Adjusting Mean Growth Percentiles for Classroom Composition

Derek C. Briggs

Ruhan Circi Kizil

Nathan Dadey

University of Colorado

Center for Assessment, Design, Research and Evaluation

November 25, 2014

Overview

In this document we describe and implement an approach that can be used to adjust a teacher's mean student growth percentile ("meanGP") for differences in classroom context. The impetus for this adjustment is the notion that just as a student growth percentile is used to compare the achievement of students to peers with similar prior year test score performance, it may be sensible to compare the aggregate growth of a teacher's students to that of teacher peers facing similar classroom contexts. More specifically, by "classroom contexts" we mean factors such as the proportion of students in a classroom from low income households, the proportion of students who receive special education services, the proportion of students who are English Language Learners, and the average academic proficiency of students as demonstrated in previous years of schooling. The computation of student growth percentiles from the Georgia Growth Model does not attempt to directly disentangle the influence of these differences in classroom contexts (although it may accomplish this indirectly by controlling for prior student achievement). To the extent that student achievement is directly influenced by classroom composition (i.e., peer effects), this may advantage certain teachers over others if they are more likely to be placed in classrooms that are "easier to teach." One symptom that suggests this may be occurring is the finding that the correlations between meanGPs and classroom context variables are moderate to strong (i.e., between 0.3 and 0.6 in absolute value).

The approach we introduce here adjusts meanGPs on the basis of a second stage regression that occurs after student growth percentiles have been computed and aggregated to the teacher level. Doing this results in adjusted meanGPs that will be uncorrelated with classroom context variables (those included in the second stage regression) by construction. Note that this approach does not necessarily eliminate bias in a teacher's meanGP estimate if other confounding variables have not been included in the model (either because they were unavailable or not thought to be important). Furthermore, the approach could actually fail to reduce bias, and even increase it, to the extent that some teachers are in fact more effective when paired with certain kinds of students. The more that this is the case, the more that the adjustment presented here will "overadjust" by essentially removing variability in meanGPs that should have been attributed to differences in teaching quality (c.f., Ballou, Sanders & Wright, 2004).

The rest of this paper is structured as follows. In the next section, we formally introduce the regression approach used to adjust a teacher's meanGP for differences in classroom context. In the sections that follow we apply the approach to a subset of Georgia teachers with student outcomes in at least one of two tested subjects, math and reading. We compare the practical impact of the adjustment in two ways. First, we examine the correlation across teachers of meanGPs and adjusted meanGPs. Second, we compare the classifications of teachers into four performance categories on the basis of meanGPs and adjusted meanGPs. We find that the ordering of teachers using either approach is largely the same, with correlations ranging between 0.92 and 0.97. The vast majority of teachers are also given the same classification under either approach: 92% and 88% retain the same classification in elementary school grades in reading and math, and about 94% and 89% retain the same classification in middle school grades in reading and math. Among the teachers that do shift classification categories, the shift is never more than one category. However, for those teachers who do shift categories, it is the teachers with the more challenging students to teach that shift up, and the teachers with the less challenging students to teach that shift down. We conclude with a brief discussion of the pros and cons of the adjusted meanGP approach.

The MeanGP Adjustment

In the overview above, the adjustment to meanGPs was described as a second stage regression. Importantly, there is much that needs to be assumed about the first stage before a second stage adjustment is sensible. To begin with, it must be assumed that student growth percentiles (SGPs) have been properly estimated for each student with a subject-specific test score in the target year and a valid test score in at least the prior year. Following this, a critical step is to compute a meanGP for each teacher. In doing so one must assume that students have been properly linked to their teacher of record for each grade/course and test subject. The nature of this link can be especially challenging when students receive instruction from multiple teachers. For students linked to more than one teacher in a given grade/year, one can either choose to count the same student multiple time in computing each teacher's meanGP, or one can assign weights to each student's SGP. Note that in either case, when teachers share students, this creates a dependence between the meanGPs of teachers in the same school that violates a central assumption of the linear regression model we specify below. In the demonstration that follows we do not make adjustments for this and other possible forms of clustering. This should not have an impact on the point estimates that are the basis for the adjustments we make to meanGPs, but it will bias the standard errors we report in the underlying teacher-level regressions downward.

