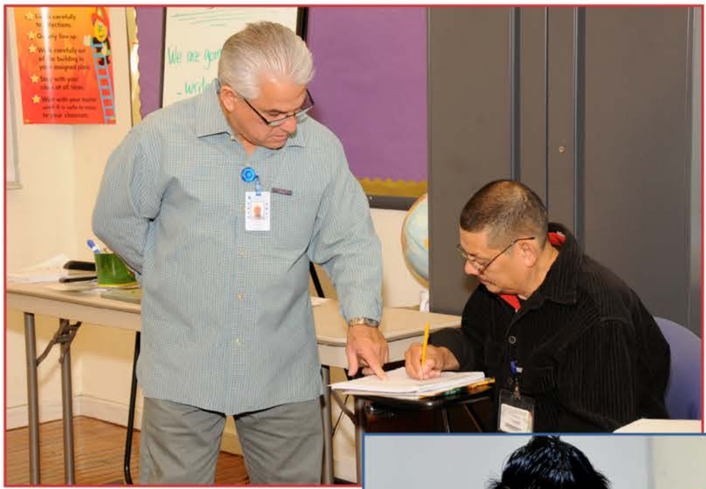
AMERICAN INSTITUTES FOR RESEARCH®

TEACHER EFFECTIVENESS IN ADULT EDUCATION

The Importance of Teacher Background Qualifications for Student Learning



Michelle Yin
Larry Condelli
Burhan Ogut
Stephanie Cronen

These materials were developed by American Institutes for Research (AIR) in partnership with the Office of Career, Technical, and Adult Education of the U.S. Department of Education as part of the Promoting Teacher Effectiveness in Adult Education Contract ED-CFO-10-A-0066. The views expressed herein do not necessarily represent the positions or policies of the Office of Career, Technical, and Adult Education or the U.S. Department of Education. No official endorsement of any product, commodity, service, or enterprise mentioned in this material is intended or implied.

Contents

Executive Summary

The federal adult education program serves more than 2 million eligible adults who lack basic literacy and English language skills.[1] Although numerous studies in K–12 have shown that measurable teacher characteristics such as certification, advanced degrees, and teacher scores on standardized tests are related to student achievement (Aaronson, Barrow, & Sander, 2007; Kane, Rockoff, & Staiger, 2006), no studies with strong statistical designs have explored teacher effectiveness in adult education.

To provide descriptive information about the characteristics of teachers in adult education and to explore whether teacher quality is associated with student achievement in adult education, the Office of Career, Technical, and Adult Education (OCTAE) contracted with American Institutes for Research (AIR) to produce four research briefs. The first brief provides research on the characteristics of adult education teachers, this second brief examines the relationships between teacher characteristics and student achievement, the third brief investigates the relationships between teacher characteristics and students transitioning into postsecondary education and the fourth brief focuses on communicating common issues with administrative data and provides recommendations from a research and evaluation perspective.

The adult education research literature lacks evidence that this type of analysis has ever been done at the student level, most likely because of the lack of appropriate data. The analyses in this second brief provide a rare look at how the characteristics of adult education teachers relate to the academic achievement of adult students using student-level data obtained from three states. Results of this study allow us to better understand adult education teachers and the adult student population.

Analytic Data and Method

To assess whether adult education teacher characteristics are correlated with student academic achievement, this study focused on the following areas:

- Teacher demographic characteristics, including gender, age, race/ethnicity, and employment status (part-time or full-time teacher)
- Teacher educational attainment
- Teacher professional development (number of hours participated in teacher PD); and
- Teacher experience, including total number of years in education and total number of years in adult education.

We used adult student–level data for the 2008–09, 2009–10, and 2010–11 program years obtained from the adult education data systems of three states. Our sample included approximately 723,000 students in adult basic education, adult secondary education, and English as a second language programs and nearly 13,000 adult education teachers from 3 states. Two

[1] Retrieved from http://wdcrobcolp01.ed.gov/CFAPPS/OVAE/NRS/reports/index.cfm, National Reporting System (NRS) database, February 11, 2013.

states from the southern region of the United States and one from the Midwest were included in this study.

Our sample is not necessarily representative of all adult education teachers and students. However, the existing research examining adult education teachers and student performance is limited; therefore, despite our limited sample, the research in this report provides important information about this population of adult education teachers and adult students.

Because the availability of data on teacher demographics, educational attainment, PD, and experience vary across states, we conducted our analysis by each state separately. In States 1 and 3, where overall student pre- and posttest scores were available, we estimate the relationships between teacher characteristics and student overall performance. In States 1 and 2, which provided student pre- and posttest scores by subject area (language, mathematics, oral English, reading, and literacy), we performed our analyses for each subject area to obtain more detailed estimates on teacher characteristics and student achievement correlations.

To gauge reliability, we employed multiple regression models: ordinary least squares (OLS) model controlling for only teacher characteristics; OLS model controlling for teacher, student, and program characteristics; student fixed effects model; and teacher random effects (RE) model controlling for teacher, student, and program characteristics. Our preferred model was the teacher RE model with which we estimated the relationship between teacher characteristics and student achievement while accounting for all other characteristics that potentially might affect student achievement and for the fact that multiple students are nested within one teacher. Results reported here are from analyses using this model.[2]

Highlights: What We Have Learned So Far

We draw our conclusions based on the results from our preferred teacher RE model as shown in Table ES-1. We did not find consistent relationships between key teacher demographics and student outcomes across three states; full-time teachers show a small positive relationship with student achievement in two states for mathematics and in one state for reading while the opposite relation was found in a third state. The mixed findings require further research in other states to understand better what the important factors are. Teacher educational attainment shows positive relationships with student achievement in two states for oral English and overall and no relationship in the third state. We also found a positive relationship between PD participation and student achievement in one state and no relationship in others. Teacher overall years of experience show a positive relationship with overall student posttest scores in State 1; data were not available in the other two. Teacher adult education experience presents a positive relationship with overall and mathematics results in State 1 and with language and mathematics results in State 2; there was no observed relationship in the third state.

[2] Full tables and discussions of various models are presented in the appended Technical Notes. We also employed student, teacher, and program fixed effects models to account for potential time-invariant student, teacher, and program characteristics while estimating teacher value-added. However, the data provided from the states did not allow us to produce much meaningful inferences from such models.

Tables A4 and A6 in Appendix A illustrate the regression results from all models on key teacher characteristics from States 1 and 3. Table A5 in Appendix A presents the regression results on key teacher characteristics by subject area from State 2.

Question 1: Are teacher demographics related to student achievement?

- Data on teacher demographics are limited in all three states, and we could not identify any common patterns among them. Although we observe positive relationships between being a female teacher and teacher age with student outcomes in state 1, we did not detect similar relationships in other states. Asian adult education teachers had a statistically significant negative relation with student performance in State 3, and students with Native American adult education teachers had higher mathematics and language posttest scores compared to students with White teachers in State 1. However, we did not observe any other statistically significant relationships between teacher race and student achievement in other states. In State 1, students with full-time teachers performed better on mathematics assessments than students with part-time teachers. We did not observe any statistically significant effects in other subjects in State 1. In State 2, students with part-time teachers scored lower in language, mathematics, reading, and literacy assessments compared to students with full-time teachers. Compared to having full-time adult teachers, however, students with part-time adult education teachers had significantly higher posttest scores in State 3.

Question 2: Is teacher professional development related to student achievement?

- PD programs varied significantly across states. Data available from the states did not allow us to gauge the quality of PD; instead, we could only evaluate whether and how teacher PD participation was correlated with student achievement using the number of hours of PD participation. We found a small but statistically significant positive relationship between hours of PD and students' oral English posttest scores in State 2. No relationship was found in other states.

Question 3: Is teacher experience related to student achievement?

- A teacher's years of experience teaching adult education was found to be significantly related to students' overall and mathematics posttest scores in State 1 and to students' language and mathematics posttest scores in State 2. In addition, students in classes with teachers who had more years of experience in any level of education showed higher posttest scores in State 1. Data on overall teacher experience were not available in States 2 and 3, and no relationship was found between adult education teaching experience and student outcomes in State 3.

Table ES-1. Summary of Regression Results on Key Teacher Characteristics Using Teacher RE Model From States 1, 2, and 3

Teacher Characteristic	State 1	State 2	State 3
Female Teacher		Higher language, mathematics, reading, and literacy posttest scores	
Older teacher	Lower oral English posttest score	†	
Native American	Higher language and mathematics posttest scores	†	
Asian		†	Lower overall posttest score
Part-time teacher	Lower mathematics posttest score	Lower language, mathematics, reading, and literacy posttest scores	Higher overall posttest score
Highest degree: professional certificate	†	†	Higher overall posttest score
Higher number of PD hours		Higher oral English posttest scores	
More years of experience	Higher overall posttest score	†	†
More years of adult education experience	Higher overall and mathematics posttest scores	Higher language and mathematics posttest scores	

† Data/category not available

Suggested Next Steps

This statistical brief and its associated briefs are the first few studies that focus on adult teachers and explore the relationship between teacher characteristics and student performance. Results of this study show that a variety of teacher characteristics are correlated with student achievement, though the magnitudes are small and the relationships can be inconsistent across states or subject areas. For example, we found inconsistent patterns of relationships across the three states on teacher employment status and PD participation. State policies and methods of data collection differ at the local level, and our models might not be stable without controlling such variations. Therefore, results from this brief should not be generalized without further analysis using data from other states.

To build a research base in adult education, it is critical to first construct a high-quality data system at both individual and aggregated levels. Currently, the National Reporting System requires states to report state-level data annually, which serve as the foundation for policy development and evaluation; yet, these data might not be sufficient for researchers to answer complicated research questions, especially when we want to drill down to the local program level. Constructing a longitudinal data system that allows for tracking students and teachers across years is important and is aligned with the data quality movement in K–16 education. This study provides valuable information on how student and teacher data are collected at the local level. The lack of standardization in data collection and the absence of longitudinal data systems limit the type of research questions that can be answered.

As discussed in detail in Section VI, “Recommendation on Data Collection,” we provide specific guidance on data collection and management to local programs. Once a reliable data system is in place, researchers should further investigate how teacher characteristics, especially employment status and PD participation, affect student performance. PD participation is one of the few malleable factors that is correlated with student performance and can be improved through coaching and technical assistance. Currently, states provide various PD offerings to teachers in the hope of improving their performances. Recording and reporting PD participation consistently will also greatly assist future research.

The adult education student population varies radically across the country. In our study, over 70% of students were Hispanic in State 2, while in State 1 only 30% of students were Hispanic. The significant difference in student population composition might be another reason why we did not detect consistent relationships between teacher characteristics and student performance. Different students have different needs in education and might require different teachers to better serve their needs. A study at the state level might be of interest to learn how teacher characteristics and student performance relationships vary across states. States that serve more disadvantaged students might require higher performing teachers or intense PD services to improve their student performance. Such studies will assist state policy makers in understanding the needs of their local programs and provide policy guidance at the state and program levels.

