Journal of Teacher Education
2020, Vol. 71(2) 247–260

DOI: 10.1177/0022487119849564
journals.sagepub.com/home/jte
SAGE

A National Study of the Differential Impact of Novice Teacher Certification on Teacher Traits and Race-Based Mathematics Achievement

Kenneth Alonzo Anderson[1]

Abstract

In this study, differential prediction of student outcomes by race and teacher traits relative to the certification levels of novice teachers was assessed. Overall, algebra achievement was higher for students who were taught by teachers with standard certificates relative to students who were taught by novice teachers with nonstandard certificates. The most conservative estimates show that findings are equivalent to approximately 8 months of additional instruction for students who were taught by teachers with standard certificates. However, the benefits of being taught by a teacher with standard certification did not translate to underrepresented racial groups. Overall, there were several differences in dispositions across certification conditions. With respect to underrepresented racial groups, only one disposition was significantly different across conditions. Teachers with nonstandard certificates reported higher emphasis on increasing mathematics interests. For underrepresented racial groups, relationships between both certification conditions and achievement were underwhelming. Recommendations to improve teacher effectiveness are provided.

Keywords

teacher preparation, school/teacher effectiveness, educational policy, mathematics teacher education, achievement

In the United States, pervasive achievement inequity exists for various racial groups. Teacher preparation programs (TPPs) can play an important role in preparing novice teachers to curtail existing achievement disparities, but little is known about how student outcomes differ by race across TPPs. TPPs often adopt various models, which include diversity courses and field experiences that are designed to develop dispositions for effective teaching in racially diverse contexts. However, much is to be learned about relationships between TPPs, teacher beliefs and dispositions, and student outcomes by race.

In this study, I summarize the large-scale literature on the relationships between TPPs, teacher dispositions, and student outcomes. I also identify some gaps in the literature and address these gaps using Boykin and Noguera's (2011) research-based scheme for promoting enhanced classroom teaching and learning to examine relationships between certification status, teacher beliefs and dispositions, and achievement disparities in mathematics for underrepresented racial groups[1]. I conclude by offering some recommendations for improving TPP capacity to address racial achievement differences.

Review of Literature

There is often great debate about the best pathways for prospective teachers to enter the profession. Despite vigorous debate about teacher pathways, Sleeter (2014) contends that due to the small amounts of available evidence, advocacy is often based on ideology. In a review of 196 studies from four top teacher education journals, Sleeter (2014) found that only 6% of the articles in 2012 examined the effects of teacher education on teaching practice or student outcomes. Although there are some limitations in the scope of current research, several large-scale studies have examined relationships between TPPs and student outcomes.

Certification Status and Student Outcomes

In a study that compared 20 university-based TPPs in Washington (state) to out-of-state certified teachers, researchers found small, but meaningful standard test score differences across TPPs (Goldhaber, Liddle, & Theobold, 2013). The authors found that differences in mathematics achievement between the least and most effective TPPs can be larger than expected differences between students who are eligible for free and reduced lunch and those who are not. In addition,

[1]Howard University, Washington, DC, USA

Corresponding Author:
Kenneth Alonzo Anderson, Department of Curriculum and Instruction, Howard University, 2441 4th St. NW, Washington, DC 20059, USA.
Email: kenneth.anderson@howard.edu

the authors found that differences in reading achievement between the least and most effective TPPs can be larger than expected differences between students with learning disabilities relative to students without learning disabilities. Washington is not a state with large amounts of teachers who enter the profession through alternative pathways, so the authors excluded uncertified teachers in the analysis. As a preface to the results, the authors provided a cautionary note indicating that student subgroups, such as race and gender, accounted for meaningful achievement differences as well. These factors were accounted for in the models, but were not discussed further.