Let the variable Y_{ij} represent a meanGP computed for teacher *i* in test subject *j*. We then specify the following regression model[1]:

$$Y_{ij} = \beta' X_{ij} + \varepsilon_{ij} \qquad (1)$$

where X_{ij} contains a set of contextual variables computed as a function of the students to which the teacher has been linked (e.g., percentage of students eligible for free and reduced lunch services), and the last term ε_{ij} is an error term that is assumed to be independent of the included covariates and independent across teachers.[2] After estimating the regression coefficients above, we then compute for each teacher

$$e_{ij} = Y_{ij} - \hat{\beta}' X_{ij}. \qquad (2)$$

In the equation above e_{ij} is a residual meanGP: the difference between the meanGP we observe for a given teacher in a given test subject (Y_{ij}) and the meanGP predicted for a

---

[1] When a teacher only has a single class of students, then teacher-level variables are the same thing as classroom-level variables. When a teacher has multiple classes, then a teacher-level and classroom level version of the same variable can diverge. A more complicated approach would be include a classroom subscript in the model.

[2] As noted above, this assumption is surely violated by the clustering of teachers within schools and school districts.

teacher with the specific set of classroom context variables in the vector X_{ij}. An "adjusted" meanGP can be computed for each teacher as

$$adjMGP_{ij} = \bar{Y}_{.j} + e_{ij}, \tag{3}$$

where e_{ij} is defined as in equation 2 and $\bar{Y}_{.j}$ is a constant: the average for all teachers with a meanGP in test subject *j*. For teachers with an observed meanGP that is higher than that of teacher peers assigned to similar collections of students, *adjMGP* will be a percentile that is greater than average; for teachers with an observed mean GP that is lower than that of similar peers, *adjMGP* will be percentile that is lower than average. In the analyses that follow we use residual meanGPs (e_{ij}) instead of adjusted meanGPs ($adjMGP_{ij}$) to make the contrast of the two approaches easier to interpret numerically and visually (i.e., scatterplots between meanGPs and *adjMGPs* would be identical to those of meanGPs and residual meanGPs, only the metric differs).

Applying the Approach to Georgia Data

Analytic Sample and Variables

To illustrate the adjusted meanGP approach, we use 2012-13 test score data and restrict our focus to teachers in schools with grades 4 and 5 ("elementary school grades") and teachers in schools with grades 6, 7 and 8 ("middle school grades"). We also restrict our focus to test outcomes in mathematics and reading for these grades. After imposing a restriction that each teacher has at least 15 students with valid subject-specific SGPs in 2012-13, the number of teachers and students by grade levels and test subject are summarized in Table 1. These students and teachers come from a total of 1,223 unique elementary schools and 478 unique middle schools.

TABLE 1. *Student and Teacher Sample Included in Analysis*

Level	Subject	# Students	# Teachers
Elementary	Math	242,436	7,504
	Reading	271,348	8,479
Middle	Math	373,671	4,632
	Reading	410,101	5,574

We specify the following second stage regression for each grade level by subject combination:

$$Y_{ij} = b_0 + b_1 FRL\% + b_2 ELL\% + b_3 SWD\% + b_4 ACHIEVE + e_{ij}. \quad (4)$$

As before, Y_{ij} represents a teacher's subject-specific MeanGP, *FRL%* indicates the percentage of students associated with a teacher who are eligible for free or reduced price lunch services, *ELL%* indicates the percentage of students that are English language learners, *SWD%* indicates the percentage of students with disabilities (students receiving special education services), and *ACHIEVE* represents students' mean prior grade achievement (computed after first standardizing test scores within grade and subject to have a mean of 0 and a standard deviation of 1). We then compute $e_{ij} = Y_{ij} - \hat{Y}_{ij}$ and compare the ordering and classifications of teachers on the basis of Y_{ij} (i.e., the observed meanGP) and e_{ij} (i.e., the residual meanGP).