I. Introduction

In the last dozen years, the availability of administrative databases that track individual student achievement over time and link students to their teachers has radically altered how research on education is conducted and has brought fundamental changes to the ways educational programs and personnel are evaluated. The availability of student-level panel data is also fundamentally changing program accountability and the measurement of teacher performance. Since then, numerous studies have been conducted on the relationship between various aspects of teacher quality and student achievement in K–12 education.

The Coleman report (Coleman et al., 1966) examined the impact of a number of teacher background characteristics, including years of experience, education level, and performance on a vocabulary test, ultimately concluding that teacher background characteristics had a larger effect on student achievement than any other general class of school effects except student body composition.

Yet, the performance of adult education teachers and students has not been well studied. Adult education has a long history in the United States, and a range of providers has long been involved in educating adult students. Over 2 million adult students who would otherwise be left outside the educational system and who lack English language proficiency, secondary school education, and the skills necessary to enter postsecondary education or become more skilled employees are enrolled in the system. This brief, the second of four briefs on adult education teachers and the performance of adult students, analyzes the relationships between teacher characteristics and adult student academic achievement.

II. Theoretical Framework

As our brief literature review indicated, the most widely studied aspect of teacher effectiveness is teacher background characteristics. Several specific background characteristics have been examined in the research literature, including degrees, coursework, credentials, experience, test scores, and the prestige ratings of teachers' undergraduate institutions. Although individual studies have found that certain aspects of teachers' backgrounds are associated with student achievement or learning, comprehensive reviews of the research literature have produced inconsistent conclusions, and there does not appear to be a consensus opinion.

Goldhaber and Brewer (1997) found that having a master's degree did not make a difference among 10th grade mathematics teachers, unless the degree was in mathematics. Mathematics teachers with a bachelor's degree in mathematics also performed better than mathematics teachers whose degree was not in mathematics. Ehrenberg and Brewer (1994) found that the average selectivity of teachers' colleges had a positive relationship with student gain scores for White and Black high school students but not for Hispanic students. Two studies by Goldhaber and Brewer (1997, 2000) found a positive relationship between certification and student gains only for mathematics or science students and only when their mathematics or science teacher's certification was in the subject taught. It did not appear to matter whether that certification was standard or provisional, but standard certification was associated with higher gains than was private school certification in mathematics.

In addition to teacher demographics, a few studies also explored the relationship between teacher experience and student achievement. Gordon, Kane, and Staiger (2006) found large gains in teacher effectiveness between the first and second years of teaching, much smaller gains between the second and third years, and no substantial improvement after the third year in the classroom. Similarly, Hanushek, Kain, O'Brien, and Rivkin (2005) contended that the only important difference was between teachers with no experience and those with at least 1 year of experience. They estimated that having a first-year teacher was roughly equivalent to having a teacher half a standard deviation down in the quality distribution. Positive effects have also been found between PD and student outcomes for PD lasting more than 14 hours (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). An average of 49 hours of PD translated into test score gains of 21 percentile points. These studies encompassed limited contexts, however—most focused on elementary reading, and most of the PD was workshop based.

Existing research in K–12 education has examined not only which aspects of teacher quality matter but also how much *teachers* matter. That is, studies have attempted to determine the proportion of the variation in student achievement and learning that can be attributed to classroom or teacher effects as opposed to other sources (e.g., school effects, effects of individual and family background characteristics). A recent review of this literature by Nye, Hedges, and Konstantopoulos (2004) examined 18 analyses from 7 studies. They reported that the proportion of the variance in student achievement gains owing to teacher effects ranged from about 0.07 to 0.21.

As the first examination of adult education teacher effectiveness, this study focuses on four aspects of teachers: teacher demographics, teacher educational attainment, teacher PD, and teacher experience. We selected variables and built models on the basis of a multilevel theoretical framework that recognized the variation in student achievement gains owing to two distinct and nested levels: teacher level and student level.

III. Data

For this study, student-level data for the 2008–09, 2009–10, and 2010–11 program years were obtained directly from three states.[3] The student-level data varied in scope, but each state included information on teachers (see Table 1), student pre- and posttest scores, student demographics, educational functioning level, and local program attended. Two states also included data on program size, program type, and program support services.

Our main interest was in estimating the relationships between a vector of teacher characteristics and student academic achievement, but to do that, we first needed to match teachers with their students for each state. As discussed in detail in the appended Technical Notes, not all states have a unique identifier for their teachers, and co-teaching is very common in most states. To solve this problem, we selected only states with unique teacher identifiers for all their teachers for the years we studied and selected a primary teacher for each student on the basis of that

[3] We initially requested data from six states, and five of them provided student, teacher, and program files for the required years. Each state also provided a detailed data dictionary, which assisted us in selecting the appropriate variables for the study. However, after intense data cleaning and preparation, only the data from three states met all criteria for this study.

student's attendance hours with each teacher. The downside of this matching method was that we introduced biases into our estimation because we assumed that student gains could be attributed only to the primary teacher. If a student benefitted from a secondary teacher, however, our estimates would be upward biased because we attributed the gains to the primary teacher.

To carry out the proposed regression analysis (discussed in detail in Section IV, "Method"), we requested from each state a longitudinal data set that contained multiple observations for the same student across 3 years and multiple observations for the same teacher. The purpose of using a longitudinal data set was to control for potential nonobservable time-invariant teacher (e.g., teaching ability, skill), student (e.g., student IQ), and program (e.g., policies that do not change over time) characteristics that might have affected student academic achievement when we estimated the effects on observable teacher characteristics. After receiving the data, we noticed that adult students in these data entered and exited the program more frequently than do K–12 students. In many cases, states treated students who reentered the program as new students, which hindered us from tracking the same students across years. Note that the limited number of repeated students for each teacher in our data set affected the reliability of our estimates from the student fixed effects (FE) model.

During our data cleaning process, we also noticed that teacher and student demographic categories were not consistent across the three states. In addition, the states collected different teacher information, and the data provided to us varied greatly across states. To obtain the most information from our analyses, we did not combine or streamline variables or categories across states. Instead, we performed separate analyses for each state.

Another unique aspect of adult education data was that students could have taken multiple posttests at different points in their enrollment. To conduct our analysis, we elected to use the most recent posttest scores as our outcome variable to evaluate student achievement. The potential bias for this method was that if the student enrolled with a different teacher after the initial posttest, we would contribute the gains to the original teacher the student had. Fortunately, the states began to recognize the importance of linking students to their teachers, and we found that less than 10% of the students had multiple posttest scores.

Our total sample size of students in adult basic education (ABE), adult secondary education (ASE), and English as a second language (ESL) programs was around 723,000 students and nearly 13,000 adult education teachers over 3 program years. The analytical sample used for the main test score analysis for adult teachers and students in all programs and all years included over 300,000 students and over 5,000 teachers.[4]

In Table 1, we present all available teacher, student, and program variables by state and year. As the table shows, the availability of variables was not consistent across states. State 1 provided the richest list of teacher variables, which was ideal for our proposed study; however, the sample size in State 1 was rather small. More reliable inferences on teacher characteristics could be drawn from States 2 and 3, which had larger sample sizes of teachers and students.

[4] The analytical sample excludes duplicates, students that cannot be linked the their teachers, and cases with missing or miscoded outcome variables. A detailed description of how data were cleaned and managed is available in the Technical Notes.

We should also point out that teacher professional development (PD) is a key predictor in student achievement in K–12 education. In adult education, however, we do not have an ideal measure that can quantify both the quality and the quantity of PD participation. The PD participation measure is usually self-reported or recorded inconsistently. In our study, we used the number of hours of PD participation as a proxy. One should be cautious when drawing conclusions because this variable does not represent the quality of PD provided in each state or local program.

Table 1. Sample Size and the Availability of Teacher, Student, and Agency Variables, by State and Year

	2008–09			2009–10			2010–11		
	State 1	State 2	State 3	State 1	State 2	State 3	State 1	State 2	State 3
Number of teachers	100	1,901	2,443	102	1,787	2,475	104	1,864	2,360
Number of observations (student level)	1,872	100,629	136,866	2,263	93,088	142,078	2,511	99,996	143,773
Teacher Variables									
Age	√		√	√		√	√		√
Gender	√	√	√	√	√	√	√	√	√
Race	√		√	√		√	√		√
Educational attainment	√	√	√	√	√	√	√	√	√
Certification level		√			√			√	
Part-time/full-time	√	√	√	√	√	√	√	√	√
Total years of experience	√			√			√		
Total years of adult education experience	√	√	√	√	√	√	√	√	√
Total PD hours	√	√	√	√	√	√	√	√	√
Average number of classes taught per week	√			√			√		
Average number of hours taught per week	√			√			√		
Average number of students taught per week	√			√			√		
Average number of planning hours per week	√			√			√		
Adult education department (ABE, ASE, ESL, etc.)	√		√	√		√	√		√
Student Variables									
Age	√	√	√	√	√	√	√	√	√
Gender									
Race	√	√	√	√	√	√	√	√	√
Attendance hours	√	√	√	√	√	√	√	√	√
Educational attainment		√	√		√	√		√	√
Special needs (learning impairment; mental impairment; etc.)			√			√			√
Employment status			√			√			√
Number of instructors			√			√			√
English is second language			√			√			√
Orientation hours		√			√			√	
Residential area (urban, rural, and other)		√			√			√	
Preassessment NRS level	√		√			√	√		√
Agency Variables									
Urbanicity	√			√			√		
Agency/program size			√			√			√
Agency/program type	√		√	√		√	√		√
Agency/program performance			√			√			√
Support services	√			√			√		

In the adult education system, students are assigned to different educational functioning levels (EFLs; similar to grade levels in K–12 education) and are given different test instruments and forms accordingly. In Table 2, for example, we present student distribution across various educational levels in State 1. Half of the students in State 1 were concentrated in low-intermediate ABE and high-intermediate ABE. In Appendix A, Tables A1–A3, we illustrate different test instruments and test areas across all three participating states. Because students are placed into different EFLs on the basis of their prior educational attainment or ability and different test instruments and forms are based on their levels, we recognized the importance of creating standardized test scores that could be compared across years, EFLs, and states.[5] Therefore, our main outcome variables, adult students' pre- and posttest score differences and posttest scores, were standardized prior to analysis.

Table 2. Pretest Assessed Educational Functioning Level in State 1: 2010

Assessed Pretest Level	Frequency	Percentage
Beg Literacy ABE	67	2.70%
Beginning ABE	355	14.10%
Low Intermediate ABE	555	22.10%
High Intermediate ABE	593	23.60%
Low Adult Secondary	197	7.90%
High Adult Secondary	64	2.60%
Completed High Adult Secondary	1	0.00%
Beg Literacy ESL	220	8.80%
Low Beginning ESL	139	5.50%
Low Intermediate ESL	67	2.70%
High Beginning ESL	110	4.40%
High Intermediate ESL	130	5.20%
Advanced ESL	6	0.20%
Completed Advanced ESL	2	0.10%
Level Not Defined	5	0.20%
Total	**2,511**	**100**

IV. Method

Selecting the Appropriate Outcome Variable: Pre- and Posttest Score Gain vs. Posttest Score

In estimating teacher effectiveness, the literature distinguishes between test-score-gain-as-outcome models and lagged-performance models. In the test-score-gain-as-outcome models, the difference between the student's current-year performance and previous-year performance is used as the outcome variable. Therefore, using student test-score gain as the outcome variable, we assumed a perfect relationship between the previous-year performance and the current-year performance. The analytical models for these kinds of models have the following form:

[5] A detailed description of how standardization was carried out in each state is available in the Technical Notes.

$$A_{it} - A_{i(t-1)} = \Delta A_i = \beta_0 + \beta_1 X_{it} + \varepsilon_{it}$$

where A represents student achievement and the difference between the current year (i.e., *t*) and the previous year (i.e., *t* – 1) serves as the outcome for the analyses.

