Article

Educational Policy
2019, Vol. 33(7) 1103–1134

DOI: 10.1177/0895904818755468
journals.sagepub.com/home/epx

Teacher Characteristics and Student Learning in Mathematics: A Comprehensive Assessment

Heather C. Hill[1], Charalambos Y. Charalambous[2], and Mark J. Chin[1]

Abstract

Diverse stakeholders have an interest in understanding how teacher characteristics—their preparation and experience, knowledge, and mind-sets and habits—relate to students' outcomes in mathematics. Past research has extensively explored this issue but often examined each characteristic in isolation. Drawing on data from roughly 300 fourth- and fifth-grade teachers, we attend to multiple teacher characteristics and find that experience, knowledge, effort invested in noninstructional activities, and participation in mathematics content/methods courses predict student outcomes. We also find imbalances in key teacher characteristics across student populations. We discuss the implications of these findings for hiring and training mathematics teachers.

Keywords

achievement, mathematics education, teacher quality, teacher qualifications

[1]Harvard University, Cambridge, MA, USA
[2]University of Cyprus, Nicosia, Cyprus

Corresponding Author:
Mark J. Chin, Center for Education Policy Research, 50 Church Street, 4th Floor, Cambridge, MA 02138, USA.
Email: mark_chin@g.harvard.edu

Identifying effective teachers has become a pressing need for many state and district policy makers. Federal legislation, such as No Child Left Behind and Race to the Top, required local educational agencies (LEAs) to both measure teacher quality and take action to dismiss or remediate those that fall below a bar for sufficient quality. LEAs themselves may wish to better identify effective and ineffective teachers, as past research has shown teacher quality to be one of the strongest institutional-level predictors of student outcomes (e.g., Chetty, Friedman, & Rockoff, 2014). Yet for LEAs interested in improving the quality of teacher candidates at entry into their system, and in improving the quality of the many teachers without student test scores and/or classroom observations, determining teacher quality can be challenging. The same is true for LEAs interested in strategically placing new teachers in certain schools to more equitably distribute teacher quality.

Such LEAs might turn to the literature on the relationship between teacher characteristics and student outcomes. The factors tested by this literature fall into three broad categories, including the preparation and experiences hypothesized to be important to teachers (e.g., postsecondary mathematics coursework, degrees, teaching experience, certification type; Clotfelter, Ladd, & Vigdor, 2007; Harris & Sass, 2011; Monk, 1994); teacher knowledge (e.g., mathematical knowledge for teaching, knowledge of students' mathematical abilities and misconceptions; Bell, Wilson, Higgins, & McCoach, 2010; Carpenter, Fennema, Peterson, & Carey, 1988; Hill, Rowan, & Ball, 2005); and teacher mind-sets and habits (e.g., efficacy, locus of control, effort invested in teaching; Bandura, 1997; Tschannen-Moran & Hoy, 2001). Across these categories, researchers have measured and tested numerous variables against student learning on standardized assessments.

While many of the studies noted above show small positive associations between measured characteristics and student outcomes, results for specific variables are decidedly mixed across the field as a whole, making the job of practitioners who want to take lessons from this literature difficult. Furthermore, most studies in this genre tend to specialize in only one category, restricting the number and type of variables tested and thus potentially obscuring important relationships. Economists, for instance, typically examine the association between student outcomes and teacher background characteristics, such as experience and degrees, by relying on district administrative data; as this implies, such papers tend not to include predictors that require data collection efforts. Studies of teachers' knowledge, on the contrary, must collect large amounts of original data to measure that knowledge, and, perhaps as a result, rarely capture more than a handful of other key variables (e.g., Baumert et al., 2010; Hill et al., 2005). The same holds for studies of teacher mind-sets and habits, as well; in fact, seldom does a single study test more than one class of characteristics (for exceptions, see Boonen, Van Damme, & Onghena, 2014; Campbell et al., 2014; Grubb, 2008; Palardy & Rumberger, 2008). Yet it seems likely that many of these factors correlate; teachers with stronger postsecondary mathematical preparation, for instance, may also have stronger mathematical knowledge and may feel more capable of teaching the subject.

To untangle these relationships, we argue for a more comprehensive comparison of these characteristics, with the aim of understanding how they relate to one another and, either individually or jointly, contribute to student outcomes. Based on evidence that teacher characteristics often vary according to the student populations they serve (e.g., Goldhaber, Lavery, & Theobald, 2015; Hill et al., 2005; Jackson, 2009; Lankford, Loeb, & Wyckoff, 2002), we also argue for an investigation of whether key teacher characteristics are distributed equitably across students and schools.

To accomplish this aim, we draw on data from roughly 300 fourth- and fifth-grade teachers of mathematics. When examining selected teacher characteristics from the three categories mentioned above, we found low to moderate correlations with student outcomes, consonant with the prior literature. Our exploration also pointed to imbalances in key teacher characteristics across student populations. In the remainder of the article, we first review evidence of each of the three categories of variables considered herein. After presenting our research questions, we outline the methods we pursued in addressing them. In the last two sections, we present the study findings and discuss their implications for teacher hiring, placement, and training.

Background

Teacher characteristics may contribute to student outcomes in important ways. Background variables, such as college- or graduate-level coursework, may demarcate experiences that provide teachers with access to curriculum materials, classroom tasks and activities, content standards, and assessment techniques that can shape practice and thus student outcomes. Experiences on the job may similarly build skill in using these elements in practice. Teacher knowledge may also inform instruction and student outcomes (Shulman, 1986); this knowledge may be acquired in coursework, in on-the-job learning, or may simply stem from differences among individuals as they enter the profession. Teacher mind-sets and work habits may contribute to student outcomes independently of the prior two categories; for instance, teachers may view mathematics as a fixed set of procedures to be learned, or as a set of practices in which to engage students, thus directly shaping their instruction. In addition, teachers may hold different views of responsibility for learning and expend different levels of effort toward securing positive student outcomes.

Below, we review evidence connecting student mathematics learning to teacher characteristics. We include a review of the limited studies that have addressed multiple categories and conclude with evidence regarding the distribution of such teacher characteristics over the population of students served by schools.

Teacher Preparation and Experience

Education production function studies have focused extensively on teachers' preparation and experience—including the number of years taught, preparation route, degrees obtained, certification type, and postsecondary coursework—as predictors of student outcomes. Among these variables, only teacher experience has shown a consistent positive relationship with student outcomes, with the gains to experience most pronounced in the early years of teaching (Boonen et al., 2014; Chetty et al., 2014; Clotfelter et al., 2007; Grubb, 2008; Harris & Sass, 2011; Kane, Rockoff, & Staiger, 2008; Papay & Kraft, 2015; Rice, 2003). Teacher attainment of a bachelor's or master's degree in education has mostly failed to show a relationship to student outcomes (e.g., Clotfelter et al., 2007; Harris & Sass, 2011; Wayne & Youngs, 2003; for an exception for master's degrees, see Guarino, Dieterle, Bargagliotti, & Mason, 2013). Findings for other variables are mixed, including for earned degrees (e.g., Aaronson, Barrow, & Sander, 2007; Harris & Sass, 2011; Rowan, Correnti, & Miller, 2002) and certification (for a review, see Cochran-Smith et al., 2012) and postsecondary mathematics content and mathematics methods coursework (e.g., Begle, 1979; Harris & Sass, 2011; Hill et al., 2005; Monk, 1994; Rice, 2003; Wayne & Youngs, 2003). For these latter variables, Harris and Sass (2011) found that neither mathematics content courses nor mathematics methods courses related to student outcomes for elementary teachers; similarly, in Hill and colleagues (2005), such courses did not predict student outcomes for elementary teachers, once controlling for teacher knowledge. In contrast, a research synthesis by Wayne and Youngs (2003) found that secondary students learn more from teachers with postsecondary and postgraduate coursework related to mathematics, while in Begle's (1979) meta-analysis, taking mathematics methods courses produced the highest percent of positive effects of all factors examined in his work. Notably, only a few of these studies (Chetty et al., 2014; Clotfelter et al., 2007; Harris & Sass, 2011; Papay & Kraft, 2015) examined within-teacher variation over time and employed designs that supported making causal inferences.