TABLE 2. *Descriptive Statistics for Teacher-Level Variables by Grade Level and Test Subject*

	Elementary School			
	Reading		Math	
	Mean	SD	Mean	SD
MeanGP	49.72	8.67	50.67	13.15
SWD%	8.26	10.05	7.59	8.63
FRL%	60.33	28.43	60.34	28.78
ELL%	7.84	14.20	8.46	14.62
Prior Achievement	-0.01	0.50	0.01	0.58
	Middle School			
	Reading		Math	
	Mean	SD	Mean	SD
MeanGP	49.48	7.39	50.05	11.21
SWD%	10.89	21.56	9.01	18.25
FRL%	60.97	26.38	61.77	25.73
ELL%	4.17	11.76	4.14	9.23
Prior Achievement	-0.01	0.6	-0.13	0.65

Table 2 provides descriptive statistics for both dependent and independent variables included in the second stage regressions. In general, the average meanGP was about 49, the average SWD% was about 8, the average FRL% was about 60, and the average ELL% was about 8 in elementary school and 4 in middle school. The average prior achievement was slightly below its within grade standardized value[3] with an SD across teachers of about 0.6.

We can examine the correlations of meanGP with each variable in Table 2 for signs that these meanGPs may be biased against teachers in certain classroom contexts. These results are presented in Table 3. For the most part, the magnitudes of these correlations are relatively small. Only two are above 0.30: the correlation with FRL% (r = -0.33) and mean prior achievement (r = 0.36) for middle school reading. In contrast, because e_{ij} is orthogonal to the collection of variables included in the regression

[3] This is because the test scores variables were standardized using the full population of test-takers, not the subset included in second-stage regressions after excluding teachers with less than 15 students.

represented by equation 4 by definition, it will be uncorrelated with *FRL%*, *ELL%*, *SWD%* and *ACHIEVE*.

TABLE 3. *Correlations Between MeanGP and Contextual Variables Across Teachers*

	Elementary School		Middle School	
	Reading	Math	Reading	Math
FRL%	-0.23	-0.21	-0.33	-0.27
SWD %	-0.08	0.01	-0.15	-0.05
ELL %	-0.01	0.02	-0.05	0.01
Prior Achievement	0.24	0.18	0.36	0.20

Results from Teacher-Level Regressions

Table 4 summarizes the results from estimating teacher-level regression models using the specification in equation 4. To facilitate comparisons across covariates holding grade level and test subject constant, as well as comparisons across grade levels and test subjects holding the covariate constant, the value in each cell in the table indicates the predicted change in a teacher's meanGP for a 1 SD increase in the value of the independent variable. (The raw results for these regressions can be found in Tables A-1 and A-2 in the appendix.) For example, a 1 SD increase in %FRL is roughly 29 for elementary school teachers, and 26 for middle school teachers. According to the regression results, holding constant ELL%, SWD% and ACHIEVE, the meanGP of an elementary school teacher with an additional 29% of students eligible for FRL services would be predicted to decrease by 1.5 percentiles in reading, and 2.3 percentiles in math. In middle school the predicted decrease in meanGP would be about the same in reading (1.3), but significantly lower in math (3.1). Of the covariates included in each regression, FRL% consistently has the largest impact on predicted meanGPs. In contrast, the marginal impact of ELL% and SWD% is much smaller. Notice however, that the impact of ELL% is actually positive—a teacher's meanGP is predicted to increase anywhere from 0.75 to 1.35 percentiles when the percent of ELL students increases by one SD (~14% in elementary school, ~10% in middle school).

TABLE 4. *Summary of Teacher-Level Regressions*

	Elementary School		Middle School	
Covariate	Reading	Math	Reading	Math
FRL %	-1.45	-2.32	-1.30	-3.12
ELL %	+0.75	+1.35	+0.75	+0.81
SWD %	-0.55	-0.48	-0.44	**-0.36**
Achieve	+1.30	+0.84	+1.67	**+0.19**
R^2	0.07	0.05	0.14	0.08

Note: The value in each cell can be interpreted as the predicted change in a teacher's meanGP for a 1 SD increase in the value of the independent variable. All estimates were statistically significant at $p < .01$ with the exception of the two values in bold face.