In contrast to test-score-gain-as-outcome models, lagged-performance models use the current-year performance as the outcome and the previous-year performance as a control variable in the analysis. By doing so, the lagged-performance models estimate an imperfect relationship between previous-year performance and current-year performance. This kind of analysis estimates an analytical model of the following form:

$$A_{it} = \beta_0 + \rho A_{i(t-1)} + \beta_1 X_{it} + \varepsilon_{it}$$

where ρ represents the decaying effect of the previous-year performance on the current-year performance. In this model, the relationship between the current-year performance and the previous-year performance is estimated to be a number between 0 and 1. The test-score-gain-as-outcome model, however, assumes that the relationship between current-year and previous-year performance is 1 (i.e., $\rho = 1$).

Even though these two models have different assumptions about the relationship between previous- and current-year performance, Harris and Sass (2006) have shown that in the K–12 setting, the score-gain model is a good approximation of the more complicated lagged dependent variable approach. Because no studies in the adult education literature use either model, let alone contrast the models, we compared these two models to find out which one was a better choice in the adult education context. Our preliminary results revealed that the correlation between pretest and posttest scores was so low that it was unreasonable to accept the assumption of the score-gain model that there is a perfect relationship between pretest and posttest performances (i.e., $\rho = 1$). Therefore, in the following analyses, we employed the lagged-performance model that uses posttest performance as the outcome while controlling for the pretest score.

Analytical Model 1: Ordinary Least Squares

As our first analytical model, which served as a baseline to compare the results from other more complicated models, we used an ordinary least squares (OLS) model with student posttest score as the outcome variable:

$$A_{ipost} = \beta_0 + \rho A_{ipre} + \sum_{p=1}^{n} \alpha_p T_{kt} + \varepsilon_{it}$$

where the subscripts *i*, *k*, and *m* denote individual students, teachers, and program site, respectively; A_{ipost} represents posttest performance of student *i*; A_{ipre} represents the same student’s pretest performance; and *T* is a vector of teacher characteristics. This model allowed us to have a first look at the unadjusted relationships between teacher characteristics and student achievement. However, it was obvious that student and program characteristics also played an important role in improving student performance. In the next model, we added in more controls and compared how estimates on teacher characteristics change:

$$A_{ipost} = \beta_0 + \rho A_{ipre} + \beta_1 X_{it} + \sum_{p=1}^{n} \alpha_p T_{kt} + \sum_{q=1}^{n} \pi_q P_{mt} + \varepsilon_{it}$$

where the subscripts *i*, *k*, and *m* denote individual students, teachers, and program site, respectively; A_{ipost} represents posttest performance of student *i*, achievement; A_{ipre} represents the same student's pretest performance; *X* is a vector of student characteristics; *T* is a vector of teacher characteristics; and *P* is a vector of program site–level characteristics.

This model estimated the relationship between a student's posttest performance and teacher characteristics while controlling for student characteristics and program site characteristics. However, the model did not take into account the nesting of students within teachers and might therefore have overstated the statistical significance of the results. Moreover, one might argue that innate student characteristics (e.g., student motivation) other than the ones we controlled for in our models might be related to a student's performance and that omitting these innate characteristics might bias the results.

Analytical Model 2: Student Fixed Effects Model

To control for student unobserved/inherent abilities, as our second model we used a student FE model:

$$A_{ipost} = \beta_0 + \rho A_{ipre} + \sum_{p=1}^{n} \alpha_p T_{kt} + \sum_{q=1}^{n} \pi_q P_{mt} + \gamma_i + \varepsilon_{it}$$

where, as previously, A_{ipost} represents posttest performance of student *i*, achievement; A_{ipre} represents the same student's pretest performance; *T* is a vector of teacher characteristics; *P* is a vector of program site characteristics; and γ_i is the student FE. Different from the previous OLS model, the student FE model did not include student characteristics because all of those variables were time invariant and were captured by the student FEs (i.e., γ_i). We compared the results from this model to those of the previous one to find out whether any unobserved student characteristics were related to student performance other than the ones we used in the OLS model.

Analytical Model 3: Teacher Random Effects Model (Two-Level Hierarchical Linear Modeling)

To take the nesting of students within teachers into account, as our last model we employed a two-level hierarchical linear modeling (HLM) model. The estimated model was of the following form:

Level 1: Student Model:

$$A_{ipost} = \beta_0 + \rho A_{ipre} + \beta_1 X_{it} + \varepsilon_{it}$$

Level 2: Teacher Model:

$$\beta_0 = \alpha_0 + \sum_{p=1}^{n} \alpha_p T_{kt} + \sum_{q=1}^{n} \pi_q P_{mt} + \nu_{kt}$$

where, as before, A_{ipost} represents the posttest performance of student i, achievement; A_{ipre} represents the same student's pretest performance; T is a vector of teacher characteristics; P is a vector of program site characteristics; and υ_{kt} is the teacher random effects (RE).

The results from this model revealed whether teacher characteristics were related to student performance when controlling for student and program characteristics and taking into account the nesting of the students within characteristics.

Results Comparison From Proposed Models

We began our analysis with a baseline OLS model that controlled for only teacher characteristics. This model allowed us to obtain unadjusted coefficients on all teacher variables of interest. By adding in more student and program variables, we were able to observe how coefficients changed. Using results obtained from State 1 as an example, Table 3 presents results from four proposed models. Our outcome variable was the student posttest score. Coefficients on pretest scores indicated that the correlation between pre- and posttest scores was much smaller than 1, confirming that a score-gain model was not appropriate for this study. Column 1 shows results from the OLS model with teacher controls only; column 2 shows results from the OLS model with student, teacher, and program controls; column 3 shows results from the student FE model; and the last column presents results from the teacher RE model (or two-level HLM where we clustered within teachers).

When we used the OLS model with teacher controls only, the coefficients on teacher age, Hispanic category, Native American category, total number of years of experience, part-time teacher status, and paid preparation time presented statistically significant effects. Teacher characteristics explained about 46% ($R^2 = 0.46$) of the variance in students' posttest scores. However, because student and program characteristics were not controlled in Model 1, some of the coefficients might have picked up effects from the uncontrolled variables and biased the estimates.

To test this possibility, we added student and program controls into Model 2 and found that the coefficients on teacher age, Native American category, and paid preparation time remained statistically significant with the same signs. However, total years of experience were no longer significant and had an opposite sign as estimated from Model 1. Similarly, part-time teachers performed significantly worse than full-time teachers according to Model 2, whereas the coefficient from Model 1 indicated a significant positive effect. The change of signs and magnitudes on coefficients from Model 2 to Model 1 implied that some effects of students and programs were picked up by teachers in Model 1, suggesting that Model 2 was more appropriate.

Although the OLS model with teacher, student, and program controls accounted for all characteristics observable and attainable by researchers, we cannot account for potential time-invariant, unobservable student learning abilities. If these student traits unobservable by researchers explained the true variation of student achievement, our estimates on teacher characteristics would be biased. Hence, in Model 3 we applied the student FE model and required multiple observations of the same student across 3 years. Unfortunately, adult students enter and exit programs very frequently, and the longitudinal data provided by our participating states did not give us enough power to estimate all teacher characteristics in the student FE model. As column 4 (Model 3) shows, only three of the teacher characteristics were estimated from the model, and we could not make any legitimate inferences for State 1 using this model. We performed the same analysis for the other states, which had much larger sample sizes, and present the results in the following section.

Last, we present results from the teacher RE (two-level HLM) model, where we controlled for teacher, student, and program characteristics and the fact that certain students are nested within the same teachers. We assumed that students were randomly assigned to different teachers and adjusted for the standard errors for each variable. Results from Models 4 and 2 are similar in terms of the coefficients' magnitudes and signs, but we observed less statistical significance because of the clustering within teachers. Overall, this is the model in which we have the most confidence, although we would have preferred to make our conclusion from student FE models while clustering within teachers. Data obtained from the states did not allow for this model.

Table 3. Regression Results From Four Proposed Models Using Data From State 1

State 1				
	(1)	(2)	(3)	(4)
VARIABLES	OLS With Teacher Controls Only	OLS With Teacher, Student, and Program Controls	Student FE Model	Teacher RE Model
Pretest score	0.651***	0.596***	0.394***	0.591***
	(0.0170)	(0.0335)	(0.0929)	(0.0423)
Female teacher	0.0155	0.0205		0.000271
	(0.0379)	(0.0502)		(0.0902)
Age	-0.00834***	-0.00664**		-0.00457
	(0.00222)	(0.00300)		(0.00509)
African American	-0.221	0.247		0.305
	(0.172)	(0.194)		(0.247)
Hispanic	-0.235***	-0.102		-0.310
	(0.0770)	(0.103)		(0.189)
Native American	0.0491	0.298***		0.115
	(0.0901)	(0.104)		(0.190)
Native American	0.171***	0.373***		0.329
	(0.0650)	(0.106)		(0.300)
Years of experience	0.0103***	-0.00511		-0.00407
	(0.00165)	(0.00322)		(0.00653)
Years of adult education experience	0.00307	0.0169***		0.0165**
	(0.00263)	(0.00410)		(0.00647)
Number of PD hours	0.000210	0.000201	-0.852*	0.000300
	(0.000888)	(0.00121)	(0.450)	(0.00185)

State 1				
	(1)	(2)	(3)	(4)
VARIABLES	OLS With Teacher Controls Only	OLS With Teacher, Student, and Program Controls	Student FE Model	Teacher RE Model
Highest degree: Master's	0.00722	-0.0311	-0.0193	0.0254
	(0.0305)	(0.0389)	(0.0190)	(0.0737)
Part-time teacher	0.0950**	-0.147**		-0.0278
	(0.0417)	(0.0749)		(0.125)
Paid preparation time	0.202***	0.198**	1.734**	0.192
	(0.0421)	(0.0864)	(0.753)	(0.174)
Observations	3,991	2,914	2,914	2,914
R-squared	0.462	0.499	0.239	0.4917

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Appendix A, we present full tables on regression coefficient results on key teacher, student, and program characteristics. Because teacher RE is our preferred model, where teacher characteristics can be estimated consistently, our discussion focuses on results from this model.

V. Results

Findings From State 1

Teacher Findings

Sample size in State 1 was rather small as shown in Table 4; however, the state collected valuable teacher information that was essential for this study. Therefore, we selected this state for the analyses, hoping to better understand the relationships between various teacher characteristics and student achievement.

The total number of adult education experiences of teachers was positively correlated with student performances. Specifically, 1 additional year as a teacher in adult education increased the student posttest score by 0.017 standard deviation. The number of hours taught per week was negatively correlated with student performance. One additional hour taught per week lowered the student posttest score by about 0.01 standard deviation.

