education sciences

MDPI

Article

Impact of Contextual Predictors on Value-Added Teacher Effectiveness Estimates

Ismail Aslantas

School of Education, Durham University, Durham DH1 1TA, UK; ismail.aslantas@durham.ac.uk

Received: 4 November 2020; Accepted: 16 December 2020; Published: 18 December 2020

Abstract: It is widely believed that the teacher is one of the most important factors influencing a student's success at school. In many countries, teachers' salaries and promotion prospects are determined by their students' performance. Value-added models (VAMs) are increasingly used to measure teacher effectiveness to reward or penalize teachers. The aim of this paper is to examine the relationship between teacher effectiveness and student academic performance, controlling for other contextual factors, such as student and school characteristics. The data are based on 7543 Grade 8 students matched with 230 teachers from one province in Turkey. To test how much progress in student academic achievement can be attributed to a teacher, a series of regression analyses were run including contextual predictors at the student, school and teacher/classroom level. The results show that approximately half of the differences in students' math test scores can be explained by their prior attainment alone (47%). Other factors, such as teacher and school characteristics explain very little the variance in students' test scores once the prior attainment is taken into account. This suggests that teachers add little to students' later performance. The implication, therefore, is that any intervention to improve students' achievement should be introduced much earlier in their school life. However, this does not mean that teachers are not important. Teachers are key to schools and student learning, even if they are not differentially effective from each other in the local (or any) school system. Therefore, systems that attempt to differentiate "effective" from "ineffective" teachers may not be fair to some teachers.

Keywords: value-added model; teacher performance evaluation; teacher accountability; effectiveness; contextual predictors

1. Introduction

With growing interest in identifying and improving student achievement through large-scale assessments, several researchers have conducted a considerable number of studies to uncover the factors that influence student academic performance. The common consensus is that of school-related factors, it is teachers that are the most significant [1–3]. Accordingly, politicians have tended to develop educational policies that consider schools and teachers responsible for the performance of pupils. One of the key concerns of decisionmakers is to ensure that effective teachers are hired in classrooms and, similarly, parents also desire their children to be enrolled in the best schools and be taught by well qualified teachers. There is, therefore, a need for a performance appraisal framework that can evaluate the role of teachers in achieving student targeted success. While it is accepted that evaluating teacher performance is beneficial in enhancing teacher development and student outcomes, it is a complex process and there is no perfect measure [4]. Since longitudinal student achievement results have become readily accessible, academics and decisionmakers have considered alternative measures that focus upon the students' achievement growth rather than the percentage of students in the class or the school reaching a threshold value determined by teacher and school effectiveness indicators.

Since teachers' performance results are heavily influenced by the context of students and other factors that are beyond the control of teachers [5], value-added models that take into account the cumulative nature of students' learning and that allow students' progress over time to be monitored, have been recently favored. Value-added models (VAMs) are statistical approaches used to estimate the effectiveness of individual teachers and schools by employing students' longitudinal test scores (and some covariates). While a variety of value-added models, with their own specific advantages and drawbacks, have been implemented to estimate an individual teacher's contribution to student attainment, the fundamental concept behind all VAMs is to determine the changes in the students' school achievement over the years. In all student standardized assessment-based VAMs, a particular teacher's performance is statistically estimated using the test scores of their students by subject and grade. Changes in student achievement in tests performed for at least two consecutive years are then attributed to teacher effects. Since students can be taught by different teachers, it is difficult to say if any improvement in student test scores is the result of any particular teacher's effort. However, by using student test scores for the same teacher for multiple years, as in this current study, it can be ensured that the effects (if any) can be attributed to the teacher. In the VAM concept, teacher effect has a special meaning that relates to evaluating the discrepancies between expected and observed student test scores [3,6].

VAMs are one of the most contentious and critical matters for education policy in the assessment of teacher performance. Theoretically, VAMs separate a particular teacher's contribution to the achievement of students from other factors that are beyond the control of the teacher. These include student characteristics, school and classroom characteristics and students' prior year test scores. On the basis of such performance evaluations, high-stakes decisions regarding teachers' professional careers are made, such as pay rises, promotions, or even the loss of a job. The VAM score estimates how much a teacher has contributed to enhancing their students' performance. Based on this score, the teacher is ranked by how much difference he/she makes in their students' achieved test scores compared to their predicted scores. Such a ranking is often used to reward or penalize teachers. However, the premise of this effectiveness evaluation has been found to be flawed by many educators [7–9]. For example, VAMs assume that students are randomly assigned to teachers, which is rarely the case in reality. Nonrandom assignment of teachers to students can bias the estimate of teacher effects [10]. Other studies argue that standardized student achievements alone are not a sufficiently reliable metric to establish the link between student learning and what teachers do. The Economic Policy Institute (EPI) recommend that standardized academic performance be considered as just one factor among many factors, for a more accurate understanding of the actual teaching performance of teachers in the classroom [11]. Despite these concerns, several scholars have come up with different models, each trying to overcome the limitations of the others.

Nevertheless, the literature agrees that lagged test scores may provide a better measure of VAM estimates. Hu [12] recorded that on average, 57% and 59% of the changes in student math and reading test scores can be explained by the closest lagged score in the related subjects, respectively. Similarly, Kersting et al. [6] found that just one test score from a previous year explained 68% of the variation in the students' actual scores. In accordance with the scope of this study, Rothstein [13] analyzed the impact of variables on the value-added estimates through modifying the models' R-squared by employing 28 contextual variables, such as ethnicity, sex, free/reduced lunch status, parental education, etc. The integration of the 28 variables into the equations contributed to changes of 0.05 and 0.01 in R^2 depending on the model to which they were applied. While existing literature suggests that lagged scores may be a better measure in explaining the variation in a student's actual test score [1,14–18], there is no consensus regarding the association between contextual predictors, such as observable student, teacher, classroom and school characteristics and student outcomes.

