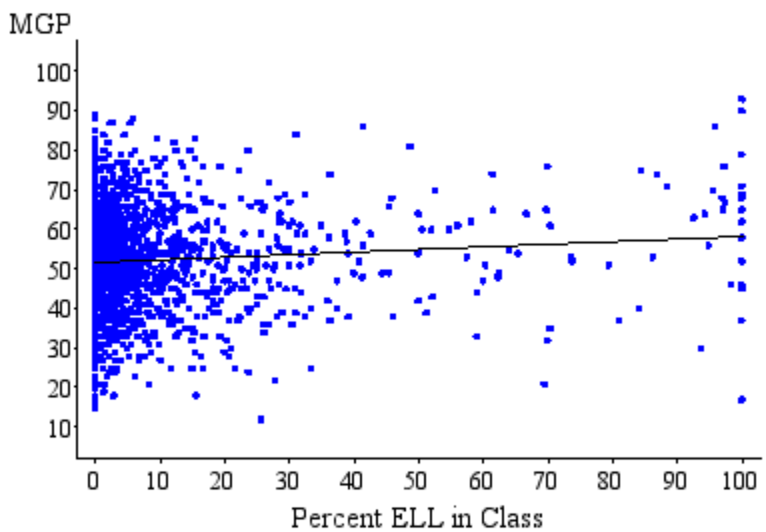


Figure H-47 Teacher MGP by Percent SWD-Math Grade 8

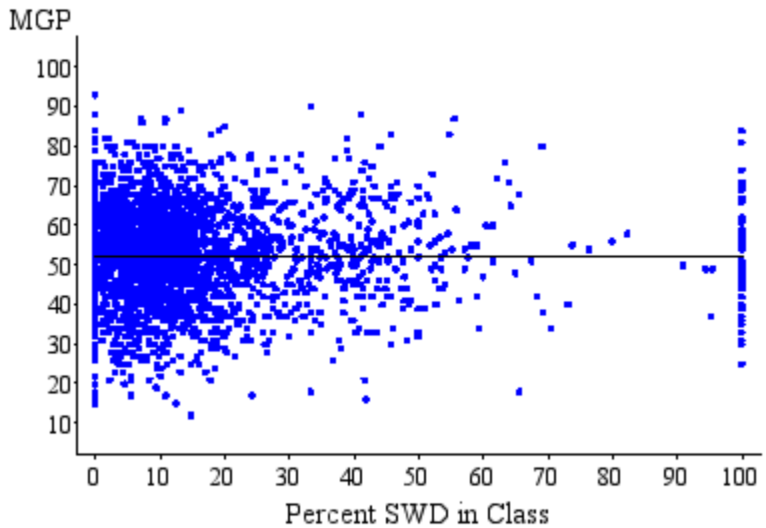


Figure H-48 Teacher MGP by Percent Economically Disadvantaged-Math Grade 8

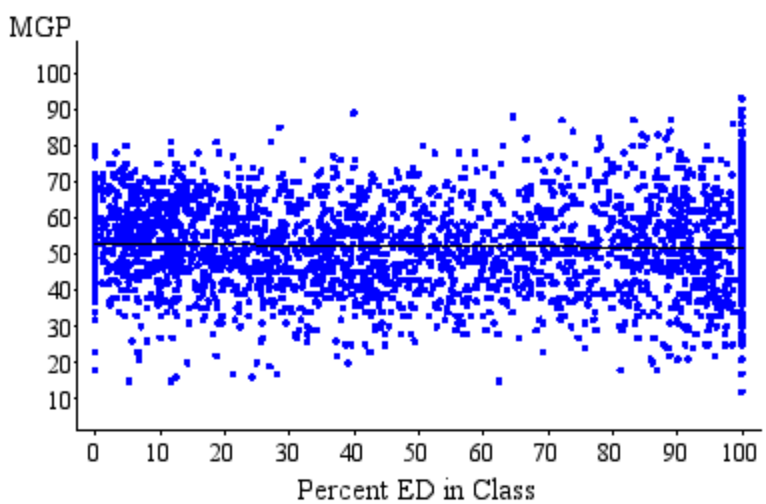


Figure H-49 Teacher MGP by Percent Nonwhite-Math Grade 8

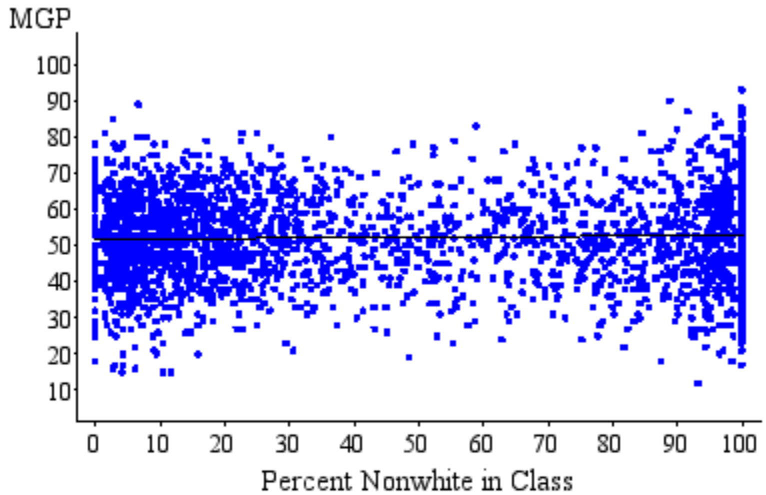
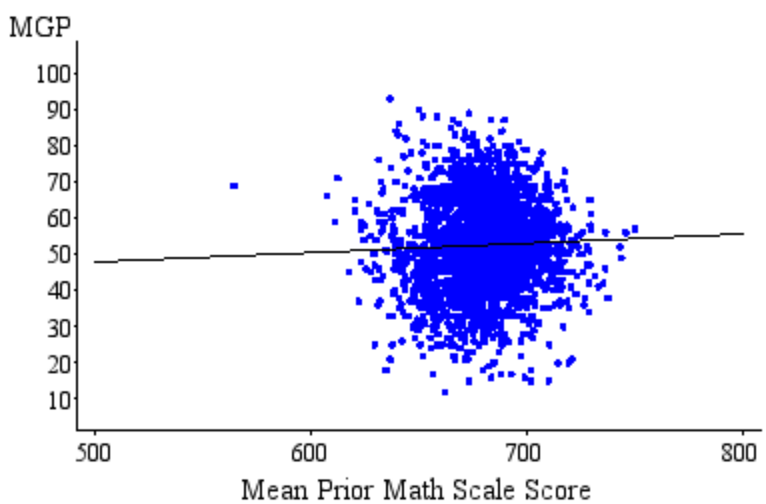


Figure H-50 Teacher MGP by Students' Mean Prior Scale Score-Math Grade 8



ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

LOCATIONS

Domestic

Washington, D.C.

Atlanta, GA

Baltimore, MD

Chapel Hill, NC

Chicago, IL

Columbus, OH

Concord, MA

Frederick, MD

Honolulu, HI

Naperville, IL

New York, NY

Portland, OR

Sacramento, CA

San Diego, CA

San Mateo, CA

Silver Spring, MD

INTERNATIONAL

Egypt

Ethiopia

Georgia

Haiti

Honduras

Kenya

Liberia

Malawi

Nicaragua

Pakistan

South Africa

Zambia

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY: 877.334.3499
www.air.org

Making Research Relevant