In Table 5, we present results from analyses by subject area in State 1. Teacher age was found to be negatively correlated with students' oral English posttest scores; students assigned to Native American teachers scored higher in language and oral English posttests compared to students assigned to White teachers; students assigned to full-time teachers scored higher in mathematics assessment than students assigned to part-time teachers; teacher years of experience was estimated to be negatively correlated with students' mathematics and reading posttest scores; years of experience teaching adult education was found to be positively correlated with student mathematics posttest score; and having paid preparation time was negatively correlated with student oral English posttest score.

Student Findings

Though teachers were a focus of our study, we report findings from the model for students. We found that student age was negatively correlated with performance. That is, younger students had higher posttest scores relative to older students. African American students, Hispanic students, and Pacific Islander students had lower posttest scores than White students. Students attending programs that provided child family support and childcare services also had higher posttest scores.

Findings From State 2

Teacher Findings

Student achievement data provided by State 2 were by testing subject area (literacy, oral English, reading, mathematics, and language); therefore, we conducted all analyses by subject area, using our proposed models. Sample size in State 2 was fairly large, with approximately 80,000 students and more than 2,000 adult education teachers. Results are presented in Table 6.

Across all subject areas, female teachers performed significantly better than male teachers in adult education. Students in classes with teachers with a master’s degree had significant higher language and oral English posttest scores compared to students with teachers having just a bachelor’s degree. Students whose teachers had elementary, secondary, or both certifications had significantly higher oral English posttest scores. Each additional year of teachers’ adult education experience increased student language and mathematics posttest scores significantly.

Student Findings

Consistent with the findings from State 1, student age was negatively correlated with posttest scores for all subjects (younger students had higher posttest scores). African American students performed significantly lower on posttests than White students across all subjects. The scores of Hispanic students were significantly lower than those of White students in oral English, reading, mathematics, and language. In contrast, while Asian students scored higher on posttests than White students in mathematics, these students scored lower in oral English, reading, and language.

We also found for State 2 that student attendance hours were significantly positively correlated with student achievement for all subjects. The number of orientation hours was also positively correlated with student performance except for literacy, reading, and mathematics. Students who were currently employed tended to perform better than those who were unemployed, and the scores of those who resided in urban areas were lower in literacy, reading, mathematics, and language than those of students residing in rural areas.

Findings From State 3

Teacher Findings

The total number of student observations from State 3 was about 218,000, and the number of teachers in our analytical sample was nearly 3,500. The size of the sample allowed us to produce precise estimates using our teacher RE models. Results in Table 4 indicate that students in classes with Asian teachers had lower posttest scores than students in classes with White teachers. Having a teacher with higher educational attainment was positively associated with student posttest scores compared to having a teacher without a degree.

Student Findings

Student age was negatively correlated with student achievement in State 3, as well. White students outperformed all other racial groups on posttests. The number of hours of student attendance was positively correlated with student performance, which was consistent with results from States 1 and 2. However, unique to State 3, students who were unemployed or not in the labor force at the time performed significantly better on posttests than those who were employed. Students with mental, physical, learning, or multiple impairments had lower test scores than students who had no disabilities. Our results also showed that program size[6] was negatively correlated with student performance, whereas the overall program performance was positively correlated with student performance.

Table 4. Regression Results From Teacher RE Models Using Data From States 1 and 3

	State 1	State 2*	State 3
Female teacher	0.000271		0.00498
	(0.0902)		(0.0130)
Age	-0.00457		0.000412
	(0.00509)		(0.000462)
African American	0.305		0.0160
	(0.247)		(0.0190)
Hispanic	-0.310		-0.0179
	(0.189)		(0.0213)
Native American	0.115		†
	(0.190)		
Native American	0.329		-0.0165
	(0.300)		(0.112)
Asian	†		-0.0661*
			(0.0346)
Part-time teacher	-0.0278		0.0614***
	(0.125)		(0.0214)
Highest degree: GED	†		0.316***
			(0.0985)
Highest degree: Associate's	†		-0.0327
			(0.0910)
Highest degree: Bachelor's			0.137**

[6] Program size is measured as the number of students enrolled in the program.

	State 1	State 2*	State 3
			(0.0647)
Highest degree: Master's	N/A		0.132**
			(0.0648)
Highest degree: Ph.D.	0.0254		0.105
	(0.0737)		(0.0746)
Highest degree: professional certificate	†		0.291***
			(0.0875)
Number of PD hours	0.000300		-5.70e-05
	(0.00185)		(0.000390)
Years of experience	-0.00407		†
	(0.00653)		
Years of adult education experience	0.0165**		0.000808
	(0.00647)		(0.000530)
Paid preparation time	0.192		†
	(0.174)		
Number of classes per week	0.0141		†
	(0.0148)		
Number of hours per week	-0.0127*		†
	(0.00707)		
Number of students taught per week	0.000452		†
	(0.00317)		
Number of preparation hours per week	0.0272		†
	(0.0174)		
Observations	2,914		218,322
R-squared	0.4917		0.4658
Number of teachers	60		3,467

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Data not available.
* State 2 provided data by subject. Therefore, we were not able to perform overall regression analyses.
N/A: Number of observations is small; no coefficients are estimated.

Table 5. Regression Results From Teacher RE Models by Subject Area Using Data From State 1

State 1	Language	Mathematics	Oral English	Reading
Female teacher	0.204	0.000779	0.00169	0.149
	(0.168)	(0.132)	(0.138)	(0.274)
Age	0.00163	-0.0108	-0.0167**	0.0208
	(0.0102)	(0.00847)	(0.00773)	(0.0175)
African American	0.310	0.0994		0.920
	(0.298)	(0.235)		(0.612)
Hispanic	-0.167	0.330	0.238	-0.276
	(0.531)	(0.360)	(0.249)	(0.784)
Native American	1.140***		0.814***	0.401
	(0.348)		(0.277)	(0.997)
Native American	0.497*	0.373*		0.177
	(0.298)	(0.220)		(0.412)
Part-time teacher	0.137	-0.255*	-0.413	0.104

State 1	Language	Mathematics	Oral English	Reading
	(0.166)	(0.145)	(0.551)	(0.456)
Highest degree: Master's	0.0228	-0.0880	-0.0927	0.0518
	(0.103)	(0.0735)	(0.279)	(0.197)
Number of PD hours	0.00112	0.000555	-0.00368	0.00513
	(0.00204)	(0.00209)	(0.00464)	(0.00588)
Years of experience	-0.00819	-0.00930*	0.0270	-0.0242*
	(0.00773)	(0.00546)	(0.0187)	(0.0147)
Years of adult education experience	0.0102	0.0234***	-0.0401	0.0172
	(0.0111)	(0.00587)	(0.0304)	(0.0177)
Paid preparation time	0.335	0.321	-1.439***	0.168
	(0.277)	(0.228)	(0.425)	(0.561)
Number of classes per week	0.0241	-0.0106	-0.0391***	0.0255
	(0.0357)	(0.0171)	(0.0141)	(0.0498)
Number of hours per week	-0.0244***	-0.0193***	0.00565	-0.0246
	(0.00914)	(0.00725)	(0.0205)	(0.0159)
Number of students taught per week	-0.000650	0.000863	-0.0160**	-0.0143
	(0.00535)	(0.00416)	(0.00746)	(0.0119)
Number of preparation hours per week	0.0984**	-0.00745	0.0276	0.0637
	(0.0425)	(0.0287)	(0.0389)	(0.0864)
Observations	596	1,094	757	315
R-squared	0.478	0.555	0.505	0.486
Number of teacher	42	41	25	39

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Data/category not available
* State 2 provided data by subject. Therefore, we were not able to perform overall regression analyses.

Table 6. Regression Results From Teacher RE Models by Subject Area Using Data From State 2

State 2	Language	Mathematics	Oral English	Reading	Literacy
Female teacher	0.0472**	0.0442**	0.0233	0.0355*	0.0350**
	(0.0211)	(0.0217)	(0.0198)	(0.0207)	(0.0148)
Part-time teacher	-0.0504*	-0.0723**	-0.0479	-0.0842***	-0.0519**
	(0.0286)	(0.0306)	(0.0351)	(0.0313)	(0.0206)
Highest degree: Master's	0.0444**	-0.00290	0.0403*	0.00970	0.0166
	(0.0211)	(0.0222)	(0.0240)	(0.0205)	(0.0142)
Highest degree: Ph.D.	-0.0734	0.106	0.0852	0.0408	0.0572
	(0.0826)	(0.0746)	(0.106)	(0.0694)	(0.0622)
Teacher certification: elementary	-0.0189	0.00359	0.0791***	-0.00857	-0.00254
	(0.0254)	(0.0269)	(0.0254)	(0.0250)	(0.0197)
Teacher certification: secondary	-0.0169	0.0589**	0.0867**	0.0287	0.0154
	(0.0270)	(0.0261)	(0.0357)	(0.0250)	(0.0186)

State 2	Language	Mathematics	Oral English	Reading	Literacy
Teacher certification: both	-0.00436	0.0446	0.0641**	0.00366	0.0369
	(0.0340)	(0.0335)	(0.0298)	(0.0310)	(0.0280)
Number of PD hours	-0.000124	-0.000367	0.000801**	-0.000473	-0.000149
	(0.000338)	(0.000323)	(0.000388)	(0.000289)	(0.000277)
Years of adult education experience	0.00239*	0.00258*	0.000241	0.00187	0.00169
	(0.00141)	(0.00152)	(0.00187)	(0.00142)	(0.00105)
Observations	38,675	40,171	77,925	37,416	53,812
R-squared	0.126	0.224	0.340	0.188	0.543
Number of teachers	1,542	1,467	2,097	1,646	1,934

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Data/category not available
* State 2 provided data by subject. Therefore, we were not able to perform overall regression analyses.

Further Analyses of Teacher-Student Race and Ethnicity Matching

In the K–12 education literature, many studies have shown that race plays a major role in student-teacher interaction (Brewer, Ehrenberg, & Goldhaber, 1994; Downey & Pribesh, 2004; Oates, 2003). Over the years, there has been a great deal of discussion in the literature about matching teacher and student by race, ethnicity, gender, and language. Some people claimed that race should not be considered when organizing or evaluating the classroom and that, therefore, matching should not even be considered. Others were willing to consider matching but disagreed about whether it was effective or not (Muller, 1998; Oates, 2003). Oates (2003) suggested that the match between teacher and students shaped teacher perceptions of student performance. Some studies suggested that White teachers assessed White students as more academically engaged than Black students (Downey & Pribesh, 2004; Ferguson, 2003). Other studies suggested that White teachers gave Black students worse evaluations than did Black teachers (Brewer et al., 1994; Downey & Pribesh, 2004; Ferguson, 2003).

To test whether teacher and student race matching plays any role in adult education, we performed analyses using data from two participating states that provided both teacher and student race/ethnicity variables. We created interaction terms between each student and teacher race combination to test whether there was any differential effect when estimated jointly. We performed four models (OLS, student FE models, teacher RE model, and teacher FE model) and report results from our preferred teacher RE and teacher FE models,[7] as shown in Tables 8 and 9. We found that in State 1, when students were assigned to teachers of the same race, only the White students and White teacher combination improved adult student performance significantly. When Hispanic students were assigned to Black teachers and White students to Hispanic teachers, the posttest scores also increased significantly compared to those of other racial combinations. These results were robust across teacher RE and teacher FE models (Table 7).