Although Alban [14] and Gagnon [15] report that adding student-level variables in the estimates improves predictability, most studies suggest that the influence of variables at the student level employed in the equations is negligible [6,12,19–23]. Gagnon [15] analyzed the use of different

predictors at the student level in the value-added teacher effectiveness estimates, including lagged scores, poverty, ethnicity, sex, English as a second language, disability status, attendance and suspension. The results show that the previous test scores in the third through to the seventh grade explain roughly 76% of the variance in test scores in the eighth grade. The researcher concluded that the lagged test scores are better predictors of potential achievement. Other student factors employed in the equations, except ethnicity, are shown to be significant predictors. Hu [12], on the other hand, used up to three years of pupil test scores in math and reading, as well as student demographic characteristics, such as sex, race, language learner status, gifted and disability status, and class size and found that the inclusion of all variables explained only 2% of the variance in the students' actual test scores.

Other studies looked at the use of teacher-level predictors in value-added estimates. These studies show a positive relationship between the teacher variables, such as Grand Point Average (GPA) score [24], permanent contract [25], being certified [26] and their prior year performance [27]. Conversely, some studies have suggested little or no advantage in using teacher-/classroom-level variables [18,28–30]. For instance, Leigh [31] examined the impact of teacher characteristics, such as gender, age, experience and DETA rating score (The Queensland Department of Education, Training and the Arts), on the teacher effectiveness estimates in terms of changes in the R-squared values of the models. Student achievement gains were regressed on the teacher characteristics, and it was found that the explained variance in the achievement gain, employing all teacher characteristics, never exceeded 1%.

In line with the scope of the study, the literature reviewed also focused on school-related variables. This revealed that adding school-level predictors makes little difference to the teacher value-added effectiveness estimates [18,24,32,33], except for one study suggesting that including the percentage of students receiving special education services at the school level in the estimates has a benefit [14]. To determine the important variables in the teacher value-added effectiveness estimates, Alban [14] employed a range of student-, teacher- and school-related variables, including sex, ethnicity, prior attainment, language learner status, teaching experience, highest qualification held, the socio-economic level of the school and the proportion of students receiving special education services. The study found that prior success is the only significant variable for each estimate, followed by gender and the percentage of students receiving special services at school.

In light of the findings of the existing literature, the aim of this study is to examine the contribution of contextual predictors and observable student, teacher/classroom and school characteristics, on teacher effectiveness estimates calculated by a value-added model.

The following research questions were addressed:

1. To what extent can the value-added estimates for mathematics teachers be explained by student characteristics, other than prior attainment?
2. To what extent can the value-added estimates for mathematics teachers be explained by school characteristics, over and above student characteristics?
3. To what extent can the value-added estimates for mathematics teachers be explained by teacher/classroom characteristics, over and above student and school characteristics?

2. Materials and Methods

2.1. Participants

The research involved 8th grade mathematics teachers. The participants of this study were 230 mathematics teachers from 145 secondary schools and their 7543 students in 8th grade during the 2016–2017 school year in the Samsun Provincial Directory of National Education, Turkey. Students that could be academically tracked from Grades 6 to 8 (Key stage 3–Years 7 to 9) were the target population in this study. However, as with most longitudinal studies, the inclusion of each lagged test score in the study results in a loss in the study population. The main reasons for these losses may be that, during the testing time, students did not take the test, or they moved away from the province. In a target

population for which 8th grade test scores are available, the average loss of employing one previous year's test score was approximately five percent.

2.2. Data

To conduct value-added teacher performance estimates, it is essential that student data can be linked to teacher data longitudinally [34,35]. The research employed the longitudinal data from schools in Samsun province, Turkey, to examine the contribution of a range of contextual variables at student, teacher/classroom and school levels on maths teacher effectiveness estimates. The Samsun Provincial Directory of National Education has been running the "Step by Step Achievement" project (SBSA) since 2015 [36] and the exam scores from 2016–2017 in maths were used as outcome variables. All the SBSA exam scores over the years, students' background information, their school and class information (i.e., teacher's name) are stored electronically in the Samsun Provincial Directorate of National Education's private electronic systems. The longitudinal test scores of all 8th grade students registered on the system were downloaded and these data were then merged with other student-level data, including their names, unique school number, gender and language learner status.

Since the SBSA did not contain teacher-/classroom-level data, these were obtained from each of the schools via the Provincial Directorate. The teacher/classroom information included gender, classroom size, number of years of teaching experience, number of years teaching in the current school, teaching appointment field, teachers' major degree subject and their highest level of qualification and its field.

School-level data were obtained from the Ministry of National Education's official website [37]. A list of all secondary schools was first downloaded from the website of Samsun Provincial Directory of National Education [38]. Only schools that were included in the project were retained. School-level data included school type (private or state-funded), school category (general, regional boarding or vocational secondary school), location (urban, suburban or rural) and school service scores. After all three data files were merged to maintain confidentiality, participants' identities were removed from the data set and identification numbers were assigned to each student, teacher, school and school location.

Data used in the study included student longitudinal achievement data, spanning up to three consecutive years, students' characteristics, teacher/classroom background information and school information. Table 1 summarizes the outcome variables and independent student, teacher/classroom and school-level variables included in this study.

Table 1. Summary statistics for data used in value-added estimates.