Other studies have investigated the relationships between TPPs and student achievement as well. In a series of studies using data from North Carolina, researchers found that teacher credentials are associated with higher student achievement, but also found that teacher credentials are inequitably distributed across schools (Clotfelter, Ladd, & Vigdor, 2006, 2010; Clotfelter, Ladd, Vigdor, & Wheeler, 2010). Findings showed that students in high poverty schools were consistently likely to be taught by teachers who earned lower scores on certification exams, did not have a regular license, or had not earned National Board Certification. Likewise, in a study of New York teachers, researchers found that initial mathematics and English language arts student achievement gains were smaller for teachers who completed reduced coursework relative to teachers who completed a university-based teacher education program (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006). Similar to Goldhaber et al.'s (2013) study, the authors account for subgroup differences in their models, but did not analyze relationships between TPPs and subgroups.

Beyond state-level studies, researchers have used nationally representative samples to show that teacher certification is positively correlated with higher achievement. Using data from more than 2,70,000 students and 10,000 schools, Lubienski, Lubienski, and Crane (2008) examined differences between public and private schools across five domains: school size, class size, school climate/parental involvement, teacher certification, and instructional practices. The authors analyzed the data using a series of 12 progressive multilevel models. One of the models examined four teacher background variables: certification status, major, first year or not, and the number of professional development activities completed. Of the teacher variables, only the certified teacher indicator, was positively and significantly related to student achievement in fourth and eighth grades. The authors also concluded that race and socioeconomic factors were the strongest predictors of achievement, but did not offer commentary or analysis relative to the relationships between teacher certification and demographics.

Positive relationships between teacher certification and student achievement have also been documented in international contexts outside of the United States as well. In a study of more than 5,80,000 students in Swedish compulsory schools, authors found that when the share of noncertified teachers increases by 1%, standardized grade point averages of students decrease by 1.8 standard deviations (Andersson, Johansson, & Waldenstrom, 2011). The authors concluded that these differences were large enough to alleviate gender differences in achievement as well as disadvantages that may be associated with immigrant status.

Contrary to findings noted above, some studies have found no differences in achievement across certification pathways. In a study of approximately 2,600 students across 63 schools in the United States over 2 years, researchers found no differences in mathematics and reading achievement between students taught by alternatively certified versus traditionally certified teachers (Constantine et al., 2009). Likewise, in a study of Arabian countries in the Gulf Cooperation Council (GCC), researchers did not find consistent relationships between certification status and test scores reported on the Trends in Mathematics and Science Study (Wiseman & Al-baker, 2013). The authors indicated that many GCC TPPs were not rigorously evaluated on a consistent basis, making it challenging to evaluate their effects. Some other factors that were associated with higher achievement in the GCC study included teacher interactions and their abilities to motivate and teach students. Similarly, Constantine et al. also acknowledged that individual teachers have differential effects on student outcomes, but were not able to identify the traits of the teacher that were associated with different levels of student achievement.

Teacher Dispositions and Student Outcomes

In the previous section, I cited research that documented relationships between TPPs and student achievement. However, more research on teacher-specific characteristics that improve achievement for underrepresented racial groups is necessary. One class of teacher traits to consider is that of dispositions. Dispositions can be defined as

> attributed characteristics of a teacher that represent a trend of a teacher's judgments and actions in ill-structured contexts (situations in which there is more than one way to solve a dilemma; even experts disagree on which way is best). Further, it is assumed that these dispositions, trends in teacher judgments and actions, develop over time when teachers participate in deliberate professional education programs . . . (Reiman & Johnson, 2007, p. 677)

In some of my prior work, I demonstrate value in assessing relationships between self-reported teacher dispositions and student outcomes. Using principal components analysis and regression analysis of national survey data, I found that when comparing high-achieving Black students to all other high-achieving students, self-reported teacher beliefs were significantly related to mathematics scores of high-achieving non-Black students. Yet, teacher beliefs were not significantly related to mathematics achievement of high-achieving

Black students, although teachers reported higher beliefs about high-achieving Black students. Teachers of non-Black high-achieving students also reported higher levels of mathematics content training than those of high-achieving Black students (K. A. Anderson, 2014). These findings highlighted potential differential prediction of race-based teacher dispositions for high-achieving students.