[7] Results from other models are available upon request.

Table 7. Results From State 1: Testing for Teacher-Student Race Matching

VARIABLES	Teacher RE Model	Teacher FE Model
Black teacher vs. Black student	0.307	0.173
	(0.276)	(0.125)
Hispanic teacher vs. Hispanic student	0.0973	0.117
	(0.0898)	(0.0952)
Black teacher vs. Hispanic student	1.007***	0.869***
	(0.296)	(0.0980)
Hispanic teacher vs. Black student	0.0823	0.107
	(0.147)	(0.148)
White teacher vs. White student	0.151*	0.170*
	(0.0843)	(0.0876)
White teacher vs. Black student	0.0810	0.110
	(0.159)	(0.163)
White teacher vs. Hispanic student	-0.00488	0.0233
	(0.116)	(0.121)
Hispanic teacher vs. White student	0.331***	0.345***
	(0.0761)	(0.0789)
Observations	2,914	2,914
R-squared	0.4905	0.441
Number of teachers	60	60

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We found that when Hispanic students were assigned to Hispanic teachers, their posttest scores increased significantly in State 3. In addition, when Hispanic students were assigned to White teachers and White students to Hispanic teachers, students' posttest scores were significantly higher than with other racial combinations (Table 8).

Table 8. Results From State 3: Testing for Teacher-Student Race Matching

VARIABLES	Teacher RE Model	Teacher FE Model
Black teacher vs. Black student	0.00523	0.0513
	(0.0541)	(0.0562)
Hispanic teacher vs. Hispanic student	-0.0488	0.00842
	(0.0458)	(0.0488)
Black teacher vs. Hispanic student	-0.0170	0.0485
	(0.0432)	(0.0464)
Hispanic teacher vs. Black student	-0.0458	-0.000105
	(0.0650)	(0.0676)
White teacher vs. White student	-0.0549	-0.00209
	(0.0352)	(0.0380)
White teacher vs. Black student	0.0113	0.0577
	(0.0467)	(0.0491)
White teacher vs. Hispanic student	-0.0215	0.0347
	(0.0340)	(0.0378)

VARIABLES	Teacher RE Model	Teacher FE Model
Black teacher vs. White student	-0.0287	0.0331
	(0.0450)	(0.0471)
Hispanic teacher vs. White student	-0.0433	0.00424
	(0.0488)	(0.0509)
Observations	211,670	211,670
R-squared	0.3225	0.323
Number of teachers	4,011	4,011

VI. Recommendations on Data Collection

To answer the research questions posed in this brief, it was essential to obtain appropriate data from the state adult education system. The NRS requires all states to have a student-level record system for reporting outcomes, attendance, and characteristics of students who attend federally funded adult education and literacy programs. The quality of NRS data systems has improved over the years as advances in technology have made data systems less expensive and more accessible. Likewise, the quality of the NRS data has improved as states gained more experience in collecting and reporting data. Consequently, a rich body of data exists among the states and local programs that can be used for secondary data analyses to answer research and policy questions.

However, using NRS data for research purposes is not straightforward. Adult education data systems in most states are designed not for research but for annual reporting to OCTAE. Also, the data systems often contain only NRS-required data elements, and the quality and subsequent usability of data vary across states. To carry out the proposed study, we recruited five states that have student longitudinal data and allow student-teacher matching. Among the five participating states, we selected three that possessed the most complete teacher information. As we cleaned and prepared the data set for analysis, we discovered a few issues with the data collection process at the state level. Therefore, we provide the following recommendations that may help states better maintain their data system and provide higher quality data for future research. Qualitative and quantitative researchers typically rely on data to conduct statistical analysis and make inferences on the basis of results derived from the data. We recognize that the current adult education data system is not designed for research purposes.

- **Use consistent categories for teachers' and students' demographic data.** Currently, states collect data that are based on their individual needs and reporting purposes. There are no standard data categories at the federal level to guide the data collection process. For instance, some states categorize their teachers into seven racial categories (White, African American, Hispanic, Native American, Native Indian, Asian, Other), whereas others categorize all teachers into four categories (White, African American, Hispanic, Other). For teacher and student education, the categories used are also not consistent within a state and across states. Having consistent categories is important not only for analytical purposes when states evaluate their own teachers and students but also for comparing their students and teachers to those of other states on different measures.
- **Create unique teacher identifiers to link student data to specific teachers.** Different from K–12 education, co-teaching is very popular in adult education, which presents a

great hurdle for researchers who are evaluating teacher effectiveness. In addition, not all states have a unique identifier for each teacher that can be used to link to student data. If the state database cannot link individual teachers to students, it is impossible to relate teacher effectiveness directly to student outcomes or attendance. Consequently, policy makers and researchers cannot effectively evaluate the performance of individual teachers.

- **Improve state longitudinal data systems.** To examine teacher effectiveness over time, researchers need longitudinal data, which will allow them to follow the same students and teachers across years. There is a growing need to establish state longitudinal data systems for reporting and research purposes. Although the states that participated in our study possessed high-quality longitudinal data systems, we noticed inconsistencies when cleaning the data sets. For instance, states might not have a unique identifier for every student. When such students exit and reenter the program, they are treated as new students, which might bias analyses because they will be treated as a separate observation.
- **Have a separate category for missing data.** States record teacher PD and teacher experience. However, it is unclear to researchers whether teachers having 0 hours of PD participation or 0 years of experience have missing data or in fact have not participated in any PD or have no prior teaching experience. It is common in education data sets to have missing data. Many statistical methods are available to adjust for biases caused by missing data. However, if missing data are mistakenly treated as "0" categories in the data collection and analytical processes, biases might be introduced into the conclusions.
- **Avoid self-reported data.** Self-reported data have been shown to lead to biases in statistical analysis. The direction of biases depends on the variable. For instance, some states use student self-reported attendance hours to evaluate the relationship between attendance and performance. Students tend to overestimate their attendance hours, which might lead to upward bias when estimating the correlation of attendance with student achievement. The more reliable alternative is to record students' participation through a third party (e.g., teacher, program director) and combine information to calculate total attendance hour.

VII. Conclusion

This study as the first attempt to explore teacher effectiveness using advanced statistical models revealed the importance of teacher quality in promoting student academic performance. Teacher gender, race, educational attainment, PD participation, and experience were found to be correlated with student performances in one or two states; however, we did not find consistent patterns across three states. It is important to recognize the limitations of our study, and we urge further research to confirm our findings.

We faced multiple data challenges when conducting the proposed analytical models that are currently commonly used in the teacher value-added literature. Among all, the lack of longitudinal data systems that allowed for perfect teacher-student matching was the biggest hurdle. Although we originally planned to estimate teacher FE models where we control for both observable and unobservable teacher characteristics and to estimate a composite score for each adult education teacher, the available data did not allow us to do so. Hence, the core of this study explored the correlation between observable teacher characteristics and student academic achievement.

The limited literature on teacher quality in adult education provided little information about which teacher variables should be included in our analytical models. High-quality research that can guide policy formation and implementation in the adult education field is needed. To conduct such research, states need to improve their data collection process. They need guidance on what data elements to include and how to record their data. For instance, states currently do not collect teacher information consistently across programs and years. Categories used to record teacher race and educational attainment are also different across states.

Due to the data quality and inconsistency across states, interpretation and conclusions drawn from our study should be applied with caution. Because all our analyses were conducted within each participating state, we do not recommend generalization. Additional studies using smaller scale but better quality data are needed to confirm our findings. For instance, researchers can use longitudinal data from local educational providers to conduct teacher value-added analysis. It is important to observe students and teachers for several years to analyze how teacher characteristics impact student achievement.

Our current study is the first step in exploring teacher quality in adult education. We have identified some basic teacher variables related to student test score gains, but further analyses are needed to confirm these findings and also to address research questions such as whether teacher quality affects student postsecondary entry and labor market outcomes. A central goal of adult education is to bridge the gap between K–16 education, further education, and the labor market. An understanding of the role of teachers—their characteristics, PD, and other factors—in promoting these outcomes will help adult education programs achieve this goal.

Appendix A

Using data from program year 2010–11, we illustrate different test instruments and test areas across all three participating states. State 1 used four test instruments covering five main subject areas (Table A1), with the TABE most used, followed by BEST Plus.

Table A1. Cross-Tabulation of Test Instrument and Subject Area in State 1: 2010

Instrument	Subject Area					
	Literacy	Language	Mathematics	Oral English	Reading	Total
BEST Literacy (BEST 2)	388	1	0	2	3	394
BEST Plus	2	0	0	278	0	280
GED	0	0	6	0	0	6
TABE	0	593	999	0	239	1,831
Total	390	594	1,005	280	242	2,511

State 2 provided data on test instrument or test form by subject area. Because students could have taken multiple tests, we were unable to combine this information. Instead, in Table A2 we present a cross-tabulation of test instrument and student educational functioning level (EFL) by subject. The BEST test was given exclusively to English as a second language (ESL) students, and the TABE was given to adult basic education (ABE) and adult secondary education (ASE) students.

Table A2. Cross-Tabulation of Test instrument and Educational Functioning Level by Subject Area in State 2: 2010

Literacy EFL	Literacy Assessment Instrument			
	BEST-B	BEST-C	BEST-D	Total
Advanced ESL	191	162	107	460
ESL Beginning High	6,722	4,164	1,735	12,621
ESL Beginning Literacy	1,606	790	312	2,708
ESL Beginning Low	14,630	7,371	2,994	24,995
ESL Intermediate High	2,964	2,302	1,066	6,332
ESL Intermediate Low	2,138	1,714	675	4,527
Total	28,251	16,503	6,889	51,643

Oral English EFL	Oral Assessment Instrument	
	BEST Plus	Total
Advanced ESL	7,350	7,350
ESL Beginning High	5,408	5,408
ESL Beginning Literacy	28,271	28,271
ESL Beginning Low	4,565	4,565
ESL Intermediate High	5,922	5,922

	Oral Assessment Instrument	
Oral English EFL	**BEST Plus**	**Total**
ESL Intermediate Low	7,502	7,502
Total	59,018	59,018

	Reading Assessment Instrument		
Reading EFL	**TABE 10**	**TABE 9**	**Total**
ABE Beginning Basic E	778	3,049	3,827
ABE Beginning Literacy	228	790	1,018
ABE Intermediate High	2,719	7,925	10,644
ABE Intermediate Low	2,167	7,419	9,586
ASE High	1,313	4,746	6,059
ASE Low	1,544	4,171	5,715
Total	8,749	28,100	36,849

	Language Assessment Instrument		
Language EFL	**TABE 10**	**TABE 9**	**Total**
ABE Beginning Basic E	2,025	6,689	8,714
ABE Beginning Literacy	469	1,898	2,367
ABE Intermediate High	2,060	7,009	9,069
ABE Intermediate Low	1,907	5,986	7,893
ASE High	1,133	2,807	3,940
ASE Low	1,098	2,761	3,859
Total	8,692	27,150	35,842

	Mathematics Assessment Instrument		
Mathematics EFL	**TABE 10**	**TABE 9**	**Total**
ABE Beginning Basic E	869	3,983	4,852
ABE Beginning Literacy	101	269	370
ABE Intermediate High	3,542	9,416	12,958
ABE Intermediate Low	2,570	9,395	11,965
ASE High	704	1,672	2,376
ASE Low	924	2,345	3,269
Total	8,710	27,080	35,790

In State 3, students were given different test instruments based on their EFL only. Because the tests were not differentiated by subject area, we were advised not to use the test scores by subject area. Instead, State 3 provided combined test scores, test forms, and test instruments (Table A3).