One of the more interesting results from these regressions is that the marginal impact of students' mean prior achievement interacts considerably by grade level and test subject. The impact on predicted meanGP for a 1 SD increase in mean prior grade achievement is a 1.3 and 1.7 percentile increase in reading for elementary and middle school levels, but only 0.8 for elementary math and just 0.2 for middle school math. Indeed, the latter result is one of only two coefficient estimates that was not statistically significant. Finally, we note that the variance in meanGP explained is quite small for all four models, with adjusted R^2 of .07, .05, .14 and .08. Only a small amount of the variance in meanGPs across teachers can be explained by the four variables included in equation 4.

Figure 1 and Table 5 depict the residual meanGP distributions by school level and test subject combination. The variability of these residuals is larger with respect to math outcomes (elementary SD = 12.8, middle SD = 10.8) than reading outcomes (elementary SD = 8.3, middle SD = 6.8). Notice that roughly 50% of teachers have residuals that are greater than 7 in absolute value for math outcomes, and greater than about 5 in absolute value for reading outcomes.

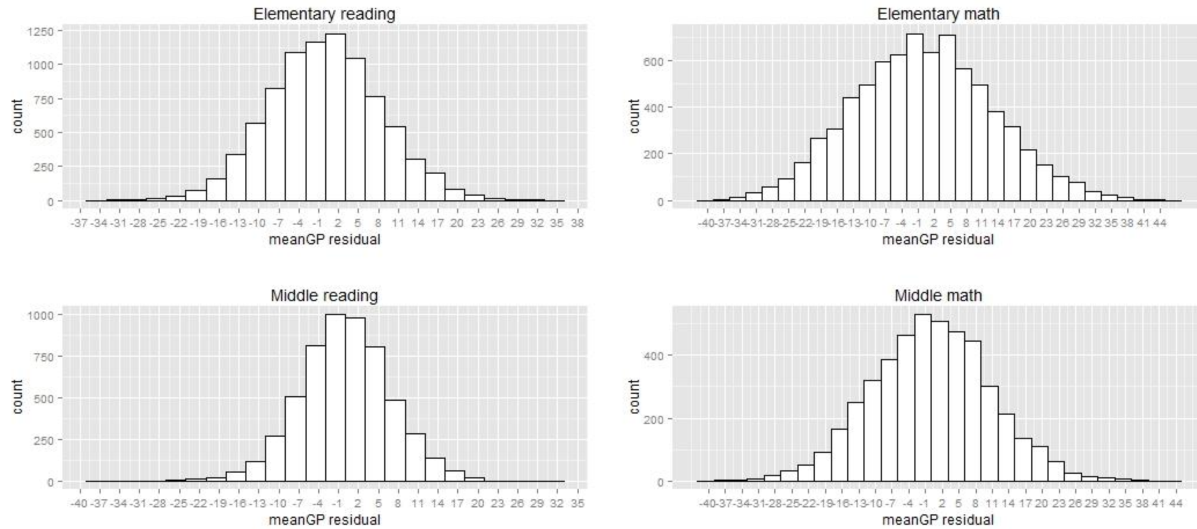

FIGURE 1. *Distributions of MeanGP Residuals by Grade Level and Test Subject*

TABLE 5. *Descriptive Statistics for MeanGP Residuals*

Level	Subject	1st Qu	Median	3rd Qu.	SD	N
Elementary	Reading	-5.65	-0.03	5.53	8.33	8479
	Math	-8.90	-0.14	8.63	12.79	7504
Middle	Reading	-4.42	-0.04	4.42	6.83	5574
	Math	-7.21	-0.02	7.29	10.75	4632

Note: “1st Qu” = “1st Quartile”; “3rd Qu = 3rd Quartile”

Comparing Teachers with Residual MeanGPs and Observed MeanGPs

The practical question is the extent to which the use of an adjusted meanGP would lead to significant differences in the ordering of teachers relative to the use of an unadjusted meanGP. As a first look at this, we present scatterplots of meanGPs and residual meanGPs by grade level and test subject in Figures 2 and 3.. In general, there is a strong linear relationship between meanGPs and residual meanGPs, with correlations of about 0.96 in elementary grade reading and math and middle grade math. The lowest correlation is 0.92 for middle level reading.