	Mean	**SD**
Outcome variable		
Math test score in Grade 8	9.18	4.348
Independent Variables		
Panel A: Student characteristics		
Female	48.3%	
Language Learner	0.2%	
Math test score in Grade 7 (t-1)	9.52	4.888
N (student)	7543	
Panel B: Teacher/Classroom characteristics		
Female	53.5%	
Class size	22.93	5.702
Percentage of female students in class	0.48	0.177
Total teaching experience	10.13	5.363
Experience in the current school	3.330	2.232
Graduation field (whether related to math?)	98.8%	
Appointment field (whether related to math?)	99.8%	
Having master or higher degree	2.9%	

Table 1. *Cont.*

		Mean	SD
	Related to math	57.9%	
Field of master's (or higher) degree	Not related to math	17.6%	
	Unknown	24.45	
Class average test score in Grade 7		9.48	2.715
Class average test score in Grade 6		8.66	2.290
N (teacher)		230	
Panel C: School characteristics			
	General	84.3%	
Categories	Regional Boarding	3.5%	
	Vocational	12.2%	
Service score (1-highest, 6-lowest score)		2.01	1.433
	Rural	17.9%	
Location	Suburban	19.8%	
	Urban	62.3%	
School average test score in Grade 7		9.48	2.193
School average test score in Grade 6		8.67	1.823
N (school)		145	

2.3. Data Analysis

The study analyzed the contribution of various contextual variables at student, teacher/classroom and school levels on teacher effectiveness estimates obtained by VAM. The multiple linear regression analysis was performed using the forward-selection approach to assess the contribution of each of the contextual variables to the value-added estimates of mathematics teachers. By using the forward-selection approach, significant changes to the model-fit can be checked by including new predictor(s) in contrast with the model employed in the previous step.

In RQ1, the basic model was generated using just one prior achievement score (t-1) and then student characteristics (gender and language learner status) were added to the next stage in order to figure out how much change exists in the model-fit when utilizing the student characteristics. The highest R-squared value can be generated if all of the predictors are used in the model, so both gender and language learner identity variables were added to the basic model with the enter method, and the highest R-squared value that can be obtained in RQ1 was found. Finally, the same student characteristics were again added to the basic model using the forward-selection method. Once a model with the largest R-squared value (determined by enter method) was proposed, the model(s) to be suggested in the next step(s) (if any) was ignored in order to include only variables with a predictive impact on the estimates. The variables excluded from predictions, because they have no predictive ability, or are too small to be considered in this research question, will also not be included in the analysis of next two research questions.

Similarly, in RQ2, to determine which school characteristics make a worthwhile contribution to mathematics teachers' value-added estimates, the model obtained in research question 1 was used as the baseline model in this research question. The same steps were followed to create the final model in research question 2. For the last stage of this study, RQ3 was built on the model derived from RQ2, to which student and teacher-/classroom-level predictors were added. Unlike with the other previous two research questions, before conducting the regression analyses, the presence of any relationship between the teacher/classroom characteristics and the mathematics teachers' value-added effectiveness scores was checked. For this, individual student residual scores (the difference between predicted and actual attainment level), obtained through the final model proposed in RQ2, were aggregated at the teacher level. The mean of the residuals at the teacher level was attributed to a teacher's individual value-added effectiveness score. Then, the effectiveness scores were correlated with the teacher/classroom characteristics. After the correlation statistics, the model generated in RQ2 was used as the baseline model for this last research question to decide whether the characteristics of teacher

and classroom contribute greatly to the value-added estimates of mathematics teachers. Then, the ultimate regression model was created by pursuing the same steps in the previous research questions.

The *p*-value and confidence interval of statistical significance tests, such as the *t*-test or chi-square, etc., are still widely reported in the social sciences, regardless of whether the data ensures the assumptions required to use these statistics. As this current study involves a nonrandom sample from the study population that also has missing data, the main assumption of reporting the *p*-value is not met [39,40]. Therefore, as the key decision on analyzing the data, *p*-values of the significance tests used are not reported in this study [41–45].

3. Results

The analyses were mainly centered on the consistency of teacher value-added effectiveness estimates regarding model specifications. A series of analyses were conducted to answer each research question. Models were created by excluding and/or including contextual variables at student, school and teacher/classroom levels and changes in model fit were evaluated by checking changes in R^2 values for each research question by conducting the forward-selection method.

3.1. Research Question 1: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by Student Characteristics, Other Than Prior Attainment?

The students' math test scores in Grade 8 were used as the outcome variable, while their gender, language learner identity and seventh grade math test scores (prior attainment at Time 1) were employed as predictors.

The baseline model indicates that the previous achievement score of students alone can explain 47% (R^2 = 0.470) of the variability in Grade 8 math test scores (Table 2). Adding other student characteristics (gender and language status) to the baseline model using the enter method increased the R^2 value of the new model to 0.471 (see in Appendix A). This increase infers that students' gender and their language status contribute 0.1% to explaining the variance in their current test scores. To find out whether this minor change in the explanation of the variation in the result variable is due to both of the two variables, or just one of them, another regression analysis was performed by employing the forward method. The forward method suggested a model using the prior attainment and gender variables with the largest R^2 values, which was the same as the highest R^2 value achievable with the full model (0.471); therefore, the language learner identity variable was excluded from the final model created using the student-level variables.

Table 2. Value-added estimates for mathematics teachers using student characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment	0.470	
Forward method			
Model 1	Prior attainment	0.470	-
Model 2	Prior attainment Gender	0.471	0.001

Dependent Variable: 8th grade math test score.

3.2. Research Question 2: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by School Characteristics, over and above Student Characteristics?

In this research question, students' eighth grade maths test scores were employed as the outcome variable. Their seventh grade maths test scores, gender and five school-level variables: school categories (general, regional boarding, vocational), service scores (1—highest, 6—lowest score), locations (rural, suburban, urban), the school-level average maths test scores for seventh grade and school-level average maths test scores for sixth grade were used as predictors in this research question.