Research has also shown that structural features of traditional TPPs, such as well-designed student teaching practica, are related to teacher effectiveness, dispositions, and teacher efficacy (D'alessio, 2018; Ronfeldt, 2012, 2015; Ronfeldt & Reininger, 2012). Qualitative studies have shown that the assessment of teacher dispositions is an essential factor for enhancing outcomes for underrepresented racial groups (Hill-Jackson & Lewis, 2010; Hollins, 2011). TPPs often offer courses and practicum experiences to build competence with racially diverse populations. However, teacher candidates, regardless of pathway, often resist discussions and program requirements that focus on race. Resistance is often related to individuals fearing that race discussions may lead to conflicts with colleagues, unfair judgment, and labeling for sharing honest opinions or experiences, or that the United States is reaching a postracial era (Milner, 2015). Milner and Laughter (2015) found that teachers have little confidence in their abilities to discuss race and are less willing to talk about race relative to their own teaching practices and identities. Using critical race theory, they found that teachers generally exhibited five common mind-sets. Specifically, teachers believe the following:

1. Acknowledging one's own racial background or the background of one's students may lead to racist labeling.
2. Treating all students the same regardless of their racial or ethnic backgrounds is appropriate.
3. Teaching children and ignoring the race of students is the right course of action because race is irrelevant.
4. Focusing on race while teaching does not matter because racism has ended.
5. Promoting postracial practices are appropriate because we live in a postracial society.

In a different study of preservice teacher dispositions, Hill-Jackson and Lewis (2010) found that dispositions of White preservice teachers, the dominant group of preservice teachers, could be categorized into two categories regarding multicultural education, advocates and resisters. Furthermore, they described five dispositions associated with advocates and resisters. Advocates may exhibit high cognitive complexity, a multifocal worldview, empathy, ethical traits, and high self-efficacy. Resisters may exhibit low cognitive complexity, a unifocal worldview, apathy, immoral traits, and low self-efficacy. The findings noted in this section led to recommendations for reformation of traditional and nontraditional TPPs to enhance dispositions about race and culture (Hill-Jackson & Lewis, 2010; Milner & Laughter, 2015).

Improving Achievement Outcomes for Underrepresented Racial Groups

Boykin and Noguera (2011) published a framework, the research-based scheme for promoting enhanced classroom teaching and learning, in their text, *Creating the Opportunity to Learn*. This framework is well-suited to address some of the critiques and findings noted in the previous sections. Their work, which is a compilation of works, emphasizes the importance of creating opportunities to learn (OTL) by summarizing primary and secondary research that is predictive of improving achievement for Black and Hispanic/Latino students. Boykin and Noguera's research-based scheme identifies three key domains: student engagement, classroom-based asset-focused factors, and guiding functions.

Boykin and Noguera (2011) indicated that student engagement includes cognitive, affective, and behavioral engagement. Cognitive engagement focuses on mental processes for comprehending complex concepts and acquiring difficult skills. Affective engagement describes emotional reactions to tasks, and behavioral engagement refers to general on-task behaviors. The authors cite several studies that documented the predictive relationships of engagement to student achievement. One notable study, using national early childhood data, demonstrated that engagement had higher effects on math gains than did instructional time (DiPerna, Lei, & Reid, 2007).

Asset-focused factors describe the "contextual conditions in which teaching, learning, engagement, and guiding functions manifest," (Boykin & Noguera, 2011, p. 69). Specific examples of asset-focused factors include interpersonal relationships between teachers and students, teachers' ability to personalize and use cultural resources in teaching situations, and teachers' ability to scaffold and support information processing. In their text, the authors describe an asset-focused factor, known as intersubjectivity, by focusing on two key components: meaningful learning and cultural concerns. Intersubjectivity "speaks to how well the values, interests, and learning priorities of the teacher are aligned with those of the students and the extent to which these aligned emphases are reflected in the curriculum," (p. 91).