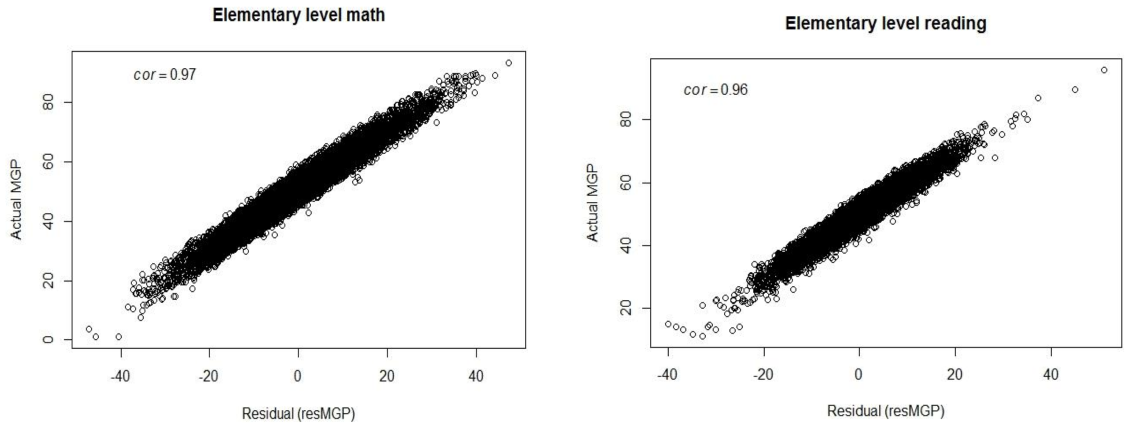

FIGURE 2. *Scatterplots of Elementary Grade Teachers' Observed MeanGPs (y-axis) and Residual MeanGPs (x-axis)*

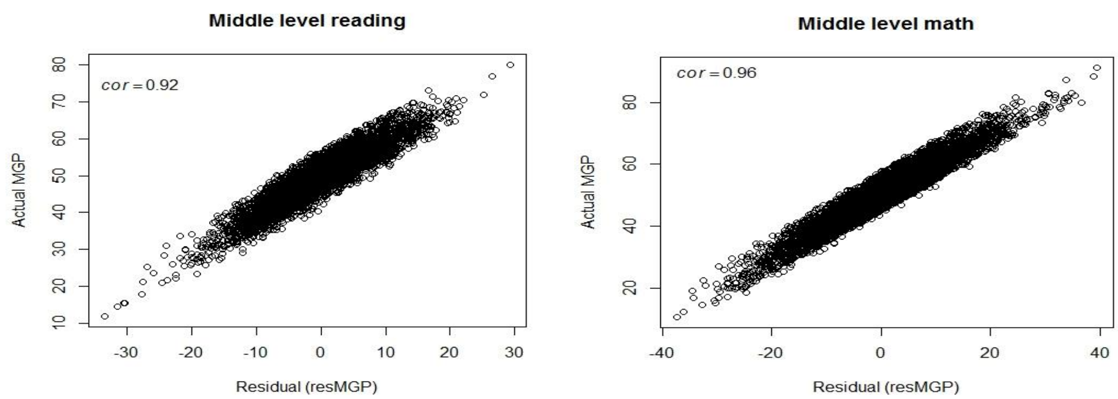

FIGURE 3. *Scatterplots of Middle Grade Teachers' Observed MeanGPs (y-axis) and Residual MeanGPs (x-axis)*

Next we compare differences in how teachers would be classified into "performance" categories using either approach. As part of Georgia's TKES model, thresholds of 30, 40 and 65 are used to convert meanGP values into four categories, where a 1 is the lowest and is indicative of a teacher that might be deemed "ineffective," a 2 represents a teacher that might ultimately be characterized as "needs development," and a 3 or 4 represent teachers that might be characterized as "effective" or "exemplary," respectively. Instead of mapping these criterion-referenced thresholds directly onto the residual meanGP distribution, we apply a normative approach to find thresholds on the residual meanGP distributions that have the same relative distances from the mean as those that exist on the meanGP distribution. That is, we find the z-scores that correspond to the meanGP thresholds of 30, 40 and 65 for each grade level and test subject and the convert these into the corresponding residual meanGP thresholds. For example, recall that the mean and SD of meanGPs for elementary reading teachers is 49.7 and 6.7 respectively. To find the z-score corresponding to a meanGP of 30 for elementary reading, we compute z = (30- 49.7)/8.67 = -2.27. The complete sets of z-scores computed in this manner are shown in Table 6 below. One thing to notice in Table 6 is that even though the MeanGP classification thresholds are fixed, because the variability of meanGPs by grade level and test subject, the corresponding z-scores can be quite different. For example, to be classified as "ineffective" in reading, a middle school teacher's meanGP would have to be 2.63 SDs below average. In contrast, to be classified as "ineffective" in math, a middle school teacher would only have to be 1.79 SDs below average.