The final model created in the previous research question was used as a baseline model for this research question, where the eighth grade students' math test scores were regressed on students' prior attainment scores (t-1) and gender. All school-related variables were added to the baseline model of RQ2 at the same time using the enter method to discover the highest R-squared value achievable at the school level. Adding all school-level variables to the baseline model increased R-squared by 19 points (R^2 = 0.490) (see in Appendix B). To minimize uncertainty and consider the variables that have a predictive impact on estimates, it is important to find out whether the 19-point increase in R-squared of the full model is due to the inclusion of all, or only some of the characteristics. Therefore, a regression analysis using the forward-selection method was conducted using the same variables employed in the full model. The forward method suggested a model with the largest R^2 value, which was the same as the highest R^2 value achievable with the full model (R^2 = 0.490), where prior attainment, school-level average test scores in Grades 6 and 7 and student gender variables were employed. School categories, service scores and school location variables were excluded from the final model created for this research question about math teachers' value-added effectiveness estimates. These exclusions can also be interpreted as giving no indication that the students' current attainment in maths is linked to the school service score, the school location and the type of school attended, once the prior attainment, school-level average test scores in Grades 6 and 7 and student gender have been taken into account (Table 3).

Table 3. Value-added estimates for mathematics teachers adding school characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment Student gender	0.471	
Forward method			
Model 1	Prior attainment	0.470	−0.001
Model 2	Prior attainment 6th grade average school test score	0.487	0.016
Model 3	Prior attainment 6th grade average school test score 7th grade average school test score	0.489	0.018
Model 4	Prior attainment 6th grade average school test score 7th grade average school test score Student gender	0.490	0.019

Dependent Variable: 8th grade maths test score.

3.3. Research Question 3: To What Extent Can the Value-Added Estimates for Mathematics Teachers Be Explained by Teacher/Classroom Characteristics, over and above Student and School Characteristics?

The research question focused on mathematics teachers' value-added estimates regarding using teacher/classroom characteristics, in addition to student- and school-level variables identified in the previous research questions. Seven observable teacher characteristics were employed as the teacher-/classroom-level predictors in this research question: gender, number of years of teaching experience, number of years teaching in the current school, teachers' major degree subject, teaching assignment field and the highest level of qualification and field, in addition to four classroom-level variables: class size, percentage of female students, and sixth and seventh grade classroom-level average maths test scores.

To determine whether there is a relationship between the teacher/classroom characteristics and teacher effectiveness scores, individual math teachers' value-added effectiveness scores were estimated by aggregating each student's residual scores (the difference between predicted and actual attainment level) obtained through the final model proposed in RQ2 at the teacher level. Pearson's r coefficients show that there is no substantial association between the teacher effectiveness score

and the teacher-/classroom-level continuous variables (see Table 4). There was barely an association between class size and teacher VAM score and, surprisingly, bigger classes had slightly more "effective" teachers (r = 0.079). Another interesting finding is that classes with a higher female student ratio are taught by less effective mathematics teachers (r = −0.101). Experience, regardless of whether in total or just in their current schools, has a negative link with their effectiveness scores, so more experienced mathematics teachers tend to be assigned a lesser effectiveness score from value-added estimates. Lastly, with very small correlation coefficients, unsurprisingly, a positive relationship appeared for classroom-level prior attainment. However, the classroom average at a two lagged year (t-2, Grade 6) was a better predictor for the effectiveness of maths teachers than the average score at the prior year (t-1, Grade 7).

Table 4. Correlation between teacher/classroom characteristics and teacher value-added effectiveness.

Teacher/Classroom Characteristics	Correlation Coefficient (r)
Class size	0.079
Percentage of female students	−0.101
Total teaching experience	−0.029
Experience in the current school	−0.017
Classroom average test score in Grade 7	0.064
Classroom average test score in Grade 6	0.106

Similarly, Cohen's effect sizes, shown in Table 5, were also calculated for each subcategory of teacher characteristics variables. On average, the value-added scores of female maths teachers were marginally worse than those of male teachers (d = −0.10). Moreover, teachers with a bachelor's degree in a field related to mathematics tend to have lower effectiveness scores, even if the effect size is very fractional (d = −0.08). More interestingly, teachers who were first appointed as primary school teachers, but later became maths teachers, have remarkably higher value-added effectiveness scores than those appointed originally as mathematics teachers (d = −0.92). This result may be attributed to the unbalanced subcategories in this variable, so this result needs to be confirmed with a variable containing a balanced subcategorical distribution. Another interesting finding is that contrary to what is believed, having a master's degree does not contribute to the effectiveness estimates for mathematics teachers, and math teachers with master's degrees even had, on average, lower effectiveness scores than those teachers with just a bachelor's degree (d = −0.25). Finally, almost no link was found between having a master's degree in a math-related field and the VAM scores of mathematics teachers (d = 0.01).

Table 5. Comparison of means for teacher/classroom characteristics.

Category	Yes	No	Cohen's d
Gender (Female)	−0.015	0.017	−0.10
Graduation field (related to math)	−0.000	0.027	−0.08
Appointment field (related to math)	−0.001	0.298	−0.92
Terminal degree (master or higher degree)	−0.080	0.002	−0.25
Field of master's degree			
Related	0.003	−0.000	0.01
Not Related	−0.312	0.002	−0.96
Unknown	−0.109	0.001	−0.34

Overall standard deviation = 0.325.

After disclosing the association between teacher-/classroom-level characteristics and teacher VAM scores, as for previous research questions, a best-fit regression model was created and had the highest R-squared that can be obtained by using as few variables as possible.