The authors provide examples of how cultural mismatch in the classroom, initiated by teachers, may affect student–teacher relationships. For example, the authors describe the communal learning preferences of many Black students. Based on several hundred hours of classroom observation, the authors found that individual and competitive classroom learning situations was most commonly observed and initiated by the teacher (Boykin, Tyler, & Miller, 2005). They also found that when communal learning was evident, it was almost always initiated by Black students, not the teacher, and was associated with negative teacher feedback. When describing their framework, the authors also cited several randomized experiments that demonstrated how culture affects achievement (Boykin, Lilja, & Tyler, 2004; Hurley, Boykin, & Allen, 2005; Hurley, Allen, & Boykin, 2009). In one of these examples, the authors assigned students to one of the three randomized conditions, interpersonal conditions with contingent rewards, group competition with contingent rewards, and communal conditions with no rewards. They found that Black students performed significantly better than White students on math estimation tasks under the communal condition without rewards and White students performed significantly better than Black students on mathematics estimation tasks under the interpersonal competition condition with rewards (see Hurley, Allen, & Boykin, 2009).

Another key factor noted in Boykin and Noguera's (2011) framework is that of guiding functions. Guiding functions include a cluster of factors that "can steer, shape, govern, and intensify fundamental engagement processes," (p. 51). Examples of guiding functions include self-efficacy, self-regulated learning, beliefs about ability, and persistence. The authors cite several examples of guiding functions and how they may manifest in the classroom. One example describes the dichotomies of fixed versus malleable beliefs about one's own ability. Describing a different randomized experiment of Black and Latino seventh graders, the authors show how beliefs about ability can affect how students respond in times of failure. They summarize work that contends that students with a fixed belief may be less likely to persist in the face of failure or may be more occupied with maintaining self-respect. Contrarily, students who maintain malleable beliefs about learning are more likely to try to learn from high-achieving students, are more task persistent, and are less anxious about novel learning experiences (see Blackwell, Trzesniewski, & Dweck, 2007).

Strengths of Boykin and Noguera's (2011) framework are that many of the factors in the research-based scheme for promoting enhanced classroom teaching and learning have direct implications for teacher certification programs and are closely linked to teacher dispositions and efficacy. In their text, Boykin and Noguera reinforce several key components that are essential to OTL. Specifically, the authors close by describing the need to overdetermine the success of underrepresented groups. They provide recommendations for teachers to develop dispositions and practices that maximize daily learning opportunities. Some of these dispositions and practices include paying attention to classroom transactions and dynamics, focusing on student assets, educating the whole child, and providing multiple pathways to success. Moreover, the authors advocate for teacher development opportunities that ensue over time, provide constructive feedback, and include job-embedded support.

Many of the components of the research-based scheme for promoting enhanced classroom teaching and learning are consistent with overarching goals of many TPPs. In fact, at least one known teacher certification program has implemented components of Boykin and Noguera's framework into their model. However, most of the evidence for the efficacy of Boykin and Noguera's framework has been documented using research on inservice teachers. No known large-scale studies investigate links between preservice teacher training and student outcomes, after their first years of teaching, based on components of Boykin and Noguera's framework. Given that Boykin and Noguera's framework represent a compilation of works, many of the components, include teacher dispositions and efficacy goals that are common to many TPPs. Some of these components include increasing mathematics interests, making content connections to business and industry, engaging students, developing problem-solving skills of students, and embedding historical and cultural relevance.