Table A3. Cross-Tabulation of Test Instrument and Educational Functioning Level in State 3: 2010

	Pretest Type					
Student's NRS level	**BEST Lite**	**BEST Plus**	**CELSA**	**TABE-M**	**TABE-R**	**Total**
ABE Beginning Basic Education	0	0	0	101	5,018	5,119
ABE Beginning Literacy	0	0	0	25	1,479	1,504
ABE Intermediate High	0	0	0	265	21,089	21,354
ABE Intermediate Low	0	0	0	321	14,959	15,280
ASE High	0	0	0	26	10,022	10,048
ASE Low	0	0	0	57	11,677	11,734
ESL Advanced	286	246	9,153	0	0	9,685
ESL Beginning Literacy	6,160	4,077	0	0	0	10,237
ESL High Beginning	8,588	859	2,656	0	0	12,103
ESL Intermediate High	3,115	568	10,009	0	0	13,692
ESL Intermediate Low	2,412	980	4,775	0	0	8,167
ESL Low Beginning	24,234	616	0	0	0	24,850
Total	44,795	7,346	26,593	795	64,244	143,773

Table A4. Full Regression Table: State 1

	(1)	(2)	(3)
VARIABLES	**OLS**	**Student FE**	**Teacher RE**
Pretest score	0.596***	0.394***	0.591***
	(0.0335)	(0.0929)	(0.0423)
Female teacher	-0.0205		-0.000271
	(0.0502)		(0.0902)
Age	-0.00664**		-0.00457
	(0.00300)		(0.00509)
African American	0.247		0.305
	(0.194)		(0.247)
Hispanic	-0.102		-0.310
	(0.103)		(0.189)
Native American	0.298***		0.115
	(0.104)		(0.190)
Native American	0.373***		0.329
	(0.106)		(0.300)
Years of experience	-0.00511		-0.00407
	(0.00322)		(0.00653)
Years of adult education experience	0.0169***		0.0165**
	(0.00410)		(0.00647)
Number of PD hours	0.000201	-0.852*	0.000300
	(0.00121)	(0.450)	(0.00185)
Highest degree: Master's	-0.0311	-0.0193	0.0254
	(0.0389)	(0.0190)	(0.0737)
Part-time teacher	-0.147**		-0.0278
	(0.0749)		(0.125)
Paid preparation time	0.198**	1.734**	0.192
	(0.0864)	(0.753)	(0.174)
Number of classes per week	0.0148	-1.849**	0.0141
	(0.00979)	(0.880)	(0.0148)
Number of hours per week	-0.0216***		-0.0127*

VARIABLES	(1) OLS	(2) Student FE	(3) Teacher RE
	(0.00370)		(0.00707)
Number of students taught per week	-0.00110		0.000452
	(0.00168)		(0.00317)
Number of preparation hours per week	0.0275***	-1.768***	0.0272
	(0.00945)	(0.647)	(0.0174)
Student: Age	-0.00257*		-0.00259**
	(0.00137)		(0.00121)
Student: Attendance hours	0.000297***		0.000399***
	(0.000113)		(0.000132)
Student: Black	-0.129***		-0.139***
	(0.0473)		(0.0433)
Student: Hispanic	-0.0964**		-0.0838**
	(0.0433)		(0.0406)
Student: Native American	-0.0990**	-0.0789	-0.0815
	(0.0449)	(0.130)	(0.0497)
Student: Pacific Islander	-0.340*	-0.0350	-0.314*
	(0.187)	(0.0230)	(0.190)
Student: two races	0.0393	-0.0343*	0.0241
	(0.169)	(0.0176)	(0.203)
Program in urban area	-0.0668	-0.837**	-0.214
	(0.0959)	(0.401)	(0.188)
Agency provides child family support	0.692***	-0.788**	0.406**
	(0.144)	(0.399)	(0.187)
Agency provides childcare	0.590***	-0.165	0.345*
	(0.122)	(0.328)	(0.184)
Agency provides transportation	0.458***	-1.052***	
	(0.136)	(0.399)	
Agency type: public school	-0.0833	-0.0682	
	(0.0774)	(0.324)	
Agency type: community college			0.126
			(0.137)
Observations	2,914	2,914	2,914
R-squared	0.499	0.239	0.4917

Table A5. Full Regression Table: State 2

Literacy	(1)	(2)	(3)
	OLS Literacy	Student FE Literacy	Teacher RE Literacy
Literacy pretest score	0.801***	0.135***	0.769***
	(0.00381)	(0.0136)	(0.00699)
Female teacher	0.0330***	0.0171	0.0350**
	(0.00603)	(0.0143)	(0.0148)
Highest degree: Master's	0.00442	-0.00889	0.0166
	(0.00611)	(0.0147)	(0.0142)
Highest degree: Ph.D.	0.00378	-0.0443	0.0572
	(0.0239)	(0.0634)	(0.0622)
Teacher certification: elementary	-0.00949	-0.0127	-0.00254
	(0.00707)	(0.0171)	(0.0197)
Teacher certification: secondary	0.00703	-0.0101	0.0154
	(0.00768)	(0.0176)	(0.0186)
Teacher certification: both	0.0152	-0.0370	0.0369
	(0.00970)	(0.0255)	(0.0280)
Part-time teacher	-0.0334***	0.0126	-0.0519**
	(0.00811)	(0.0220)	(0.0206)
Years of adult education experience	0.00124***	-1.49e-05	0.00169
	(0.000418)	(0.00106)	(0.00105)
Number of PD hours	-7.17e-05	-0.000199	-0.000149
	(0.000111)	(0.000235)	(0.000277)
Student: Age	-0.00352***	-0.0123	-0.00401***
	(0.000262)	(0.0173)	(0.000297)
Student: Black	-0.215***		-0.143***
	(0.0408)		(0.0423)
Student: Hispanic	-0.142***		-0.0322
	(0.0244)		(0.0251)
Student: Asian	-0.0790***		0.0117
	(0.0269)		(0.0278)
Student: Hawaiian	0.0879		0.152*
	(0.0976)		(0.0918)
Student: attendance hours	0.000594***	0.000363***	0.000579***
	(2.79e-05)	(6.17e-05)	(4.27e-05)
Student: orientation hours	0.00276***	-0.00653***	0.00180
	(0.000689)	(0.00167)	(0.00121)
Student: never attend school	-0.0169	0.155***	-0.0809***
	(0.0151)	(0.0342)	(0.0225)
Student: obtain degree from other country	-0.0400**	0.0404	-0.0550***
	(0.0185)	(0.0370)	(0.0210)
Student: reside in urban area	-0.0954***	-0.0460	-0.0259*
	(0.00873)	(0.0338)	(0.0145)
Student: reside in urban area with high unemployment rate	-0.00229	-0.0364	0.0223
	(0.0113)	(0.0443)	(0.0218)
Student: reside in other area	-0.108***	-0.0382	-0.0974***
	(0.0232)	(0.0763)	(0.0301)
Student: unemployed	0.0478***	-0.0183	0.0210**
	(0.00884)	(0.0247)	(0.00974)
Student: not in labor force	0.000764	-0.0245	0.00185
	(0.00596)	(0.0201)	(0.00663)
Observations	53,812	53,812	53,812
R-squared	0.619	0.254	0.543
Number of staff			1,934
Number of participants		45,504	

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Oral English	**(4)**	**(5)**	**(6)**
	OLS Oral	**Student FE Oral**	**Teacher RE Oral**
Oral pretest score	0.692***	0.151***	0.650***
	(0.00385)	(0.0101)	(0.00817)
Female teacher	0.00827	0.0344***	0.0223
	(0.00599)	(0.0125)	(0.0198)
Highest degree: Master's	0.00105	0.0334**	0.0403*
	(0.00638)	(0.0136)	(0.0240)
Highest degree: Ph.D.	-0.000495	0.0283	0.0852
	(0.0274)	(0.0830)	(0.106)
Teacher certification: elementary	0.0340***	-0.0116	0.0791***
	(0.00730)	(0.0152)	(0.0254)
Teacher certification: secondary	0.0688***	0.0199	0.0867**
	(0.00822)	(0.0167)	(0.0357)
Teacher certification: both	0.0784***	0.0204	0.0641**
	(0.0104)	(0.0223)	(0.0298)
Part-time teacher	-0.140***	-0.00525	-0.0479
	(0.00873)	(0.0217)	(0.0351)
Years of adult education experience	0.000246	-0.000561	0.000241
	(0.000427)	(0.000819)	(0.00187)
Number of PD hours	0.000790***	5.47e-05	0.000801**
	(0.000122)	(0.000253)	(0.000388)
Student: age	-0.00406***	-0.0233	-0.00403***
	(0.000258)	(0.0153)	(0.000293)
Student: Black	-0.162***		-0.111***
	(0.0389)		(0.0421)
Student: Hispanic	-0.207***		-0.143***
	(0.0281)		(0.0296)
Student: Asian	-0.205***		-0.147***
	(0.0302)		(0.0304)
Student: Hawaiian	-0.0688		-0.0487
	(0.122)		(0.113)
Student: attendance hours	0.000574***	0.000395***	0.000718***
	(3.17e-05)	(5.78e-05)	(5.86e-05)
Student: orientation hours	0.0102***	-0.0103***	0.00534***
	(0.000749)	(0.00172)	(0.00154)
Student: never attend school	-0.106***	-0.0568	-0.134***
	(0.0175)	(0.0360)	(0.0258)
Student: obtain degree from other country	-0.116***	0.00380	-0.104***
	(0.0208)	(0.0441)	(0.0222)
Student: reside in urban area	-0.0886***	0.00334	-0.00518
	(0.00926)	(0.0314)	(0.0163)
Student: reside in urban area with high unemployment rate	0.00535	0.00859	0.0586**
	(0.0120)	(0.0405)	(0.0246)
Student: reside in other area	-0.215***	-0.103	-0.117***
	(0.0272)	(0.0736)	(0.0399)
Student: unemployed	0.0310***	-0.00561	0.00814
	(0.00934)	(0.0222)	(0.00962)
Student: not in labor force	-0.0193***	-0.0196	-0.0215***
	(0.00604)	(0.0173)	(0.00688)
Observations	77,925	77,925	77,925
R-squared	0.423	0.175	0.34
Number of staff			2,097
Number of participants		65,142	