The final model created in RQ2 was again used as a baseline model in this research question (see Table 6). Using the enter method, all teacher-/classroom-related variables were added to this baseline

model at the same time to discover the highest R-squared value achievable at teacher/classroom levels. Adding all teacher-/classroom-level characteristics to the baseline model increased the variance explained by 12 points (R^2 = 0.502). To include only variables that have predictive impacts on estimates, another regression analysis was carried out with the forward method, employing the same predictors used in the full model.

Table 6. Value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Model	Predictors Used	R-Squared (R^2)	Changes in R^2
Baseline Model	Prior attainment 6th grade average school test score 7th grade average school test score Student gender	0.490	
Forward method			
Model 1	Prior attainment	0.470	−0.020
Model 2	Prior attainment 6th grade average classroom test score	0.497	0.007
Model 3	Prior attainment 6th grade average classroom test score 7th grade average classroom test score	0.500	0.010
Model 4	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender	0.501	0.011
Model 5	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender Class size	0.501	0.011
Model 6	Prior attainment 6th grade average classroom test score 7th grade average classroom test score Student gender Class size Percentage of female students	0.502	0.012

Dependent Variable: 8th grade maths test score.

The forward method proposed a model with the largest R^2 value achievable at teacher/classroom levels (revealed by the full model), which includes students' prior attainment scores (t-1), classroom-level average test scores in Grades 6 and 7, student gender, class size and percentage of female students' variables (R^2 = 0.502) (see Appendix C). School-level variables and sixth and seventh grade average school test scores were included in the final model in RQ2. However, once the teacher/class characteristics had been taken into account, these variables' predictive power on value-added estimates disappeared. Therefore, these variables were removed from the final model of this research question. In addition, teacher gender, major degree subject, teaching field, number of years of teaching experience, number of years teaching in the current school and highest level of qualification and field variables were also excluded from the final model, as they did not contribute to the variance explained in the students' current test scores. Interestingly, these exclusions show that none of the teacher characteristics are considerably related to the achievements of students, but all classroom characteristics are included in the final model. Therefore, it can be interpreted that student achievement is more related to class characteristics than the observable characteristics of the teacher.

The full list of standardized coefficients for the predictors employed in the final model were also investigated (see Table 7). The overall conclusion is that when the hold of the other factors is constant, there is the largest positive relationship between students' prior mathematics attainment and their recent mathematics outcomes; for each correct answer increase in the prior mathematics test score, the number of the correct answers in the Grade 8 test increases by 0.615. The second largest

relationship was found between the eighth grade maths score and the class average maths score at two lagged years (t-2). For each correct answer increase in the sixth grade maths average classroom test score, the recent maths attainment increases on average by 0.269. Interestingly, a negative relationship was revealed between the nearest prior year's classroom average (t-1) and the eighth grade math scores. When the negative standardized coefficient accounts for each one point increase in the seventh grade maths average classroom test score, the recent maths attainment decreases by 0.118, on average. In the current maths test, it is expected that females would be 0.039 points higher than male students. Another unexpected conclusion related to class size was that the standardized coeffect indicated that, on average, larger classes were more successful in the eighth grade maths test. Finally, although female students were, on average, more successful in the mathematics test in Grade 8, the classes with more female students were less successful in the same test. For each increase in the percentage of female students in the classroom, the number of correct answers in the eighth grade maths test decreases by 0.024.

Table 7. Standardized regression coefficients for the variables used in the final model.

Variables	Standardized Coefficient β
Prior attainment (t-1)	0.615
6th grade average classroom test score	0.269
7th grade average classroom test score	−0.118
Student gender (female)	0.039
Class size	0.026
Percentage of female students	−0.024

$R^2 = 0.502$.

4. Discussion

A student's academic achievement is linked with various student-related, teacher-/classroom-related and school-related factors and in a growing number of studies there is a consensus that, among school-related resources, the most crucial factor is teachers [1–3,46]. Consequently, the accountability of teachers for students' school achievement is among the educational issues on which policymakers and researchers have recently focused.

However, teacher effectiveness is not an easy attribute to measure. Recently, measures of teacher effectiveness have relied on more objective measures, such as student performance in high-stakes tests. The use of student data in measuring teacher performance is one of the most controversial and important issues of educational policy. Significant decisions affecting teachers' professional careers, such as pay raises, promotions, or redundancies, are made based on such performance evaluations. Schools and teachers are penalized and even shamed based on such measures. A well known example of a teacher performance evaluation measure based on student achievement is that of value-added models. VAMs, theoretically, isolate a particular teacher's relative effectiveness on his/her students' achievement from other factors outside the teacher's control. These include student, school, teacher, classroom characteristics and students' prior year test scores. The main purpose of this study is to examine the contribution of contextual predictors, such as observable student, teacher/classroom and school characteristics to teacher value-added effectiveness estimates.

Consistent with the findings of the literature review, this study suggests that the strongest student-related factor in explaining the variation in a student's current test score is their nearest prior attainment (math score at Grade 7). The results show that approximately half of the variance in the students' Grade 8 math test scores (47%) can be explained by their Grade 7 math results alone. The literature in the review also agrees that the strongest contribution to value-added estimates is from previous years' test scores. For example, Hu [12] wrote that the nearest prior year's achievement score alone accounted for an average of 57% and 59% of the variance in students' current achievements in

math and reading, respectively, while Kersting et al. [6] found that 68% of the variance in the students' current scores was explained by controlling only test scores from one previous year.