Teacher Knowledge

A number of studies have empirically linked aspects of teacher mathematical knowledge to student outcomes (for a discussion of teacher knowledge types, see Shulman, 1986). Some studies have focused on mathematics content knowledge in its relatively pure form, finding an association between teachers' competence in basic mathematics skills and student outcomes (e.g., Metzler & Woessmann, 2012; Mullens, Murnane, & Willett, 1996). Other studies, drawing on Shulman's (1986) framework, have shown that teachers' pedagogical content knowledge in mathematics better predicts student learning than pure, or basic, content knowledge (Baumert et al., 2010; Campbell et al., 2014). Still other studies have focused on measuring Ball, Thames, and Phelps's (2008) notion of specialized content knowledge, finding that students perform better when their teacher has more capacity to provide mathematical explanations, evaluate alternative solution methods, and visually model the content (Hill, Kapitula, & Umland, 2011; Hill et al., 2005; Rockoff, Jacob, Kane, & Staiger, 2011). A third category of studies has measured teachers' knowledge of students, including the accuracy with which teachers can predict their students' performance (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Helmke & Schrader, 1987) and the extent to which teachers can recognize, anticipate, or interpret common student misconceptions (Baumert et al., 2010; Bell et al., 2010; Carpenter et al., 1988). These knowledge-of-students measures typically either form a composite measure of teacher knowledge, meaning their unique effect on student learning cannot be disentangled, or they only inconsistently predict student outcomes. From all these studies, only one (Metzler & Woessmann, 2012) used a rigorous design, studying within-teacher variation over two subject matters and thus making causal claims about the effect of teacher knowledge on student performance.

Teacher Mind-Sets and Habits

Scholars have also extensively explored teacher mind-sets and habits as possible contributors to student outcomes, with strong interest in teachers' attitudes about what constitutes disciplinary knowledge, beliefs about how instruction should occur, and levels of enthusiasm and confidence about subject matter (Begle, 1979; Ernest, 1989; Fang, 1996). Given the growing interest around intelligence and mind-sets not as fixed entities but as traits that can be developed through effort and persistence (Dweck, 2006; Molden & Dweck, 2006), scholars have again turned to investigating the role of mind-sets and habits for learning, not only for students, but for teachers alike. In the latter case, scholars have taken up whether teachers believe in their capacities to influence the learning of all their students through their teaching. One such construct is perceived efficacy, defined as teachers' perceptions of their ability to organize and execute teaching that promotes learning (Bandura, 1997; Charalambous, Philippou, & Kyriakides, 2008). Teacher efficacy beliefs have consistently positively related to teachers' behavior in the classroom and the quality of their instruction (Justice, Mashburn, Hamre, & Pianta, 2008; Stipek, 2012; Tschannen-Moran & Hoy, 2001; Tschannen-Moran, Hoy, & Hoy, 1998). They have also predicted students' learning outcomes, both cognitive (e.g., Palardy & Rumberger, 2008) and affective (e.g., Anderson, Greene, & Loewen, 1988; Soodak & Podell, 1996).

Teacher locus of control taps the extent to which teachers feel they can influence their students' outcomes or whether, alternatively, they believe that those outcomes mostly hinge on non-classroom and school factors (e.g., students' socioeconomic background and parental support). Drawing on Rotter's (1966) work on internal versus external control of reinforcement, researchers have explored the relationship between teachers' locus of control and student learning, often documenting a positive relationship between the two constructs (Berman & McLaughlin, 1977; Rose & Medway, 1981).

A central tenet in the growth mind-set theory is the role that effort plays in improvement. However, scholars have studied teacher effort less extensively than efficacy and locus of control. In the only study in this category that used a research design that allowed for making causal inferences, Lavy (2009) observed that teacher effort mediates the positive impact of teacher merit pay on students' outcomes; survey data suggest that after-school tutoring plays a strong role, in particular, in producing improved outcomes. However, to our knowledge, few other studies have captured this variable.

Studies Measuring Multiple Categories

Within each of the three categories above, scholars have identified teacher characteristics that significantly and positively predict student outcomes. However, we could locate only four studies that examine the relationship between student outcomes and variables in more than one category; in all these studies, characteristics from different categories were concurrently fit into the same model(s). Investigating teacher preparation and experience, teacher attitudes, and self-reported practices, Palardy and Rumberger (2008) found that some self-reported instructional practices and attitudes (i.e., efficacy) predicted student outcomes, but teacher background characteristics did not. In Boonen and colleagues' (2014) work, teacher experience and job satisfaction—a background characteristic and attitude, respectively—predicted

Flemish students' mathematics outcomes. Grubb (2008) reported positive relationships between a variety of teacher background and preparation characteristics (e.g., experience, teaching in-field, education track), teacher efficacy beliefs, and student outcomes in mathematics in the NELS:88 data. Finally, Campbell et al. (2014) found that teacher knowledge positively associated with student outcomes, special education certification negatively associated with those outcomes, and teacher attitudes and beliefs largely had no effects outside interactions with knowledge itself.

Studies Examining the Distribution of Teacher Characteristics Across Students

Schools serving higher proportions of non-White and impoverished students have traditionally employed less qualified teachers, where qualified has been defined as individuals who are fully certified, hold advanced degrees, have prior teaching experience, and score better on certification exams or other standardized teaching-related exams (Hill and Lubienski, 2007; Jackson, 2009; Lankford et al., 2002). Such studies often used U.S. state or national-level data to examine teacher sorting; scholars have less often studied sorting within urban districts or metropolitan areas. Choi (2010) demonstrated that disadvantaged minorities and free/reduced-price lunch (FRPL) recipients in the Los Angeles Unified School District received instruction from, on average, less qualified teachers. Schultz (2014) replicated this result for the St. Louis metropolitan area, while a recent report from the Organization of Economic Co-Operation and Development (Schleicher & Organisation for Economic Co-Operation and Development, 2014) corroborated this finding internationally, reporting higher concentrations of unqualified teachers in schools serving disadvantaged students in several countries, including Belgium, Chile, the Czech Republic, Iceland, Luxembourg, the Netherlands, and Slovenia.

Synthesizing results from three studies that collectively examined the distribution of effective teachers across schools and districts across 17 U.S. states, Max and Glazerman (2014) have also reported that, on average, disadvantaged students in Grades 3 to 8 receive less effective teaching in mathematics than their counterparts. This difference amounted to 2 weeks of learning (which was calculated to be equivalent to 2%-3% of the achievement gap between disadvantaged and nondisadvantaged students) and varied across districts from less than a week in some districts to almost 8 weeks in others. Middle-grade students in the lowest poverty schools were additionally twice as likely to get teachers with value-added scores in the top 20% of their district compared with their counterparts in the highest poverty schools.