TABLE 6. *Z-score thresholds that correspond to meanGP thresholds used by Georgia to Classify Teachers*

MeanGP	Elementary School		Middle School	
Threshold	Reading	Math	Reading	Math
30	-2.27	-1.57	-2.63	-1.79
40	-1.12	-0.81	-1.28	-0.90
65	1.76	1.09	2.10	1.33

Next we use these z-scores to derive corresponding thresholds on the residual meanGP distributions. For example, with respect to elementary reading we ask what residual meanGP value is -2.27 SDs below average? Since the mean of residual meanGPs is 0, this is akin to solving the following equation -2.27 = x/8.3. The answer is -18.9: any teacher with a residual meanGP that is 18.9 percentiles below average would be in the lowest performance category (i.e., "ineffective"). The same process is used for all other threshold values.

Once teachers have been classified according to their location on each distribution, we can examine the crosstabulation to see how many teachers would shift categories when going from an observed meanGP to a residual meanGP. The complete crosstabs are provided in Tables 7a-10a. The number of teachers in each cell is shown as a percentage of the total number of teacher across all categories. To summarize, for elementary grade teachers, 92.6% and 87.8% remain in the same performance category in reading and math respectively; for middle grade teachers, 93.8% and 89.0% remained in the same performance categories. Among the small percentage of teachers that shifted categories, no teacher shifted more than one category in a positive or negative direction.

However, among those teachers that switch categories, some interesting trends are evident and these are displayed in Tables 7b-10b. Across grade levels and content domains, those teachers that move up a category when growth is evaluated on the basis of a residual meanGP in place of observed meanGP, tend to have large proportions of students who are eligible for free and reduced price lunch services and a mean level of prior grade test performance that is far below average. For example, Table 8b shows that there were 472 teachers in elementary school grades who shifted up one category (e.g., from 1 to 2, 2 to 3 or 3 to 4) when examining evidence of student growth in mathematics on the basis of residualized meanGPs. The prior grade achievement for the students of these teachers was roughly half an SD below average, and the proportion of students who were FRL eligible was between 85 and 91%. By contrast, there were 442 teachers who shifted down by one category. The prior grade achievement for the students of these teachers was typically more than half an SD above average, and the proportion of students who were FRL eligible was between 20 and 34%. These results are essentially what one would predict on the basis of the meanGP adjustment for classroom composition: teachers with more challenging classroom contexts benefit relative to teachers with less challenging classroom contexts.

TABLE 7a. *Differences in Teacher Classifications from MeanGP to Residual MeanGP: Elementary Grade Reading*

	Residual MeanGP			
MeanGP	1 (Low)	2	3	4 (High)
1 (Low)	**0.80**	0.32	0.00	0.00
2	0.32	**9.27**	2.57	0.00
3	0.00	2.21	**79.36**	1.16
4 (High)	0.00	0.00	0.84	**3.16**

Note: Values in cells represent percentages out of total number of teachers (N = 8,479)

TABLE 7b. *Characteristics of Students Associated with Teachers Who Switch Categories: Elementary Grade Reading*

	Shifts to Higher Classifications			Shifts to Lower Classifications		
	1 → 2	2 → 3	3 → 4	2 → 1	3→ 2	4 → 3
N	27	218	98	27	187	71
meanGP	28.4	38.6	63.4	31.8	41.8	66.0
FRL%	92.8	89.6	81.5	39.0	30.2	23.3
SWD%	17.1	14.0	10.6	5.8	7.2	4.3
ELL%	2.7	4.3	7.6	19.0	6.6	3.5
Achieve	-0.76	-0.60	-0.49	0.32	0.53	0.76