Another conclusion reached in this study is that, in agreement with the literature, student gender is not an important factor in explaining differences in the teacher effectiveness estimates. Similar to this current study, Heistad [22] found that adding gender to a model that already controlled for students' prior attainment and race increased the explanatory power by 0.1% to 0.4% depending on the testing year. Tobe [26] excluded gender in the analyses as it did not make a statistically significant contribution to explaining the variance in their model. This current study of teacher effectiveness estimates, which employed data from Turkey, came to the same conclusion as the studies in the review—that is, that the strongest predictor in explaining students' test scores in the current year is their prior test scores, and a student's gender is not an important factor in the value-added teacher effectiveness estimates when the prior attainment is controlled [12,13,21].

Adding the language learner identity variable (i.e., whether Turkish is a second language or not) in value-added estimates also showed no impact on students' test scores. This variable was therefore removed from further analysis. This is likely because only a very small proportion (0.2%) of the students in the study population identified Turkish as their second language. However, the literature in the review also revealed that students' language identity did not considerably contribute to the teacher value-added effectiveness estimates [12,20,21,23].

Regarding changes in the R^2 value, the findings suggest that the most important school-related variables, in terms of explaining the variation in students' current test scores, are the average math test scores in Grades 6 and 7. Adding sixth and seventh grade average math test scores to the model obtained in RQ1 increased the R^2 of the new model to the highest R^2 value that could be achieved with the school characteristics. Therefore, it was determined that the other three characteristics, category, service scores and location, do not have a considerable predictive impact on value-added estimates. This result can also be interpreted to mean that there is no indication that students' current attainment in math is linked to school service score, school location and the type of school attended, once their prior attainment, sixth and seventh grade average test scores and gender have been taken into account. Contrary to this, Sander and Horn [47] summarized previous research findings and reported that cumulative academic gains for schools in the entire state are unrelated to schools' average attainment. Supporting the finding about the ineffectiveness of school-level variables, the retrieved literature on the effectiveness of teachers also indicates that school-level measures, the percentage of students receiving free/reduced-price lunches, class size, racial/ethnic composition, students with special educational needs and those with English as a second language (ESL) accounted for very little of the variance in student attainment [33]. Ballou et al. [32] suggest that controlling the percentage of students eligible for free/reduced-price lunches in schools has a substantial impact on TVAAS (the Tennessee Value-Added Assessment System) estimates in some grades and subjects; however, the authors also emphasize the precision of the models used and advise caution in terms of this finding. Alban [14], using the socio-economic level of the school, the percentage of students receiving special services, enrolment, and mobility and ethnicity as school-level predictors, found that prior achievement is the only significant variable in each estimation, followed by the percentage of students receiving special education.

Many recent studies report that the value-added estimates of teachers are loosely tied to the observable characteristics of teachers. To determine whether a relationship exists between teacher/classroom characteristics and the effectiveness of mathematics teachers, the residual scores of individual students were aggregated at the teacher level, and the aggregated residuals at the teacher level (teacher effectiveness score) were then correlated with teacher/classroom characteristics. The results suggest that no considerable relationship exists between teacher effectiveness and the teacher-/classroom-level variables. Interestingly, the experience of teacher, regardless of whether over their total career or only in their current schools, has a negative relationship with the effectiveness scores of teachers; thus, more experienced mathematics teachers tend to acquire lower effectiveness scores from value-added estimates. The findings related to teacher/classroom characteristics are consistent

with previous research in that no clear relationship can be found between teacher/class characteristics and teacher effectiveness [24,29,48].

The existing literature presents mixed findings on experience; therefore, the association between experience and effectiveness remains unanswered. In the current research, both total and current teacher experience have a slightly negative relationship with teachers' effectiveness scores. Conversely, Wayne and Youngs [49] found that most studies claim a positive relationship in this respect [34,50]. More robust evidence is therefore needed to ascertain whether a close relationship exists between teacher effectiveness and experience.

Aside from correlation analyses, the changes in R^2 indicate that the most important teacher-/classroom-level variables in explaining variations in students' current test scores are classroom-level average test scores in Grades 6 and 7, class size and the percentage of variables associated with female students. The exclusion of the other teacher-/classroom-level variables, teacher gender, graduation field, teaching assignment field, total teaching experience, experience in the current school, terminal degree and field of terminal degree indicate the lack of obvious links with students' current math attainment when the proposed predictors in the final model (model 6) were used. Interestingly, although the final model contains all classroom-level characteristics, consistent with a large body of literature, the findings also reveal that none of the teacher characteristics are directly linked to students' attainment in mathematics [1,13,18,27,29,31,33].

Nye et al. [30] examined the impact of teacher characteristics, experience and education on teacher effectiveness estimates and found that none of them have a considerable impact. Their contribution to the variance explained in each estimate never exceeded 5%. Conversely, few studies have revealed the impacts of certain teacher characteristics on the effectiveness estimates. For instance, Kukla-Acevedo [24] found that among teacher characteristics employed, only the overall undergraduate performance (GPA) of teachers in math has a consistent and positive impact on students' mathematics achievement. Goel and Barooah [25] suggested that only permanent employment status (tenured) positively affects teacher effectiveness estimates. Moreover, Tobe [26] reported that apart from certification by the state, none of the other teacher characteristics have an impact on students' attainment. Munoz et al. [51] also discovered that teacher experience is the only teacher characteristic that has predictive power on teacher effectiveness estimates.

The finding related to class size based on the standardized coefficient is another surprising conclusion drawn from this current study. Only a modest correlation was found between class size and teacher effectiveness, and, interestingly, larger classes performed better in eighth grade math tests on average. Taking into account debates on class size, with support from more precise evidence, this finding may contribute to efforts towards a reconsideration of the policy of class size reduction, which involves considerable costs [2,52]. On the other hand, completely ignoring this policy is not advocated, considering the existing evidence from many studies suggesting that reducing class size can increase the academic success of students [53,54].