In a more recent study, Goldhaber et al. (2015) examined this issue more comprehensively by using a set of indicators of teacher quality—experience, licensure exam scores, and value-added estimates of effectiveness—across a set of indicators of student disadvantage, including FRPL status, underrepresented minority, and low prior academic performance. Although focusing on just one U.S. state, their study showed unequal distribution of almost every single indicator of teaching quality across every indicator of student disadvantage, a pattern that held for every school level examined (elementary, middle school, and high school). Much of this inequitable distribution of teaching quality appears to be due to teacher and student sorting across districts and schools rather than to inequitable distribution across classrooms within schools. This key role of the teacher and student sorting across districts appears in a recent brief (Goldhaber, Quince, & Theobald, 2016) that searched for explanations for the different estimates of the teacher quality gaps—namely, the unequal distribution of effective teachers across students—reported in studies conducted from 2013 to 2016. Regardless of the specifications of the value-added model employed in these studies, the subject, and the grade-level examined, districts that served more disadvantaged students tended to have lower average teacher quality. Collectively, these studies highlight not only the unequal distribution of teacher quality among disadvantaged and nondisadvantaged students but also pinpoint the sorting of students across districts as a key contributor to these inequalities.

Research Questions

Our research extends the knowledge base described above by simultaneously using multiple indicators of teacher characteristics and experiences to predict student mathematics outcomes. We do so because many extant studies explore only a small number of variables and measures; this leads to the possibility of omitted variable bias and the misidentification of important characteristics. Furthermore, even the studies reviewed above that contain more than one type of teacher indicator typically do not report the relationships *between* teacher characteristics.

We focus exclusively on teacher characteristics even though we would typically expect that more proximal measures to student learning (i.e., instructional quality) would best predict student learning in mathematics. Focusing on characteristics, however, asks a different question: What kinds of knowledge and experiences might be associated with successful teaching? Answering this question provides information to agencies seeking to hire teachers in ways that maximize student outcomes, and to place teachers within districts in equitable ways. Specifically, we ask the following research questions:

Research Question 1: How do measures describing teacher preparation and experience, teacher knowledge, and teacher mind-sets and habits correlate with one another and relate to student outcomes in mathematics when concurrently examined?

Research Question 2: How are key teacher characteristics distributed across student populations within districts?

Because our dataset includes two different kinds of mathematics tests—standardized state tests as well as a content-aligned project-administered test—we can examine the consistency of our findings, important given prior research reporting divergent results across different assessments (Lockwood et al., 2007; Papay, 2011).

Data and Methods

Sample

Our data come from the National Center for Teacher Effectiveness main study, which spanned three academic years, from 2010-2011 to 2012-2013. The study, which developed and validated several measures of mathematics teacher effectiveness, collected data from fourth- and fifth-grade teachers and their students in four large urban East Coast public school districts. The project recruited 583 teachers across the four districts, of which 328 matriculated into the study. After excluding majority special education classrooms and those with excessive missing student data at student baseline assessment, we arrived at an analytic sample of 306 teachers and 10,233 students over the three study years.

Our teacher sample was generally experienced, with an average self-reported 10.22 years (*SD* = 7.23 years) in teaching at entry into our study. Most of the sample was traditionally certified (86%), and roughly half had a bachelor's degree in education (53%). A small proportion had a mathematics-specific certification (15%), and a relatively large fraction reported possessing a master's degree (76%). Student demographics reflected those in most urban settings, with 64% of students eligible for FRPL, 10% qualified for special education (SPED), and 20% designated as English language learners (ELLs) at the time of the study. A notable percentage of the participating students were either Black (40%) or Hispanic (23%).

Data Sources and Reduction

Data collection relied upon several instruments, including a background and experience questionnaire administered once to teachers in their first year of study participation; a fall questionnaire, administered each school year and comprising questions measuring teachers' mathematical knowledge as well as questions related to teachers' mind-sets and habits; and a spring questionnaire, administered each school year and containing items assessing teachers' knowledge of students. As noted above, we gauged student performance with both a project-developed mathematics assessment (see Hickman, Fu, & Hill, 2012) and with state standardized test scores. Districts provided the latter scores in addition to student demographic information.

Below, we describe the teacher measures used in this study. We selected these measures, which we organize into the three categories described in our literature review, based on prior theoretical and empirical evidence supporting their importance for student learning.

Teacher Preparation and Experience Measures

We used teacher responses to eight survey items to develop the following measures related to preparation and experience:

- A dichotomous variable indicating novice teachers (i.e., those with no more than 2 years of experience);
- Ordinal variables, with responses ranging from 1 ("no classes") to 4 ("six or more classes"), indicating teachers' reported number of undergraduate or graduate-level classes covering college-level mathematics topics (mathematics courses), mathematics content for teachers (mathematics content courses), and methods for teaching mathematics (mathematics methods courses);
- A dichotomous variable indicating a traditional pathway into the profession (traditionally certified) as opposed to participation in an alternative certification program (e.g., Teach for America) or no participation in any formal training;
- A dichotomous variable indicating possession of a bachelor's degree in education;
- A dichotomous variable indicating possession of a certificate in the teaching of elementary mathematics; and
- A dichotomous variable indicating possession of any master's degree.

Teacher Knowledge Measures

The fall and spring teacher questionnaires supplied two measures of teacher knowledge. The first was MKT/STEL, built from items assessing teachers' Mathematical Knowledge for Teaching (MKT; Hill et al., 2005) and items from a state Test of Education Licensure (STEL). We originally hoped these two types of items would form two measures, one representing teachers' basic mathematical competence (STEL) and one representing teachers' specialized mathematical knowledge (MKT), but a factor analysis suggested these dimensions could not be distinguished (Charalambous, Hill, McGinn & Chin, 2017, manuscript in preparation). The second knowledge measure tapped teachers' *accuracy* in predicting student performance on items from the project test. We presented teachers with items from the project-administered mathematics assessment, and then asked what percent of their students would answer the item correctly. Using these data, we calculated the absolute difference between the teacher estimate and actual percentage correct within the teacher's classroom. We then estimated a multilevel model, with each difference as the dependent variable, that crossed fixed item effects with random teacher effects while including weights for the number of students in each classroom; we adjusted the random effects from this model—the accuracy scores—for the classroom composition of students on evidence that teachers of low-performing students may receive higher difference scores because teachers are generally overoptimistic regarding student outcomes rather than a true difference in accuracy (see Hill & Chin, under review). The MKT/STEL measure possessed a reliability of .92; the adjusted intraclass correlations of the teacher accuracy scores ranged from .71 to .79.[1] For more information on the construction and validity of these knowledge measures, please see (Hill & Chin, under review).

Teacher Mind-Sets and Habits Measures

Responses to items on the fall questionnaire allowed us to estimate scores reflecting teacher efficacy, locus of control, and effort invested in noninstructional activities. We adapted efficacy items from Tschannen-Moran and Hoy (2001); these items describe teachers' assessment of their ability to carry out common classroom activities (e.g., crafting good questions). We selected locus of control items from Hoy and Woolfolk (1993) and Dweck, Chiu, and Hong (1995); these items capture teacher beliefs about, for example, whether or not students can change their intelligence or learn new things. Project staff created effort items, which captured the amount of time spent on noninstructional activities like grading homework or securing resources for students. Efficacy and locus of control were subject-independent metrics, whereas the effort measure was specifically tied to mathematics. Table 1 shows the items related to each measure and the internal consistencies of composites. Because the study asked teachers about efficacy and effort in multiple years, we leveraged this additional information using the following equation:

Table 1. Descriptions of Teacher Mind-Sets and Habits Measures.