TABLE 8a. *Differences in Teacher Classification from MeanGP to Residual MeanGP: Elementary Grade Math*

	Residual MeanGP			
MeanGP	1 (Low)	2	3	4 (High)
1 (Low)	**4.46**	1.33	0.00	0.00
2	1.32	**11.90**	2.84	0.00
3	0.00	2.60	**59.55**	2.12
4 (High)	0.00	0.00	1.97	**11.90**

Note: Values in cells represent percentages out of total number of teachers (N = 7,504)

TABLE 8b. *Characteristics of Students Associated with Teachers Who Switch Categories: Elementary Grade Math*

	Shifts to Higher Classifications			Shifts to Lower Classifications		
	1 → 2	2 → 3	3 → 4	2 → 1	3→ 2	4 → 3
N	100	213	159	99	195	148
meanGP	28.2	38.3	63.3	32.1	42.0	67.0
FRL%	90.7	89.2	85.1	33.8	29.5	20.3
SWD%	7.3	6.7	6.6	7.1	8.7	7.3
ELL%	3.5	3.4	6.0	9.8	9.6	4.6
Achieve	-0.59	-0.52	-0.44	0.43	0.53	0.82

TABLE 9a. *Differences in Teacher Classification from MeanGP to Residual MeanGP: Middle School Reading*

	Residual MeanGP			
MeanGP	1 (Low)	2	3	4 (High)
1 (Low)	**0.59**	0.20	0.00	0.00
2	0.18	**6.35**	2.19	0.00
3	0.00	1.99	**85.88**	0.99
4 (High)	0.00	0.00	0.68	**0.95**

Note: Values in cells represent percentages out of total number of teachers (N = 5,574)

TABLE 9b. *Characteristics of Students Associated with Teachers Who Switch Categories: Middle School Grade Reading*

	Shifts to Higher Classifications			Shifts to Lower Classifications		
	1 → 2	2 → 3	3 → 4	2 → 1	3→ 2	4 → 3
N	11	122	55	10	111	38
meanGP	28.1	38.6	62.8	31.9	41.9	66.7
FRL%	89.1	88.1	76.9	45.1	37.5	18.6
SWD%	52.3	31.3	28.5	6.3	6.4	2.7
ELL%	1.57	4.74	8.67	20.57	1.33	0.76
Achieve	-1.17	-0.86	-0.7	0.08	0.45	1.02

TABLE 10a. *Differences in Teacher Classification from MeanGP to Residual MeanGP: Middle Grade Math*

	Residual MeanGP			
MeanGP	1 (Low)	2	3	4 (High)
1 (Low)	**2.76**	0.67	0.00	0.00
2	0.71	**11.77**	2.46	0.00
3	0.00	2.89	**67.44**	1.84
4 (High)	0.00	0.00	2.40	**7.06**

Note: Values in cells represent percentages out of total number of teachers (N = 4,632)

TABLE 10b. *Characteristics of Students Associated with Teachers Who Switch Categories: Middle School Math*

	Shifts to Higher Classifications			Shifts to Lower Classifications		
	1 → 2	2 → 3	3 → 4	2 → 1	3→ 2	4 → 3
N	31	114	85	33	134	111
meanGP	28.3	38.4	63.1	32.0	42.0	66.9
FRL%	92.4	87.4	85.5	40.1	34.6	22.1
SWD%	16.9	13.4	12.0	9.6	5.2	7.0
ELL%	1.8	2.3	4.2	8.1	4.1	2.2
Achieve	-0.76	-0.54	-0.51	0.37	0.4	0.65

To Adjust or Not to Adjust?

Although the practical impact of adjusting meanGPs using the second stage regression approach illustrated here appears to be rather small with roughly 90% of teachers staying in the same performance category, when applied to thousands of teachers, even a small impact can have meaningful implications. For example, among the 7,504 and 4,632 elementary and middle school teachers linked to student performance in math, 914 and 508 respectively would have performance categories that shift by one when going from meanGP to adjusted meanGP. These teachers are roughly evenly split among those who shift up and those that shift down. Those that shift up tend to have students who are, on the whole, more challenging to teach. Those that shift down tend to have students who are, on the whole, less challenging to teach.