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Reading	(7)	(8)	(9)
	OLS R-reading	Student FE Reading	Teacher RE Reading
Reading pretest score	0.446***	0.185***	0.429***
	(0.00618)	(0.0208)	(0.00824)
Female teacher	0.0546***	0.120***	0.0355*
	(0.0105)	(0.0418)	(0.0207)
Highest degree: Master's	-0.00338	-0.0420	0.00970
	(0.0104)	(0.0408)	(0.0205)
Highest degree: Ph.D.	0.0233	0.00846	0.0408
	(0.0300)	(0.117)	(0.0694)
Teacher certification: elementary	-0.0178	-0.0432	-0.00857
	(0.0128)	(0.0524)	(0.0250)
Teacher certification: secondary	-0.0115	-0.0167	0.0287
	(0.0125)	(0.0517)	(0.0250)
Teacher certification: both	0.0286*	0.0414	0.00366
	(0.0163)	(0.0721)	(0.0310)
Part-time teacher	-0.0858***	-0.0509	-0.0842***
	(0.0127)	(0.0580)	(0.0313)
Years of adult education experience	0.00329***	-0.00559*	0.00187
	(0.000687)	(0.00302)	(0.00142)
Number of PD hours	-0.000520***	-0.00121**	-0.000473
	(0.000179)	(0.000595)	(0.000289)
Student: age	-0.00400***	-0.126***	-0.00356***
	(0.000402)	(0.0382)	(0.000444)
Student: Black	-0.138***		-0.150***
	(0.0165)		(0.0170)
Student: Hispanic	-0.139***		-0.106***
	(0.0138)		(0.0152)
Student: Asian	-0.130***		-0.173***
	(0.0316)		(0.0331)
Student: Hawaiian	-0.0964		-0.111
	(0.0983)		(0.0926)
Student: attendance hours	0.000273***	0.000442***	0.000275***
	(4.81e-05)	(0.000148)	(6.71e-05)
Student: orientation hours	-7.80e-05	-0.000272	0.00153
	(0.00105)	(0.00422)	(0.00158)
Student: never attend school	-0.113*	-0.0404	-0.199***
	(0.0611)	(0.142)	(0.0722)
Student: obtain degree from other country	-0.0855**	-0.00563	-0.131***
	(0.0405)	(0.101)	(0.0424)
Student: reside in urban area	-0.0384***	0.0945	-0.0584***
	(0.0125)	(0.0883)	(0.0189)
Student: reside in urban area with high unemployment rate	-0.0903***	-0.0279	-0.0978***
	(0.0172)	(0.122)	(0.0252)
Student: reside in other area	-0.300***	0.188	-0.185***
	(0.0373)	(0.254)	(0.0633)
Student: unemployed	0.0109	0.0107	0.00395
	(0.0130)	(0.0601)	(0.0132)
Student: not in labor force	-0.00984	-0.00822	-0.0101
	(0.0108)	(0.0503)	(0.0125)
Observations	37,416	37,416	37,416
R-squared	0.213	0.062	0.188
Number of staff			1,646
Number of participants		33,781	

Robust standard errors in parentheses; *** *p* < 0.01, ** *p* < 0.05, * *p* < 0.1

Mathematics	(10)	(11)	(12)
	OLS Mathematics	Student FE Mathematics	Teacher RE Mathematics
Mathematics pretest score	0.473***	0.166***	0.455***
	(0.00571)	(0.0186)	(0.00760)
Female teacher	0.0459***	0.0143	0.0442**
	(0.00990)	(0.0422)	(0.0217)
Highest degree: Master's	-0.00753	-0.0229	-0.00290
	(0.00969)	(0.0443)	(0.0222)
Highest degree: Ph.D.	0.107***	0.216	0.106
	(0.0291)	(0.136)	(0.0746)
Teacher certification: elementary	-0.0249**	-0.0228	0.00359
	(0.0120)	(0.0510)	(0.0269)
Teacher certification: secondary	-0.000812	-0.0239	0.0589**
	(0.0116)	(0.0493)	(0.0261)
Teacher certification: both	0.0573***	0.0505	0.0446
	(0.0152)	(0.0647)	(0.0335)
Part-time teacher	-0.0794***	-0.0528	-0.0723**
	(0.0120)	(0.0607)	(0.0306)
Years of adult education experience	0.00357***	-0.00227	0.00258*
	(0.000636)	(0.00319)	(0.00152)
Number of PD hours	-0.000305*	-0.000673	-0.000367
	(0.000171)	(0.000535)	(0.000323)
Student: age	-0.00414***	-0.0773**	-0.00339***
	(0.000393)	(0.0368)	(0.000445)
Student: Black	-0.178***		-0.181***
	(0.0149)		(0.0150)
Student: Hispanic	-0.110***		-0.0768***
	(0.0117)		(0.0127)
Student: Asian	0.132***		0.107***
	(0.0364)		(0.0380)
Student: Hawaiian	-0.102		-0.112
	(0.101)		(0.0974)
Student: attendance hours	0.000413***	0.000414***	0.000484***
	(4.89e-05)	(0.000159)	(7.51e-05)
Student: orientation hours	-0.00135	0.000477	0.000837
	(0.000953)	(0.00368)	(0.00152)
Student: never attend school	-0.135**	0.0403	-0.184**
	(0.0687)	(0.190)	(0.0786)
Student: obtain degree from other country	-0.0518	0.150	-0.0694
	(0.0452)	(0.105)	(0.0481)
Student: reside in urban area	-0.0484***	-0.0397	-0.0501***
	(0.0112)	(0.0917)	(0.0175)
Student: reside in urban area with high unemployment rate	-0.124***	-0.167	-0.136***
	(0.0165)	(0.120)	(0.0309)
Student: reside in other area	-0.381***	-0.147	-0.167
	(0.0372)	(0.272)	(0.105)
Student: unemployed	-0.0129	0.0569	-0.00200
	(0.0118)	(0.0497)	(0.0122)
Student: not in labor force	-0.00843	0.0738	-0.00670
	(0.0102)	(0.0454)	(0.0110)
Observations	40,171	40,171	40,171
R-squared	0.246	0.057	0.224
Number of staff			1,467
Number of participants		36,449	

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Language	**(13)**	**(14)**	**(15)**
	OLS Language	**Student FE Language**	**Teacher RE Language**
Language pretest score	0.354***	0.0918***	0.343***
	(0.00609)	(0.0191)	(0.00773)
Female teacher	0.0769***	0.0904**	0.0472**
	(0.0107)	(0.0436)	(0.0211)
Highest degree: Master's	0.0257**	0.0775*	0.0444**
	(0.0106)	(0.0428)	(0.0211)
Highest degree: Ph.D.	-0.0312	-0.0612	-0.0734
	(0.0325)	(0.180)	(0.0826)
Teacher certification: elementary	-0.0558***	-0.0585	-0.0189
	(0.0130)	(0.0546)	(0.0254)
Teacher certification: secondary	-0.0557***	-0.101*	-0.0169
	(0.0129)	(0.0527)	(0.0270)
Teacher certification: both	0.0160	-0.128*	-0.00436
	(0.0168)	(0.0689)	(0.0340)
Part-time teacher	-0.0386***	0.102	-0.0504*
	(0.0131)	(0.0652)	(0.0286)
Years of adult education **experience**	0.00366***	0.00187	0.00239*
	(0.000677)	(0.00304)	(0.00141)
Number of PD hours	-0.000163	-0.000290	-0.000124
	(0.000181)	(0.000577)	(0.000338)
Student: age	-0.00333***	-0.0911**	-0.00305***
	(0.000414)	(0.0404)	(0.000463)
Student: Black	-0.161***		-0.166***
	(0.0167)		(0.0185)
Student: Hispanic	-0.129***		-0.101***
	(0.0133)		(0.0150)
Student: Asian	-0.0340		-0.0508
	(0.0323)		(0.0334)
Student: Hawaiian	0.0742		0.0738
	(0.0903)		(0.0883)
Student: attendance hours	0.000337***	0.000366**	0.000330***
	(4.88e-05)	(0.000150)	(6.47e-05)
Student: orientation hours	0.00287***	-0.00539	0.00410**
	(0.00106)	(0.00396)	(0.00166)
Student: never attend school	-0.0302	0.0666	-0.103
	(0.0622)	(0.162)	(0.0671)
Student: obtain degree from other country	-0.0222	-0.0114	-0.0618
	(0.0412)	(0.117)	(0.0436)
Student: reside in urban area	-0.0399***	-0.147*	-0.0717***
	(0.0123)	(0.0852)	(0.0189)
Student: reside in urban area with high unemployment rate	-0.0997***	-0.0378	-0.0892***
	(0.0184)	(0.139)	(0.0309)
Student: reside in other area	-0.346***	-0.475	-0.290***
	(0.0362)	(0.331)	(0.0703)
Student: unemployed	0.0163	0.129**	0.0201
	(0.0133)	(0.0600)	(0.0138)
Student: not in labor force	-0.0200*	0.0987**	-0.0131
	(0.0110)	(0.0490)	(0.0116)
Observations	38,675	38,675	38,675
R-squared	0.142	0.033	0.126
Number of staff			1,542
Number of participants		34,916	

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A6. Full Regression Table: State 3

VARIABLES	(1) OLS	(2) Student FE	(3) Teacher RE
z-Score pretest	0.517***	0.189***	0.479***
	(0.00246)	(0.00797)	(0.00765)
Female teacher	-0.00527	0.0142**	0.00498
	(0.00323)	(0.00672)	(0.0130)
Teacher: age	0.00117***	0.000356	0.000412
	(0.000129)	(0.000261)	(0.000462)
Teacher: Black	0.0247***	-0.00855	0.0160
	(0.00483)	(0.0118)	(0.0190)
Teacher: Hispanic	-0.00697	0.0257**	-0.0179
	(0.00504)	(0.0119)	(0.0213)
Teacher: Asian	-0.0984***	-0.0664***	-0.0661*
	(0.00899)	(0.0219)	(0.0346)
Teacher: Native American	-0.0281	0.167***	-0.0165
	(0.0230)	(0.0577)	(0.112)
Teacher: Pacific Islander	-0.120***	-0.138	0.0201
	(0.0371)	(0.0937)	(0.159)
Teacher: other race	0.0593***	-0.0225	0.0869**
	(0.0111)	(0.0217)	(0.0342)
Highest degree: GED	0.267***	-0.0225	0.316***
	(0.0418)	(0.0601)	(0.0985)
Highest degree: Associate’s	0.0686**	-0.0171	-0.0327
	(0.0312)	(0.0610)	(0.0910)
Highest degree: Bachelor’s	0.121***	-0.0209	0.137**
	(0.0281)	(0.0581)	(0.0647)
Highest degree: Master’s	0.133***	-0.0272	0.132**
	(0.0281)	(0.0583)	(0.0648)
Highest degree: Ph.D.	0.0678**	-0.0167	0.105
	(0.0300)	(0.0625)	(0.0746)
Highest degree: professional certificate	0.217***	0.00914	0.291***
	(0.0456)	(0.0794)	(0.0875)
Part-time teacher	0.0419***	-0.0216**	0.0614***
	(0.00489)	(0.00904)	(0.0214)
Years of adult education experience	0.000640***	0.000130	0.000808
	(0.000136)	(0.000264)	(0.000530)
Number of PD hours	0.000488***	0.000124	-5.70e-05
	(0.000135)	(0.000253)	(0.000390)
Student: age	-0.00507***	1.227***	-0.00549***
	(0.000121)	(0.0398)	(0.000233)
Student: Black	-0.162***		-0.165***
	(0.00489)		(0.00916)
Student: Hispanic	-0.132***		-0.118***
	(0.00415)		(0.00845)
Student: Asian	-0.136***		-0.120***
	(0.00636)		(0.0112)
Student: **Native American**	-0.167***		-0.147***
	(0.0266)		(0.0337)
Student: other race	-0.0874***	-0.0652	-0.0430
	(0.0198)	(0.0745)	(0.0307)
Student: Pacific Islander	-0.0546		-0.0408
	(0.0527)		(0.0703)
Student: attendance hours	0.000510***	0.000787***	0.000738***
	(1.44e-05)	(6.05e-05)	(4.32e-05)
Student: part-time	0.00758*	-0.0129	0.00383