	Items	Alpha 2010-2011	Alpha 2011-2012
Efficacy[a]	Belief in ability to, for example, craft good questions for students, provide alternative explanations or examples to confused students, use a variety of assessment strategies to help students learn, control disruptive behavior	.66	.86
Locus of control	Belief in, for example, whether or not students can change their intelligence; whether or not students learn new things		.93
Effort	Time spent per week, for example, on grading math assignments, gathering and organizing math lesson material, reviewing the content of specific math lessons, helping students learn any subject after school hours	.79	.73

[a]For the efficacy measure, items and scales changed between 2010-2011 and 2011-2012. Scores are thus standardized within school year.

$$TQ_{yt} = \beta_0 + \alpha_y + \mu_t + \varepsilon_{yt} \tag{1}$$

The outcome, TQ_{yt}, represents the average of teacher t's responses, within year y, across the items of each respective construct. The model controls for differences in average response level across years using year fixed effects, α_y. The random effect for teacher t, μ_t, comprises each teacher's score on effort or efficacy.[2]

Student Mathematics Tests

We employed two measures of student learning in mathematics. First, districts supplied student scores on state mathematics tests for the years of the study and for up to 2 years prior; many schools and teachers experienced these as high-stakes assessments due to No Child Left Behind regulations. These state tests ranged in content, from two that primarily focused on basic skills and problem-solving to one—used in two study districts—that required more complex thinking and communication about mathematics (Lynch, Chin & Blazer, 2017). Second, sampled students completed a project-developed mathematics test in the spring semester of each school year. Project staff designed this assessment in partnership with the Educational Testing Service to include cognitively challenging and mathematically complex problems. The staff hoped that the assessment would prove more reflective of current standards for student learning (i.e., Common Core Standards for Mathematics) and would more strongly align to the study's mathematics-specific knowledge measures. Student-level reliabilities for this test ranged from .82 to .89. Depending on the academic year, student-level correlations between the state and project tests ranged from .69 to .75 and correlations of teacher value-added scores based on these tests ranged from .29 to .53.

Scoring and Imputation of Missing Data

For ease of interpretation, we standardized students' test scores. Specifically, we standardized students' project-developed mathematics test scores across districts to have a mean of zero and an *SD* of 1; we similarly standardized state mathematics test scores, but did so within district because the assessments differed across each context. We de-meaned the teacher scores described above by district to account for the sometimes sizable differences in teacher characteristics across those districts. Doing so also mirrored the scoring of students' state standardized test performance, which was standardized within district to account for differences in tests. Roughly 63% of the teachers used in our analyses had complete data; 95% of teachers were missing four variables at most. For cases of missing data, we imputed scores using the district mean and included a dichotomous indicator designating whether a teacher was missing data from a specific source (i.e., from the background, fall, or spring questionnaire).

Analysis Strategies

We used three primary strategies in our analyses. We began by correlating teacher scores from all the measures listed above. Doing so describes the relationship of the teacher characteristics to one another and also serves as a check for multicollinearity. Next, we predicted student performance on both the state and project mathematics tests using the following multilevel model, which nests students within teacher–year combinations, which are subsequently nested within teacher:

$$y_{spcgyt} = \beta_0 + \alpha X_{sy-1} + \delta D_{sy} + \phi P_{pcgyt} + \kappa C_{cgyt} + \eta + \omega\theta_t + \mu_t + \nu_{yt} + \varepsilon_{spcgyt} \qquad (2)$$

The outcome, y_{spcgyt}, represents either the state or project test performance of student s, in classroom p, in cohort (i.e., school, year, and grade) c, taking the test for grade g, in year y, taught by teacher t. Equation 2 contains the following controls:

- X_{sy-1}, a vector of controls for student prior test performance;
- D_{sy}, a vector of controls for student demographic information (i.e., race or ethnicity, gender, FRPL-eligibility; SPED status; and ELL status);
- P_{pcgyt}, classroom-level averages of X_{sy-1} and D_{sy} to capture the effects of a student's peers;
- C_{cgyt}, cohort-level averages of X_{sy-1} and D_{sy} to capture the effect of a student's cohort;
- η, district and grade-by-year fixed effects;
- θ_t, a vector of teacher-level scores for teacher characteristic measures[3];
- μ_t, a random effect on test performance for being taught by teacher t; and
- ν_{yt}, a random effect on test performance for being taught by teacher t in year y.

The model represented by Equation 2 contains controls used by many states and districts when estimating teacher value-added scores. One advantage of this model is that it uses multiple classroom-years to construct teacher value-added scores; prior research shows that doing so results in less biased and more reliable score estimates (Goldhaber & Hansen, 2013; Koedel & Betts, 2011). We included classroom- and cohort-average student demographics as well as district fixed effects to account for the sorting of teachers to student populations. As our data show evidence of such sorting (discussed below), we considered these controls appropriate; to further control for sorting, we also conducted specification checks including models with school fixed effects. Results (available upon request) largely replicated those presented here.

We recovered our primary parameters of interest, coefficients representing the relationship between specific teacher characteristics to student test performance, from ω. To make interpretation of these coefficients easier, we standardized teacher scores on the mathematics courses, knowledge, and mind-sets and habits measures across the teacher sample prior to model estimation. We estimated Equation 2 four times for each outcome, testing the variables first by category and then in an omnibus model. As we had a relatively large number of variables given our sample size, we set a slightly higher threshold for statistical significance, referring to estimates with p values between .10 and .05 as marginally significant. For each model estimation, we conducted a Wald test to examine the joint significance of each category's variables in predicting outcomes and also reported the amount of variance in teacher effects explained by each category (an "adjusted pseudo R-squared"; see Bacher-Hicks, Chin, Hill & Straiger, 2018).

Finally, as noted above, prior research has demonstrated imbalances in teacher qualifications across student populations. To investigate this issue, we followed Goldhaber et al. (2015) and calculated teacher quality gaps between advantaged and disadvantaged students (i.e., students who were Black/Hispanic, FRPL-eligible, ELL, and/or in the bottom quartile of prior-year state mathematics test performance), and estimated the significance of these gaps using two-sample *t* tests. Specifically, we compared the percentage of disadvantaged students in our sample taught by teachers that performed poorly, as compared with other teachers in the same district, on our key measures of quality to the percentage for advantaged students.

Results

Examining Associations Among Teacher Characteristics and With Student Outcomes

We start by discussing the correlations among teacher characteristics. Table 2 shows few notable correlations that arose between our independent variables.

Teachers' reports of completing mathematics content for teachers and mathematics methods courses correlated strongly (r = .67); correlations between these variables and college-level mathematics courses were moderate (r = .44, .48). Consistent with the conventional training and certification processes in most states, traditionally certified teachers more often possessed a bachelor's degree in education (r = .59). This, along with evidence of multicollinearity in our regressions, led us to combine the mathematics content/methods courses variables and traditional certification/education bachelor's degree metrics in our analyses below. Overall, novice teachers in our sample took fewer courses (mathematics courses, r = –.25; mathematics content courses, r = –.34; mathematics methods courses, r = –.32), and often did not possess a master's degree (r = –.37). These patterns reflect what we would expect intuitively, as newer teachers have had less time to attain these additional milestones. Novice teachers also reported feeling less efficacious (r = –.22).