Additional Opportunities for Large-Scale Teacher Preparation Research

Large-scale research that assesses interactions between TPPs, teacher dispositions, and student outcomes by race is limited. Teacher and student voice, through self-report, is an underutilized and low-cost approach to assessing teacher dispositions and assessing relationships to student achievement, thereby highlighting opportunities for additional large-scale teacher preparation research. Researchers have found that simply asking students questions can provide valuable insight and reduce bias in estimates in educational settings. When analyzing results from random and nonrandom assignment of psychology students to a mathematics versus vocabulary training condition, researchers found that topical preference was a superior and necessary domain for reducing bias, especially relative to mathematics outcomes (Steiner, Cook, Shadish, & Clark, 2010). In addition, the authors found that the combination of demographics, proxy pretests, and prior academic achievement performed similarly in predicting student choices for selecting training conditions, but were not necessarily superior to simply asking students which topic they preferred. Given the value of teacher-reported dispositions and efficacy and the low-cost of survey administration, connecting teacher-reported dispositions to administrative data and student outcomes provide an opportunity to explore the prevalence of Boykin and Noguera's framework across TPPs. Specifically, comparisons across certification condition and race-based subgroups would provide direction for future research on enhancing outcomes for underrepresented groups. Accordingly, this study addresses this gap.

Much of the existing large-scale TPP research makes common assumptions or uses common methodological approaches, especially as it relates to race-based outcomes. Yet, more than 20 years ago, Cochran-Smith (1995) called on teachers and teacher educators to re-consider assumptions that are made about race and other groups in schools. Many large-scale researchers have advocated for models that use statistical controls that essentially eliminate relationships between student demographics and student growth (Ehlert, Koedel, Parsons, & Podgursky, 2014). The idea is to use statistical controls to "level the playing field" to draw conclusions about effectiveness, regardless of student demographics. Statistical controls are often operationalized using variables such as percentages of racial groups in given samples.

Statistical controls often enhance estimation, but rarely lead to specific recommendations, nuanced insight, or emergent better practices for racial/ethnic groups. Instead conclusions are often described in terms of average or overall student effects. Moreover, researchers have asserted that simply using race/ethnicity as control variables implies inherent low expectations (K. A. Anderson, 2015, 2016; Brown, 2011; Timmermans, Doolard, & de Wolf, 2011). In some cases, data limitations may prevent researchers from disentangling outcomes by race. Thus, I am not wholly advocating for the abandonment of the use of statistical controls for demographics, but I am advocating for an expansion of methodological approaches used in large-scale TPP research to initiate policy and practice that is accountable for demographics, especially race.

Another common practice in large-scale TPP research is to assume that relationships between TPPs and student outcomes are linear. A challenge with assuming linearity is that such models may be misspecified in practice (Lin, Li, & Sun, 2014), possibly leading to biased effect estimates. However, additional approaches that relax parametric assumptions, such as the use of propensity score methods, (Price, Spriggs, & Swinton, 2011) may be useful in exploring large-scale TPP research. Accordingly, this study appeals to Raudenbush's (2015) appeal for "a wide range of related research and a coherent theory of action" (p. 141) by providing a fresh approach to large-scale TPP research. Specifically, I investigate outcomes using a range of methodologies and statistical assumptions, allowing readers to make informed judgments and assess consistency of results about TPPs and race-based outcomes across conditions.

The literature review and limitations of existing large-scale TPPs were used to guide this study. Moreover, components of Boykin and Noguera's (2011) framework were used to assist in selecting variables and examining relationships across certification conditions. As the goal of the study is to examine relationships between teacher certification, teacher traits, and student achievement, with a specific interest in examining differential prediction for underrepresented racial groups, the following research questions were developed.

> **Research Question 1:** What are the differences in mathematics outcomes for students who were taught by novice teachers who held standard certificates compared with students who were taught by novice teachers who held nonstandard certificates?
>
> **Research Question 2**: Does certification status of novice teachers differentially predict mathematics outcomes for underrepresented racial groups versus represented racial groups?
>
> **Research Question 3:** What are the differences in teacher-reported instructional practices and beliefs for novice teachers who held standard certificates compared with novice teachers who held nonstandard certificates.
>
> **Research Question 4**: Does certification status of novice teachers differentially predict teacher-reported instructional practices and beliefs for underrepresented racial groups versus represented racial groups?