VARIABLES	(1) OLS	(2) Student FE	(3) Teacher RE
	(0.00454)	(0.0156)	(0.00648)
Student: unemployed	0.0242***	0.00753	0.0140**
	(0.00350)	(0.0143)	(0.00548)
Student: not in labor force	0.0173***	0.00231	0.0136*
	(0.00495)	(0.0191)	(0.00764)
Student: learning impaired	-0.0378	0.0631	-0.214***
	(0.0299)	(0.0979)	(0.0235)
Student: multiple disability	-0.0424		-0.171***
	(0.0375)		(0.0458)
Student: no disability	0.207***	0.0760	0.0137
	(0.0267)	(0.101)	(0.00876)
Student: physically impaired	0.117***	-0.00805	-0.0683**
	(0.0326)	(0.120)	(0.0283)
Student: mentally impaired		-0.0741	-0.127***
		(0.122)	(0.0465)
Program size	5.98e-06***	4.27e-05**	-1.49e-05***
	(7.74e-07)	(1.78e-05)	(5.55e-06)
Program performance	0.616***	0.0259	0.395***
	(0.0154)	(0.0822)	(0.0663)
Program type: community college	-0.136***		-0.0960***
	(0.00569)		(0.0212)
Program type: **community-based organization**	-0.368***		-0.403***
	(0.0200)		(0.0693)
Program type: local education agency	-0.0798***		-0.0577**
	(0.00658)		(0.0237)
Observations	298,818	218,322	218,322
R-squared	0.443	0.141	0.4658
Number of instructors			3,467
Number of students		130,522	

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Technical Notes

Data Cleaning Procedures

State 1

Student, teacher, and course data were received in several Microsoft Excel files and exported into the data file format to be used in the analyses. Files were grouped under teacher, student, demographics, and student outcomes (assessment scores, GED, employment, and postsecondary). Each group contained several files: 5 files for teacher data, 4 files for student demographics data, and 11 files for student outcomes data. To merge the data sets within and across groups, we examined each data file to identify unique identifiers that we could use to merge information across data sets. During this process, we communicated with the state several times to ensure that the variables identified to be used in linking information from different data files were accurate. All variables used in the analyses were checked for out-of-range data and data entry errors.

Separate data files for teachers' demographic, experience, and professional development (PD) information were received. If a teacher had more than one record (i.e., degree, PD), data were transformed into a format with only one observation per teacher, with the multiple observations collapsed into categories such as highest degree obtained and total number of PD hours. Then, all the teacher's files were merged, resulting in one teacher file with a unique teacher identification number. Because the purpose of the analyses was to study the importance of teacher background qualifications for student learning, teachers with all the demographic and experience variables missing were excluded from the data set. These teachers were mostly volunteers because data files did not include background information for volunteer teachers.

Similar to teacher data, separate files for student demographics, attendance, assessment, and job market outcomes were received. Duplicate records containing student background characteristics data were removed using the unique student identification numbers. If a student had multiple records (i.e., monthly attendance data), data were collapsed into one record per student in terms of total number of days attended in a school year. Individual files for student data were merged to create one student file that included a unique student identifier.

States 2 and 3

We worked directly with States 2 and 3 in preparing the data sets to be used in the analyses. States were first asked to fill out a survey of information available in their systems. The survey was divided into three subsections: teacher variables, class- and program-level information, and student variables. Teacher variables included information about teacher demographics, experience, education, and PD. Class- and program-level variables included information on course (e.g., type, educational functioning level) and program characteristics. Student variables included information on student demographics, assessments, attendance, and job market outcomes. After states reported back on the availability of data, they were informed about how the data sets should be constructed and were sent mock data files. States were asked to follow the decisions we made when processing State 1 data. Specifically, they were asked to do the following:

- Provide a data set that included one observation per student and teacher.
- Identify a primary teacher. If more than one teacher taught the class, identify the teacher who taught the class most in terms of numbers or hours taught as the primary teacher for that class.
- Identify a primary class for each student. If a student was enrolled in more than one course in the same subject, identify the class that the student attended most in terms of number of enrollment hours as the primary course.
- Identify pretest and posttest scores for each assessment.

During this process, we communicated with the states and answered their questions to clarify the type of information needed in the data as well as the format of the data files.

Student, Teacher, and Class Data Match

Examining teacher effectiveness by using student test scores as the outcome required linking student data with teacher data and class data, using unique identifiers. In the following section, we describe how we created the final data files we used in our analyses.

State 1

The data we received included unique teacher identification numbers, student identification numbers, and course identification numbers. As mentioned earlier, we merged separate teacher files (background variables, experience, and PD) using teacher identification numbers. Similarly, we used student identification numbers to merge student data files.

Class data files had unique course identifiers that were associated with teacher identifiers for those teachers teaching the course. However, some courses were taught by more than one teacher. In these cases, we identified a primary teacher by using the numbers of hours taught and assigning the teacher with the highest number as the primary teacher for the class. We used only the observations where we could identify the primary teacher.

The last piece of information needed to create a cross-walk between teachers, students, and class was the link between students and classes. For this purpose, we used the data file that identified what class or classes a student was taking as well as his or her attendance. If a student was enrolled in more than one class within a subject, we used the attendance information to assign a primary class to the student. The course that the student attended most was identified as the primary course, and the records for the other course were excluded from the data set.

Finally, using the data that linked teachers to courses and the data that linked students to courses, we were able to create a cross-walk file that linked students to teachers. Using this cross-walk, we merged student and teacher files and created one file that had student-, teacher-, and class-level information to be used in the analyses.

The final step in data preparation was figuring out the pretest and posttest so that we could measure students' growth. Student assessment data included multiple records for the assessment scores without identifying the pretest and posttest. To identify the pretest and posttest, scale scores were sorted within a year, subject, educational functioning level, and student by the assessment date. The score from the earliest test was determined to be the pretest, and the score from the latest test taken was assigned to be the posttest.

States 2 and 3

As part of working together with States 2 and 3 to prepare the data sets to be used in the analyses, we informed the states about how to link the student, teacher, and class data files to create one data set for each school year. States 2 and 3 were asked to do the following:

- Provide a data set that included one observation per student and information for the primary teacher-, class-, and program-level information. If a student was enrolled in more than one subject, states were asked to enter the information related to that assessment in separate variables (i.e., reading scale score, mathematics scale score).
- Identify a pretest and a posttest for each assessment. States were asked to enter information for pretests and posttests on different variables (i.e., pretest reading scale score, posttest reading scale score).

During this process, we answered questions from the states and clarified how the data files should be formatted. States 2 and 3 provided separate files for each school year (2008, 2009, and 2010) in Microsoft Excel format. These files were transferred into the data format to be used in the analyses.

Data were checked for inconsistencies and out-of-range responses. Variable names and formats (e.g., numeric, string) across years were standardized. Then, we combined the data from different years into one file for each state that included year information.

Standardization of Student Pre- and Posttest Scores

Our main outcome variables were students' pre- and posttest scores. Because the test type and test form[8] used generally differed across educational functional levels and years, it was necessary to standardize test scores prior to analysis. Specifically, all test scores were standardized to *z* scores. We first calculated the mean μ, and standard deviation σ of a set of scores by test type (t), form (f), functioning level (l), and year (y). We then standardized each individual i's score, x_i, by converting it into a *z* score by using the following formula on each individual score:

$$z_{i,t,f,l,y} = \frac{x_i - \mu_{t,f,l,y}}{\sigma_{t,f,l,y}} \qquad (1)$$

where $z_{i,t,f,l,y}$ is the standardized score for student i using test type t and test form f in functioning level l and year y; $\mu_{t,f,l,y}$ is the mean score for test type t and test form f in functioning level l and year y; and $\sigma_{t,f,l,y}$ is the standard deviation for test type t and test form f in functioning level l and year y.

[8] We standardized individual test scores within each state because data availability varied. In State 1, test scores are standardized by year, subtest form, and instrument; in State 2, test scores are standardized by year and educational functioning level; and in State 3, pretest scores are standardized by year, pretest form, and pretest type, while posttest scores are standardized by year, posttest form, and posttest type.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public Schools. *Journal of Labor Economics*, *25*, 95–135.

Brewer, D. J., Ehrenberg, R. G., & Goldhaber, D. D. (1994). *Do teachers' race, gender, and ethnicity matter? Evidence from NELS 88*. Paper presented at the National Bureau of Economic Research.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267–282.

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review, 13*(1), 1–17.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test K–12 test score gap. *Urban Education, 38*(4), 460.

Goldhaber, D., & Brewer, D. (1997, July). Evaluating the effect of teacher degree level on educational performance. In W. Fowler (Ed.), *Developments in school finance, 1996* (pp. 199–210). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs97/97535l.pdf

Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–146.

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: The Brookings Institution.

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005, February). *The market for teacher quality* (NBER Working Paper 11154). Cambridge, MA: National Bureau of Economic Research.

Harris, D., & Sass, T. R. (2006). *Value added models and measurement of teacher quality*. Unpublished manuscript, Florida State University at Tallahassee.

Kane, T., Rockoff, J., & Staiger, D. (2006, April). *What does certification tell us about teacher effectiveness? Evidence from New York City* (NBER Working Paper 12155). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12155.pdf

Muller, C. (1998). The minimum competency exam requirements: Teachers' and students' expectations, and academic performance. *Social Psychology of Education, 2,* 192–216.

Nye, B., Hedges, L., & Konstantopoulos, S. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237–257.

Oates, G. (2003). Teacher-student racial congruence, teacher perceptions, and test performance. *Social Science Quarterly, 84*(3), 508–525.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf

ABOUT AMERICAN INSTITUTES FOR RESEARCH

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

LOCATIONS

Domestic

Washington, D.C.

Atlanta, GA

Austin, TX

Baltimore, MD

Cayce, SC

Chapel Hill, NC

Chicago, IL

Columbus, OH

Frederick, MD

Honolulu, HI

Indianapolis, IN

Metairie, LA

Naperville, IL

New York, NY

Rockville, MD

Sacramento, CA

San Mateo, CA

Waltham, MA

International

Egypt

Honduras

Ivory Coast

Kyrgyzstan

Liberia

Tajikistan

Zambia

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Making Research Relevant