Teachers who reported receiving their bachelor's degree in education were less likely to also have completed a master's degree (r = –.31). There was a notable relationship between reported mathematics courses and effort (i.e., time spent grading papers, preparing for class, and tutoring; r = .22), perhaps a sign that some teachers in our sample had more time or inclination to invest in their work. Contrary to expectations, teacher completion of mathematics

Table 2. Correlations Between Teacher Characteristics.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Novice	1												
2. Math courses	−.25	1											
3. Math content courses	−.34	.48	1										
4. Math methods courses	−.32	.44	.67	1									
5. Traditional certification	−.11	−.12	.11	.09	1								
6. Education bachelor's	−.16	−.04	.13	.24	.59	1							
7. Elementary math certification	−.04	.14	.09	.09	−.43	.15	1						
8. Master's	−.37	0	−.01	0	.12	−.31	−.12	1					
9. MKT/STEL	.11	.04	.02	.03	.07	−.05	.01	.04	1				
10. Accuracy	.1	−.05	−.07	−.07	.1	−.02	−.03	.11	.24	1			
11. Efficacy	−.22	.1	.01	.05	−.1	0	.11	−.08	.04	−.08	1		
12. Locus of control	0	0	.02	.03	.1	.05	.1	.01	0	−.02	−.19	1	
13. Effort	.01	.22	.14	.14	−.08	.01	−.01	−.12	−.14	−.09	.07	−.2	1

Note. Light gray cells indicate correlations between .30 and .50. Dark gray cells indicate correlations greater than .50. Tetrachoric, polychoric, and polyserial correlations reported when appropriate. MKT = Mathematical Knowledge for Teaching; STEL = State Test of Education Licensure.

content or methods courses did not strongly associate with our teacher knowledge measure.

Looking further into the measures from the knowledge and mind-sets and beliefs categories, we observed few additional patterns. As predicted, the two measures of teachers' knowledge correlated with one another, though more weakly than expected ($r = .24$). Teacher efficacy also negatively but weakly correlated with locus of control ($r = -.19$). The former observed relationship reflects what might be expected intuitively: Our locus-of-control variable measures the endorsement of a view of fixed intelligence, and teachers appear to feel more efficacious when they also feel they can influence student learning. Some relationships were remarkable in their absence. Against expectations, mathematical content knowledge did not relate to perceived teacher efficacy; similarly, efficacy only weakly related to teacher accuracy in predicting student performance on the project test.

Next, we explore how different measures relate to student outcomes. Table 3 shows regressions predicting student performance on both the state and project tests. Among self-reported teacher preparation and experiences, no measures were significant in both the state and project mathematics tests' final models. This includes teachers' completion of mathematics content and/or methods courses, which positively predicted student performance on both tests, though only the relationship to the project test appeared significant. Similarly, students taught by novice teachers performed more poorly on both the state and project tests, corroborating findings of prior research, yet only the former outcome proved significant. Beyond these variables, no others within this category significantly related to student outcomes, including the possession of a master's degree, the possession of elementary mathematics certification, and the combined measure for being traditionally certified and possessing a bachelor's degree in education. Noticeably, despite the low overall variance explained by these variables, the Wald test indicated that the measures together were jointly significant for predicting performance on the state test, and were marginally significant for the project test.

In the knowledge category, teachers' MKT/STEL and accuracy scores predicted student outcomes on both assessments, with the point estimate and significance for accuracy slightly higher. Despite again observing low explained variance, the significance of the Wald test examining the joint significance of both knowledge measures for both outcomes further supports theories positing the importance of teacher knowledge for student learning. Although these associations are not causal, these results suggest that teaching-related mathematical knowledge and predictive accuracy, though correlated with one another, may be individually important, and thus contribute separately to student growth.

Table 3. Predicting Student Mathematics Test Performance Using Teacher Characteristics.

	State test				Project test			
	Prep. and experiences	Knowledge	Mind-sets and habits	All	Prep. and experiences	Knowledge	Mind-sets and habits	All
Novice	−0.107*			−0.121*	−0.022			−0.035
	(0.046)			(0.045)	(0.046)			(0.046)
Math courses	0.004			−0.004	0.001			−0.002
	(0.014)			(0.014)	(0.014)			(0.014)
Math content/methods courses	0.012			0.012	0.020**			0.021**
	(0.008)			(0.008)	(0.008)			(0.007)
Trad. cert./Ed. bachelor's	0.028			0.027	−0.003			−0.004
	(0.020)			(0.020)	(0.019)			(0.019)
El. math cert.	−0.004			0.002	−0.045			−0.041
	(0.035)			(0.034)	(0.033)			(0.033)
Master's	0.010			0.008	−0.009			−0.016
	(0.031)			(0.030)	(0.029)			(0.029)
MKT/STEL		0.017		0.023†		0.023†		0.023†
		(0.013)		(0.012)		(0.012)		(0.012)
Accuracy		0.023†		0.027*		0.030*		0.034**
		(0.012)		(0.012)		(0.012)		(0.012)
Efficacy			−0.001	−0.001			0.002	0.004
			(0.012)	(0.012)			(0.012)	(0.012)
Locus of control			0.004	0.001			−0.004	−0.006
			(0.012)	(0.012)			(0.012)	(0.011)
Effort			0.032**	0.034**			0.008	0.005
			(0.012)	(0.013)			(0.012)	(0.012)
Variance explained	−0.002	0.039	0.048	0.093	0.047	0.062	0.004	0.121
Wald test p value	.016	.034	.078	.001	.070	.002	.859	.004

Note. The number of students and teachers in each model is 10,233 and 306, respectively. MKT = Mathematical Knowledge for Teaching; STEL = State Test of Education Licensure.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

In the mind-sets and habits category, neither the efficacy nor the locus-of-control measures predicted performance on the state or project assessment in the final analysis. By contrast, teachers' self-reported effort—the number of hours spent grading mathematics homework, preparing mathematics lessons, and tutoring students outside of regular school hours—predicted performance on the state but not on the project test. To check the intuition that tutoring may have driven this result, perhaps as teachers helped students prepare for state assessments, we removed the tutoring item from the scale and found the same result (b = .035, p < .01). Despite this striking finding, this category explained again only a small (5%) amount of variance in teacher effects on student state test performance and the Wald test was just marginally significant.

To determine the extent to which the significant relationships between our measures of teacher characteristics and student test scores is meaningful, we make two comparisons. First, we compare the coefficient sizes in Table 3 with the size of the teacher-level *SD* in student scores on the state tests (0.16) and project-developed tests (0.14). Second, we compare the relationship between teacher characteristics and test scores with those of key student characteristics and test scores, such as FRPL-eligibility ($\beta_{State} = -.05; \beta_{Project-developed} = -.04$). Through these comparisons, we see that the coefficient on state test performance for a novice teacher ($\beta = -.12$) is sizable; students taught by a more veteran teacher in our sample had test score gains of nearly three quarters of those taught by a teacher 1-*SD* above average for raising student state test scores; this coefficient also completely closes the gain-gap for students who are FRPL-eligible. On the contrary, being taught by a teacher 1-*SD* above average on other significant teacher predictors such as MKT/STEL, accuracy, and self-reported effort does not yield commensurate gains. In each case, students taught by an above-average teacher gain less than one fifth of that observed for students taught by teacher 1-*SD* above average for raising student test scores. However, teacher performance on these measures still accounts for a significant proportion of the gain-gap associated with FRPL-eligibility.