To sum up, the study reveals that, although approximately half of the variance in student current math test scores can be explained by their prior attainment alone (47%), including other contextual predictors, such as teacher or school characteristics, makes a very limited contribution to teachers' value-added estimates. Classroom-level average test scores in one and two lagged years, student gender, class size and percentage of female students variables are the other contextual predictors that contribute to the value-added effectiveness estimates, but R^2 change analyses suggests that the contribution of the contextual variables to the variance explained in math has never exceeded 4%. This result shows that student achievement mostly depends on the performance of the student in early education. The focus of education and investment, therefore, should clearly be in the early years.

Funding: This research was funded by the Ministry of National Education, Turkey.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The author declares no conflict of interest.

Appendix A

Full model of value-added estimates for mathematics teachers using student characteristics.

Table A1. Value-added estimates for mathematics teachers using student characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment Gender Language Learner ID	0.471

Appendix B

Full model of value-added estimates for mathematics teachers adding school characteristics.

Table A2. Value-added estimates for mathematics teachers adding school characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment Student gender School categories Service score School locations 7th grade average school test score 6th grade average school test score	0.490

Appendix C

Full model of value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Table A3. Value-added estimates for mathematics teachers adding teacher/classroom characteristics.

Model	Predictors Used	R-Squared (R^2)
Full Model	Prior attainment 6th grade average school test score 7th grade average school test score Student gender Teacher gender Class size Percentage of female students Graduation field Appointment field Total teaching experience Experience in the current school Terminal degree Field of terminal degree 7th grade average classroom test score 6th grade average classroom test score	0.502

References

1. Aaronson, D.; Barrow, L.; Sander, W. Teachers and student achievement in the Chicago public high schools. *J. Labor Econ.* **2007**, *25*, 95–135. [CrossRef]
2. Rivkin, S.G.; Hanushek, E.A.; Kain, J.F. Teachers, schools, and academic achievement. *Econometrica* **2005**, *73*, 417–458. [CrossRef]
3. Sanders, W.L.; Wright, S.P.; Horn, S.P. Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *J. Pers. Eval. Educ.* **1997**, *11*, 57–67. [CrossRef]

4. See, B.H. Challenges in Using research evidence in improving teaching quality. *Bera Res. Intell.* **2020**, *144*, 20–21.
5. Wei, H.; Hembry, T.; Murphy, D.L.; McBride, Y. *Value-Added Models in the Evaluation of Teacher Effectiveness: A Comparison of Models and Outcomes*; Pearson: New York, NY, USA, 2012.
6. Kersting, N.B.; Chen, M.; Stigler, J.W. Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Educ. Policy Anal. Arch.* **2013**, *21*, 1–39.
7. Rothstein, J. *Can Value-Added Models Identify Teachers' Impacts*; (Policy Brief); Institute for Research on Labor and Employment: Berkeley, CA, USA, 2016.
8. McCaffrey, D.F.; Koretz, D.; Lockwood, J.R.; Hamilton, L.S. *The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness*; RAND Corporation: Boston, MA, USA, 2004.
9. Pivovarova, M.; Amrein-Beardsley, A.; Broatch, J. Value-added models (VAMs): Caveat emptor. *Stat. Public Policy* **2016**, *3*, 1–9. [CrossRef]
10. Rothstein, J. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Q. J. Econ.* **2010**, *125*, 175–214.
11. Baker, E.L.; Barton, P.E.; Darling-Hammond, L.; Haertel, E.; Ladd, H.F.; Linn, R.L.; Ravitch, D.; Rothstein, R.; Shavelson, R.J.; Shepard, L.A. *Problems with the Use of Student Test Scores to Evaluate Teachers*; Economic Policy Institute: Washington, DC, USA, 2010.
12. Hu, J. Teacher Evaluation Based on an Aspect of Classroom Practice and on Student Achievement: A Relational Analysis between Student Learning Objectives and Value-Added Modeling. Ph.D. Thesis, Boston College, Boston, MA, USA, 2015.
13. Rothstein, J. Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Educ. Financ. Policy* **2009**, *4*, 537–571. [CrossRef]
14. Alban, T.R. Evaluating School and Teacher Effectiveness: A Comparison of Analytic Models. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2002.
15. Gagnon, D.J. Understanding the Distribution of Teacher Effectiveness. Ph.D. Thesis, University of New Hampshire, Durham, NH, USA, 2014.
16. Gallagher, H.A. The Relationship between Measures of Teacher Quality and Student Achievement: The Case of Vaughn Elementary. Ph.D. Thesis, The University of Wisconsin–Madison, Madison, WI, USA, 2002.
17. Kane, T.J.; Staiger, D.O. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation*; The National Bureau of Economic Research: Cambridge, MA, USA, 2008.
18. Slater, H.; Davies, N.M.; Burgess, S. Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxf. Bull. Econ. Stat.* **2012**, *74*, 629–645. [CrossRef]
19. Chetty, R.; Friedman, J.N.; Rockoff, J.E. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *Am. Econ. Rev.* **2014**, *104*, 2593–2632. [CrossRef]
20. Cunningham, P.L. The Effects of Value-Added Modeling Decisions on Estimates of Teacher Effectiveness. Ph.D. Thesis, The University of Iowa, Iowa City, IA, USA, 2014.
21. Ehlert, M.; Koedel, C.; Parsons, E.; Podgursky, M. The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Stat. Public Policy* **2014**, *1*, 19–27. [CrossRef]
22. Heistad, D.J. Stability and Correlates of Teacher Effects in Grade Two Reading Achievement. Ph.D. Thesis, University of Minnesota, Minneapolis, MN, USA, 1999.
23. Johnson, M.T.; Lipscomb, S.; Gill, B. Sensitivity of teacher value-added estimates to student and peer control variables. *J. Res. Educ. Eff.* **2015**, *8*, 60–83. [CrossRef]
24. Kukla-Acevedo, S. Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Econ. Educ. Rev.* **2009**, *28*, 49–57. [CrossRef]
25. Goel, D.; Barooah, B. *Drivers of Student Performance: Evidence from Higher Secondary Public Schools in Delhi*; Global Labor Organization (GLO): Geneva, Switzerland, 2018.
26. Tobe, P.F. An Investigation of the Differential Impact of Teacher Characteristics and Attitudes on Student Mathematics Achievement Using a Value-Added Approach. Ph.D. Thesis, University of Houston, Houston, TX, USA, 2008.