At first blush, there appears to be evidence that for each grade level and test subject, teachers in performance category 2 (“needs development”) on the basis on an unadjusted meanGP would be more likely to move up to category 3 (“effective”) on the basis of an unadjusted meanGP than they would be to move down to category 1 (“ineffective”). However, this can be attributed to a regression to the mean effect at work as the converse is also true: teachers in performance category 3 on the basis of an unadjusted meanGP are more likely to shift down to a category 2 on the basis of an adjusted meanGP than they would be to shift up to a category 4 ("exemplary"). The basis for these regression effects is the fact that the means of each distribution are closer to categories 2 and 3 than to 1 and 4. It can also be argued that the shifts in classifications observed when shifting from unadjusted to adjusted meanGPs are likely to be far smaller than the shifts we would expect to observe just on the basis of chance differences in the student cohorts that teachers are assigned from year to year.

The biggest advantage of the adjusted meanGP approach is that it addresses a legitimate concern that teachers are not being fairly compared on the basis of the unique classroom contexts they may face. Doing this adjustment as a distinct second stage at the teacher level after growth percentiles have been computed for student in a first stage also lends itself to a relatively straightforward explanation of the process. Students are being compared to students with similar prior achievement, and teachers are then being compared to teachers with similar classroom contexts.

There are a number of possible challenges that need to be taken into account, and technical/logistical questions that would need to be answered before such an approach could be implemented at scale. First, it is not immediately clear how to decide when to stop including contextual variables in the second stage meanGP regression. The four variables included in this illustration are likely candidates, but many others would be possible, and the more variables that are included, the more difficult it may be to ascertain (and explain) just which teachers are being compared to one another. Second, it is not clear how many unique regressions should be specified. At one extreme, it would be possible to have as many regression models as there are unique test subjects and grades/courses. At the other extreme, a single combined regression could be specified after combining all available meanGPs for each teacher across test subjects. In this illustration, four separate regression models were specified, and each resulted in somewhat different adjustments for the contextual variables included (See Table 4). It might prove difficult to explain, for example, why the regression adjustment for %FRL has two to three times the weight in math as it does for reading. Third, a decision would need to made as to whether the adjustment should be allowed to vary from year to year as cohorts of students enter and exit the system, or whether a single adjustment should be computed on the basis of an average across multiple years of data in a matter akin to the baseline referencing approach used to compute meanGPs.

Finally, as noted at the outset, there is no guarantee that the adjusted meanGP approach is necessarily removing clear sources of bias if some of the factors included in the regression are in fact correlated with teaching quality. The adjusted meanGP approach can certainly make it appear that there is no correlation between poverty and teacher effectiveness, but this may obfuscate fundamental inequities in the quality of teachers found in more disadvantaged schools. Nonetheless, the empirical results presented here suggest that even if the adjusted approach is "overcorrecting," this is only having a practical consequence for a small proportion of all teachers.

References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37–65.

Appendix

Table A-1. Elementary School Teacher-Level Regression Results

	Reading			Math		
	Estimate	SE	t-value	Estimate	SE	t-value
Constant	52.59	0.33	158.89	54.71	0.48	111.75
FRL%	-0.05	0.00	-9.67	-0.08	0.01	-11.73
ELL%	0.05	0.01	7.76	0.09	0.01	8.02
SWD%	-0.05	0.01	-5.63	0.02	0.02	1.02
ACHIEVE	2.32	0.25	8.97	1.96	0.35	5.58
Adj R^2	.076			.054		

Table A-2. Middle School Teacher-Level Regression Results

	Reading			Math		
	Estimate	SE	t-value	Estimate	SE	t-value
Constant	52.73	0.35	152.40	57.17	0.56	102.42
FRL%	-0.05	0.01	-8.95	-0.12	0.01	-13.30
ELL%	0.04	0.01	4.93	0.09	0.02	5.33
SWD%	-0.02	0.00	-3.80	-0.02	0.01	-1.99
ACHIEVE	2.78	0.26	10.84	0.29	0.37	0.80
Adj R^2	.144			.078		