Finally, as noted above, we assessed the consistency of relationships across the two different student tests (i.e., criterion stability). Findings here perhaps shed light on the mostly mixed results from prior studies of the education production function. Conflicting results, for instance, could have been caused by studies occurring in states with different teacher education and certification pathways (some more effective than others), or, as here, by differences in the sensitivity of the tests used to measure teacher characteristics. Our results suggest conflicting findings for three variables—teacher experience, enrollment in mathematics content and/or methods courses, and effort. Notably, however, nonsignificant findings were similarly consistent. Thus, we found more consistency across tests within the same sample than suggested by the aggregate findings from prior literature.

Investigating the Distribution of Teacher Characteristics Across Student Populations

To examine the distribution of key teacher characteristics across student populations within districts (our second research question), Table 4 displays the exposure of disadvantaged students to teacher characteristics associated with poorer performance, as judged by our models above. Black and Hispanic students, FRPL-eligible students, ELL-students, and low-performing students were all exposed to novice teachers and teachers performing in the bottom quartile of their districts on the MKT/STEL and accuracy measures more frequently than their more advantaged counterparts. These findings were unsurprising and matched intuition and prior research (Hill & Lubienski, 2007; Loeb & Reininger, 2004). Conversely, we found the opposite pattern for teacher effort; disadvantaged students were less frequently exposed to teachers in the bottom quartile of their district on this measure. These findings suggest that teachers may adjust their behaviors in response to the needs of students they teach. Interestingly, exposure to teachers with less mathematics content and/or methods coursework was generally evenly distributed across groups of students, with the exception of the exposure gap between FRPL- versus non–FRPL-eligible students. Mathematics content and/or methods courses may be prescribed by local or state guidelines, or incentivized by districts themselves, resulting in a relatively even distribution across student populations. Overall, however, the results depicted in Table 4 suggest that the teacher-level characteristics we identified in earlier analyses as important for student learning varied inequitably across among students in our sample. We return to these results and their implications for policy in our "Discussion" section.

Finally, we conducted a number of checks of our results, looking for interaction effects between key variables (e.g., whether effort yielded differential benefits by teacher knowledge) and between key variables and student populations (e.g., whether there were consistent patterns in the association between resources within student subgroups). We found no significant interaction effects.

Discussion

We initiated our work by noting that although policy makers may wish for clear guidance regarding characteristics of effective teachers, scholarship on

Table 4. Exposure Rates to Low-Performing Teachers on Key Teacher Characteristics.

Panel A	Race/ethnicity			FRPL-eligibility		
	Black/ Hispanic	Non–Black/ Hispanic	Difference	FRPL	Non-FRPL	Difference
Novice Teacher	0.07	0.03	0.04***	0.07	0.03	0.04***
Low Math Content/Methods Courses Teacher	0.45	0.42	0.03	0.41	0. 45	−0.04***
Low MKT/STEL Teacher	0.23	0.16	0.07***	0.23	0.17	0.06***
Low Accuracy Teacher	0.22	0.20	0.02*	0.22	0.19	0.03**
Low Effort Teacher	0.21	0.26	−0.05***	0.21	0.26	−0.05***

Panel B	ELL status			Quartile of prior state test performance		
	ELL	Non-ELL	Difference	Lowest	Non-lowest	Difference
Novice Teacher	0.09	0.05	0.04***	0.07	0.05	0.02***
Low Math Content/Methods Courses Teacher	0.43	0.43	0.00	0.44	0.42	0.02
Low MKT/STEL Teacher	0.27	0.19	0.08***	0.25	0.19	0.06***
Low Accuracy Teacher	0.25	0.20	0.05***	0.24	0.20	0.04***
Low Effort Teacher	0.19	0.24	−0.05***	0.19	0.24	−0.05***

Note. FRPL = free or reduced-price lunch; MKT = Mathematical Knowledge for Teaching; STEL = State Test of Education Licensure; ELL = English language learner.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$ (two-sample *t* test).

this topic has returned mixed results and, in many cases, studies that examine only a handful of characteristics in isolation (for exceptions, see, for example, Boonen et al., 2014; Campbell et al., 2014; Grubb, 2008; Palardy & Rumberger, 2008). By bringing together characteristics from three main categories—teacher preparation and experience, knowledge, and mind-sets and habits—we assessed their joint associations with student learning. We also explored the criterion consistency of our findings, as we considered student learning as measured on a state standardized assessment and on a project mathematics assessment. In doing so, our study not only complements but also extends existing approaches examining the contribution of teacher-level characteristics to student learning in two significant ways. First, it explores a noticeably more comprehensive list of teacher attributes compared with those considered in prior studies; second, we examine the distribution of key teacher characteristics across student populations within districts, building on the work of Choi (2010), Goldhaber et al. (2015), and Schultz (2014). Finally, our work also allowed us to examine correlations between teacher characteristics.

We found most correlations in our data to be mild, at .20 or low. Variables that represent teacher preparation and experiences proved to be one exception, with factors such as coursework, an education major, certification, experience, and a master's degree correlating at .30 or higher. This may shed light on the mixed evidence for many of these variables in the extant economics of education literature, where omitted variable bias may lead to disparate results. These correlational analyses also revealed that expected relationships between such variables do not always materialize. This was particularly interesting in the case of mathematics coursework, teacher knowledge, and efficacy, where we expected to see strong relationships.

By and large, results from both of the student-level outcomes under consideration pointed to the same characteristics as potentially important. Consistent with some prior literature (Monk, 1994; Rice, 2003) and the attention paid to such courses in educational systems worldwide (Tatto et al., 2008), we found that the completion of mathematics content and/or methods courses positively related with student learning on both outcomes, with an observed significant relationship for the project test. This is remarkable in an era in which many teacher preparation programs—particularly alternative entry pathways—do not feature content-specific teaching coursework. It suggests that such coursework may be an important support for elementary teachers (Sleeter, 2014); however, the possibility that selection effects (e.g., teachers more comfortable with mathematics may enroll in more such courses) influenced our findings cannot be ruled out. Because this work took place in only four districts, we also cannot be sure whether the associations we observe are specific to the teacher education programs that serve these districts, or whether this association between coursework and student outcomes holds generally.

Similarly, in line with recent research findings on teacher knowledge, the two teacher mathematics knowledge measures we employed—teacher accuracy in predicting student mathematics test performance and MKT/STEL—positively related to student outcomes. This supports the importance of teacher knowledge of content and its teaching and of what students know and do not know—both components of Shulman's conceptualization of teacher knowledge and Ball, Thames, and Phelps's (2008) notion of MKT. The models in this article improve upon those offered in most prior research, in that they are well controlled for related teacher characteristics and knowledge, suggesting that these associations were not driven by omitted correlates such as efficacy and mathematics coursework. Interestingly, the variables representing mathematics methods/content courses and teacher knowledge did not relate to one another, suggesting that each had an independent pathway through which they related to student outcomes.