27. Goldhaber, D.; Hansen, M. *Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions*; National Center for Analysis of Longitudinal Data in Education Research (CALDER): Washington, DC, USA, 2010.
28. Buddin, R. *Measuring Teacher and School Effectiveness at Improving Student Achievement in Los Angeles Elementary Schools*; University Library of Munich: München, Germany, 2011.
29. Muñoz, M.A.; Chang, F.C. The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *J. Pers. Eval. Educ.* **2007**, *20*, 147–164. [CrossRef]
30. Nye, B.; Konstantopoulos, S.; Hedges, L.V. How large are teacher effects? *Educ. Eval. Policy Anal.* **2004**, *26*, 237. [CrossRef]
31. Leigh, A. Estimating teacher effectiveness from two-year changes in students' test scores. *Econ. Educ. Rev.* **2010**, *29*, 480–488. [CrossRef]
32. Ballou, D.; Sander, W.; Wright, P. Controlling for student background in value-added assessment of teachers. *J. Educ. Behav. Stat.* **2004**, *29*, 37–65. [CrossRef]
33. Germuth, A.A. Comparing Results from Value-Added HLM and OLS Models to Assess Teacher Effectiveness. Ph.D. Thesis, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 2003.
34. Clotfelter, C.T.; Ladd, H.F.; Vigdor, J.L. Teacher-student matching and the assessment of teacher effectiveness. *J. Hum. Resour.* **2006**, *XLI*, 778–820. [CrossRef]
35. Goldhaber, D. Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *J. Hum. Resour.* **2007**, *42*, 765–794. [CrossRef]
36. Step by Step Achievement Project. Available online: https://samsun.meb.gov.tr/www/adim-adim-basari-projesi/icerik/570 (accessed on 10 November 2017).
37. Schools and Other Institutions. Available online: https://www.meb.gov.tr/baglantilar/okullar/index.php (accessed on 11 November 2017).
38. The List of Institutions. Available online: https://mebbis.meb.gov.tr/KurumListesi.aspx (accessed on 11 November 2017).
39. Figueiredo Filho, D.B.; Paranhos, R.; Rocha, E.C.; Batista, M.; Silva, J.A.; Santos, M.L.W.D.; Marino, J.G. When is statistical significance not significant? *Braz. Political Sci. Rev.* **2013**, *7*, 31–55. [CrossRef]
40. Gorard, S. Significance testing with incompletely randomised cases cannot possibly work. *Int. J. Sci. Res. Methodol.* **2018**, *11*, 42–51.
41. Cohen, J. The earth is round ($p < 0.05$). *Am. Psychol.* **1994**, *49*, 997–1003.
42. Gorard, S. Damaging real lives through obstinacy: Re-emphasising why significance testing is wrong. *Sociol. Res. Online* **2016**, *21*, 102–115. [CrossRef]
43. Gorard, S. Do we really need confidence intervals in the new statistics? *Int. J. Soc. Res. Methodol.* **2019**, *22*, 281–291. [CrossRef]
44. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350. [CrossRef]
45. White, P.; Gorard, S. Against inferential statistics: How and why current statistics teaching gets it wrong. *Stat. Educ. Res. J.* **2017**, *16*, 5565.
46. Opper, I.M. Does helping John help Sue? Evidence of spillovers in education. *Am. Econ. Rev.* **2019**, *109*, 1080–1115. [CrossRef]
47. Sanders, W.L.; Horn, S.P. Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *J. Pers. Eval. Educ.* **1998**, *12*, 247–256. [CrossRef]
48. Buddin, R.; Zamarro, G. Teacher qualifications and student achievement in urban elementary schools. *J. Urban. Econ.* **2009**, *66*, 103–115. [CrossRef]
49. Wayne, A.J.; Youngs, P. Teacher characteristics and student achievement gains: A review. *Rev. Educ. Res.* **2003**, *73*, 89–122. [CrossRef]
50. Rice, J.K. *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*; Economic Policy Institute: Washington, DC, USA, 2003.
51. Munoz, M.A.; Prather, J.R.; Stronge, J.H. Exploring teacher effectiveness using hierarchical linear models: Student- and classroom-level predictors and cross-year stability in elementary school reading. *Plan. Chang.* **2011**, *42*, 241–273.

52. Hanushek, E.A. The failure of input-based schooling policies. *Econ. J. Lond.* **2003**, *113*, F64–F98. [CrossRef]
53. Krueger, A.B. Economic Considerations and Class Size. *Econ. J. Lond.* **2003**, *113*, F34–F63. [CrossRef]
54. Nye, B.A.; Archilles, C.M.; Boyd-Zaharias, J.; Fulton, B.; Wallenhorst, M. Small is far better. *Res. Sch.* **1994**, *1*, 9–20.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).