As in other reports (e.g., Kane et al., 2008), lack of teaching experience related negatively to student outcomes. Here, however, the measure was only significant for the state test. One intriguing possibility is that the association between novice teachers and student outcomes may result as much from familiarity with the state standardized test as it does from lagging effectiveness in the classroom. Novice teachers may not optimally adjust their curriculum and pacing to align with the state assessment; they also may be unfamiliar with question formats and topics. A similar study found no relationship between experience and state test scores, with a positive effect observed between novice teachers and an alternative test in mathematics (Kane & Staiger, 2012). Thus, this is an issue for further analyses, potentially via a review of available evidence on this topic.

Our findings also suggest attention to teacher effort, measured as investment in noninstructional work hours, which here had a positive association with student learning as measured by state test results. Interestingly, this positive association did not appear to be driven by tutoring alone. If students benefit when their teacher spends more time grading papers and preparing for class, then arranging for a greater ratio of noninstructional to instructional work hours in U.S. schools—and cultivating knowledge about the productive use of that time—becomes imperative. However, because we cannot make a causal attribution in this study, this issue warrants further investigation.

Several teacher characteristics thought to predict increased student performance in mathematics did not do so in this sample. Teacher self-efficacy was one such characteristic. Although this is a frequently studied teacher belief, and even though we used a widely disseminated metric, it neither correlated strongly with hypothetically related constructs (e.g., teacher knowledge, locus of control) nor appeared related to student outcomes. We also note, parenthetically, that the version of the self-efficacy instrument used here seemed to us more a self-report of teaching expertise than efficacy as envisioned by theorists such as Bandura (1997), for whom the construct also included aspects of grit and task persistence. Locus of control also saw relationships close to zero.

We found imbalances of key teacher characteristics across populations of students. Specifically, students from disadvantaged groups more frequently had novice teachers and those with lower knowledge scores. These findings, which hearken to sociologist Robert Merton's (1968) notion of accumulated advantage—"the rich get richer and the poor get poorer"—align with other related findings and scholarly discussions both in the United States (Darling-Hammond, 2010) and worldwide (Schleicher & Organization for Economic Co-Operation and Development, 2014). This research implies that the students in most need of high-quality teachers are not afforded them. Although this imbalance—a significant problem—cannot explain the entire achievement gap between privileged and less privileged students, it undoubtedly explains some. Advocates for better hiring and placement practices (e.g., Rutledge, Harris, Thompson, & Ingle, 2008) are correct in noting that this is a solvable problem (Liu & Johnson, 2006), and in fact, the Every Student Succeeds Act (ESSA) requires that states define ineffective teachers and ensure that poor and minority students are not taught disproportionately by such teachers. In mathematics, metrics such as teacher certification test scores in mathematics (related, potentially, to MKT), teachers' content/methods coursework, and years of experience could prove relatively inexpensive ways for states to collect information about inequities in distribution of the teaching workforce, and to incent districts to act upon such inequities by publicizing that information while also supporting districts with poor-quality teachers make improvements in their hiring process.

In larger perspective, these findings suggest that despite some consistent patterns, there does not seem to be one teacher characteristic that exhibited a strong relationship to teacher effectiveness in mathematics. Even variables found to significantly predict student mathematics outcomes had small coefficients. In addition, our variables explained a modest, at best, percentage of the variance in student learning, even in conjunction with one another. This result mirrors outcomes from similar studies, including those that examine teacher characteristics and those that measure instructional quality via observation rubrics (Kane & Staiger, 2012; Stronge, Ward, & Grant, 2011). One perspective on these findings is that there remain unexamined teacher-level variables that help explain student outcomes; another is that factors beyond the scope of this study, including classroom climate and enacted instructional practice, explain student outcomes; a third is that given the complexity of the education production function and noise present in student test score data, no combination of measured factors will result in a model with strong predictive capability.

Because this study is not experimental, we cannot rule out the possibility that the above findings (and nonfindings) are driven by selection effects. Individuals with more aptitude for teaching may invest in more content and methods coursework, and teachers with more mathematical knowledge may be hired into more affluent schools serving more academically advanced students. However, our models include many controls at the teacher, student, and school levels, and recent work in the economics of education has suggested that including prior student achievement in such models serves as an adequate control for teacher sorting (Chetty et al., 2014). This suggests that while not causal, our findings can help formulate strong hypotheses for future research.

Our findings can certainly be useful, even descriptively, to LEAs interested in hiring, retaining, and remunerating the most qualified candidates. One suggestion would be to screen for easily observable variables, including teachers' mathematical knowledge and background in mathematics-specific coursework, and teaching experience. LEAs may also search for proxies for teacher effort, such as existing metrics of conscientiousness (e.g., McIlveen & Perera, 2016). By contrast, LEAs should not hire based on certification, certification specialization in mathematics, or advanced degrees. The master's degree finding also suggests that LEAs may wish to rethink automatic salary increases that often accompany the acquisition of advanced degrees (see also Roza & Miller, 2009).

Similarly, the two aspects of teacher knowledge found to contribute to student outcomes—mathematical knowledge and accuracy in predicting student outcomes—appear amenable to improvement through professional development programs, particularly those focused on mathematics content and formative assessment (e.g., see Lang, Schoen, LaVenia, & Oberlin, 2014). Although more research is needed to further validate these relationships, the results of our study, along with those of qualitative studies documenting positive associations between teachers' knowledge of student thinking and instructional quality (e.g., Bray, 2011; Even & Tirosh, 2008), suggest that LEAs ought to support teachers' development and deepen their respective knowledge in these domains.

Finally, this study provides guidance for LEAs interested in ensuring that teacher expertise is distributed equitably over student populations. LEAs may make use of data collected at the point of hire (e.g., course transcripts, content-specific certification test scores) or in administrative data (e.g., experience) to understand the distribution of teacher characteristics across student populations. LEAs may also wish to collect additional data on teachers' accuracy in assessing student understanding, and around teacher effort. This advice applies to LEAs engaged in meeting ESSA reporting requirements regarding teacher quality and student populations.

The lack of a single "silver-bullet" teacher characteristic predicting student outcomes also contains lessons for research, namely that future research studies of this type should contain as many measures as is practically feasible. Without extensive coverage of key teacher traits, models may suffer from omitted variable bias. The results of past research, which show conflicting evidence regarding key variables such as mathematics methods and content courses and teacher efficacy, also indicate that replication research of this sort is warranted, ideally with larger and more representative datasets. Scholars may also wish to design studies that capture variability in key variables—for example, to discern the relative effectiveness of specific forms of mathematics-related teacher preparation coursework (see, for example, Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009) and specific forms of teacher knowledge and teaching experience. With such research, we could sharpen our lessons and suggestions for practitioners and policy makers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported in part by Grant R305C090023 from the Institute of Education Sciences.

Notes

1. Intraclass correlations were adjusted for the modal number of *accuracy* items teachers responded to. This adjustment provided an estimate of reliability that was more reflective of measure scores, which incorporated teacher responses to several items as opposed to a single item.
2. Because we asked teachers to answer questions regarding locus of control only in 2011-2012, we did not use the model from Equation 1 to estimate scores for teachers on this measure.
3. Teacher accuracy scores were included in the model at the teacher-grade level, and the indicator for being a novice teacher was considered at the teacher–year level.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*, 95-135.

Anderson, R. N., Greene, M. L., & Loewen, P. S. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research*, *34*, 148-165.

Bacher-Hicks, A., Chin, M. J., Hill, H. C., & Staiger, D. O. (2018). *Explaining teacher effects on achievement using commonly found teacher-level predictors*. Manuscript submitted for publication.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*, 389-407.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133-180.

Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature*. Washington, DC: Mathematical Association of America and the Council of Teachers of Mathematics.

Bell, C. A., Wilson, S. M., Higgins, T., & McCoach, D. B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education*, *41*, 479-512.

Berman, P., & McLaughlin, M. W. (1977). *Federal programs supporting educational change, Volume VII: Factors affecting implementation and continuation*. Santa Monica, CA: The RAND Corporation.

Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: Which aspects matter most? *School Effectiveness and School Improvement*, *25*, 126-152.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*, 416-440.

Bray, W. S. (2011). A collective case study of the influence of teachers' beliefs and knowledge on error-handling practices during class discussion of mathematics. *Journal for Research in Mathematics Education*, *42*, 2-38.

Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., . . . Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, *45*, 419-459.

Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, *19*, 385-401.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*, 499-531.

Charalambous, C. Y., Philippou, G. N., & Kyriakides, L. (2008). Tracing the development of preservice teachers' efficacy beliefs in teaching mathematics during fieldwork. *Educational Studies in Mathematics, 67*(2), 125-142.

Charalambous, C. Y., Hill, H. C., McGinn, D., & Chin, M. J. (2017). *Teacher knowledge and student learning: Bringing together two different conceptualizations of teacher knowledge*. Manuscript in preparation

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, *104*, 2633-2679.

Choi, D. (2010). The impact of competing definitions of quality on the geographical distributions of teachers. *Educational Policy*, *24*, 359-397.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*, 673-682.

Cochran-Smith, M., Cannady, M., McEachern, K., Mitchell, K., Piazza, P., Power, C., & Ryan, A. (2012). Teachers' education and outcomes: Mapping the research terrain. *Teachers College Record*, *114*(10), 1-49.

Darling-Hammond, L. (2010). *The Flat World and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*, 267-285.

Ernest, P. (1989). The knowledge, beliefs, and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching*, *15*, 13-33.

Even, R., & Tirosh, D. (2008). Teacher knowledge and understanding of students' mathematical learning and thinking. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 202-223). New York, NY: Routledge.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*, 47-65.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, *80*, 589-612.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*, 293-307. doi:10.3102/0013189X15592622

Goldhaber, D., Quince, V., & Theobald, R. (2016). *Reconciling different estimates of teacher quality gaps based on value added*. National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.calder-center.org/sites/default/files/Full%20Text_0.pdf

Grubb, W. N. (2008). Multiple resources, multiple outcomes: Testing the "improved" school finance with NELS88. *American Educational Research Journal*, *45*, 104-144.

Guarino, C., Dieterle, S. G., Bargagliotti, A. E., & Mason, W. M. (2013). What can we learn about effective early mathematics teaching? A framework for estimating causal effects using longitudinal survey data. *Journal of Research on Educational Effectiveness*, *6*, 164-198.

Hickman, J. J., Fu, J., & Hill, H. C. (2012). *Creation and dissemination of upper-elementary mathematics assessment modules*. Princeton, NJ: ETS.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*, 798-812.

Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, *3*, 91-98.

Hill, H. C., & Chin, M. J. (under review). *Connecting teachers' knowledge of students, instruction, and achievement outcomes*.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794-831.

Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy, 21*(5), 747-768.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371-406.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, *93*, 355-372.

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, *27*, 213-256.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*, 51-68.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*, 615-631.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations, student surveys, and achievement gains*. Seattle, WA: The Measures of Effective Teaching Project, The Bill and Melinda Gates Foundation.

Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, *6*, 18-42.

Lang, L. B., Schoen, R. R., LaVenia, M., & Oberlin, M. (2014). *Mathematics Formative Assessment System—Common Core State Standards: A randomized* *field trial in kindergarten and first grade*. Paper presented at the Annual Meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*, 37-62.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *The American Economic Review*, *99*, 1979-2011.

Liu, E., & Johnson, S. M. (2006). New teachers' experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly*, *42*, 324-360.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, *44*, 47-67.

Loeb, S., & Reininger, M. (2004). *Public policy and teacher labor markets: What we know and why it matters*. East Lansing, MI: The Education Policy Center at Michigan State University.

Lynch, K., Chin, M., & Blazar, D. (2017). Relationships between observations of elementary mathematics instruction and student achievement: Exploring variability across districts. *American Journal of Education*, *123*(4), 615-646.

Max, J., & Glazerman, S. (2014). *Do disadvantaged students get less effective teaching? Key findings from recent Institute of Education Sciences studies*. National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/pubs/20144010/pdf/20144010.pdf

McIlveen, P., & Perera, H. N. (2016). Career optimism mediates the effect of personality on teachers' career engagement. *Journal of Career Assessment*, *24*, 623-636.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*, 56-63.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, *99*, 486-496.

Molden, D. C., & Dweck, C. (2006). Meaning in psychology: A lay theories approach to self-regulation, social perception, and social development. *American Psychologist*, *61*, 192-203.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*, 125-145.

Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, *40*, 139-157.

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, *30*, 111-140.

Papay, J. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcomes measures. *American Educational Research Journal*, *48*, 163-193.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105-119.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*, 43-74.

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *The Journal of Educational Research*, *74*, 185-190.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*, 1-28.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *The Teachers College Record*, *104*, 1525-1567.

Roza, M., & Miller, R. (2009). *Separation of degrees: State-by-state analysis of teacher compensation for Master's degrees*. Seattle, WA: Center on Reinventing Public Education.

Rutledge, S. A., Harris, D. N., Thompson, C. T., & Ingle, W. K. (2008). Certify, blink, hire: An examination of the process and tools of teacher screening and selection. *Leadership and Policy in Schools*, *7*, 237-263.

Schleicher, A. & Organisation for Economic Co-Operation and Development. (2014). *Equity, excellence and inclusiveness in education: Policy lessons from around the world*. Paris, France: Organisation for Economic Co-Operation and Development.

Schultz, L. M. (2014). Inequitable dispersion: Mapping the distribution of highly qualified teachers in St. Louis metropolitan elementary schools. *Education Policy Analysis Archives*, *22*(90), 1-20.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Sleeter, C. (2014). Toward teacher education research that informs policy. *Educational Researcher*, *43*, 146-153.

Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education*, *12*, 401-411.

Stipek, D. (2012). Context matters: Effects of student characteristics and perceived administrative and parental support on teacher self-efficacy. *The Elementary School Journal*, *112*, 590-606.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, *62*, 339-355.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual*

Framework. East Lansing: Teacher Education and Development International Study Center, College of Education, Michigan State University.
Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783-805.
Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*, 202-248.
Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89-122.

Author Biographies

Heather C. Hill is the Jerome T. Murphy Professor in Education at the Harvard Graduate School of Education. Her primary work focuses on teacher and teaching quality and the effects of policies aimed at improving both. She is also known for developing instruments for measuring teachers' mathematical knowledge for teaching (MKT) and the mathematical quality of instruction (MQI) within classrooms.

Charalambos Y. Charalambous is an Assistant Professor in Educational Research and Evaluation at the Department of Education of the University of Cyprus. His main research interests center on issues of teaching/teacher effectiveness, contributors to instructional quality, and the effects of instruction on student learning.

Mark J. Chin is a PhD Candidate in Education Policy and Program Evaluation at Harvard University. His research interests center on the experiences of students of color and first-and second-generation students in US K-12 educational contexts. His work focuses on explorations of how race, racism, and assimilative pressures cause inequality in these students' outcomes.