Method

Sample

Data from the High School Longitudinal Study of 2009 (HSLS:09; Ingels, Dalton, Holder, Lauff, & Burns, 2011) were used in this study. The HSLS:09 research team collected data from a nationally representative sample of high school students to assess high school, postsecondary education, career, and adulthood transitions over several years. One of the strengths of the HSLS:09 database is that administrator, student, and teacher survey data were collected and can be matched to individual student achievement. Novice teachers who had 5 or fewer years of experience were the teachers of interest in this study. A total of 2,599 teachers were retained for this study. Of these teachers, roughly 15% of the teachers in the analytic sample held nonstandard certifications (n = 383 vs. n = 2,216).

Variable Descriptions

Certification status. Students were excluded if their corresponding teachers did not provide information about their certification status. Certification status was classified into two groups, standard or nonstandard. Standard certification (coded as 1) consisted of teachers who were granted regular or advanced state certificates or were granted certificates, but were on some initial probationary period for novice teachers. Nonstandard certification (coded as 0) consisted of teachers who were issued certificates that required additional coursework, tests, or completion of a certification program, and teachers who held no certificate.

Achievement. Data from the ninth-grade year (base year; 2009-2010) and the eleventh-grade year (first follow-up; 2011-2012) were used in this study to examine differences in sustained mathematics knowledge of students who were taught by novice teachers who held standard versus nonstandard certificates. Sustained mathematics knowledge was assessed by examining standardized algebra scores (M = 50, SD = 10) 2 years after having taken Algebra I in ninth grade. Given the design of HSLS:09, students were given the same algebra test in ninth and eleventh grades, thus, the initial test (limited to students who took the algebra exam at the beginning of the algebra course) was used as a pretest.

Race. Underrepresented racial[2] groups or students who were, American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, were included as control variables to assess overall effects of TPP on student outcomes. These students were compared with Asian, More than One Race, and White students.

Guiding functions. To control for variables that are especially critical in reducing bias (see Steiner et al., 2010), mathematics efficacy, interests, identity, and utility were included as covariates. To account for prior experiences before and after ninth-grade Algebra I, the guiding functions variables were included twice, once as a pretest in ninth grade and again in eleventh grade. These variables were created by the HSLS:09 research team (Ingels et al., 2011) using principal components analysis of student survey data. The results were standardized for the entire sample using a mean of zero and standard deviation of one.

Dispositions. Differences in teacher guiding functions and dispositions about instructional practices by preparation pathway were assessed to provide better guidance for practice. These dispositions were collected through self-reports of survey items ranking a teachers' emphasis on a Likert-type scale ranging from no emphasis to heavy emphasis. Several dispositions that aligned to Boykin and Noguera's framework (shown parenthetically) were retained for this study. Specifically, increasing mathematics interests (guiding functions), standardized test preparation (asset: information processing quality), business or industry applications of mathematics (meaningful learning), speedy and accuracy of computations (asset: information processing quality), teaching mathematics concepts (student engagement), developing computational skills (asset: information processing quality), developing problem-solving skills (cognitive engagement), mathematical reasoning (cognitive engagement), preparation for future mathematics study (intersubjectivity), logical structure of mathematics (cognitive engagement), history and nature of mathematics (culture/intersubjectivity), or effectively explaining mathematics ideas (student engagement) aligned to the framework.

Pre- and postmath experiences. To make equivalent comparisons, students were included only if they were enrolled in Algebra I in ninth grade during the base year. To control for variations in exposure to mathematics content, the rigor of mathematics courses taken in eighth and eleventh grades were also included. Measures of prior achievement included standardized algebra pretest scores and letter grades (A-D and below D) earned in the eighth-grade math course.

Demographics. A continuous measure of socioeconomic status (SES) and a categorical gender variable provided by the HSLS:09 team (Ingels et al., 2011) were included as additional covariates.

Analysis

A series of six regression models that contain a variety of approaches (linear, nonlinear, and nonparametric) were compared to address some of the gaps in the literature and assess the consistency of teacher certification estimates. Specifically, variants of the following models were assessed:

$$\mathrm{Y}_i = \beta_0 + \beta_1 Certified_i + \beta_2 URG_i + \beta_3 (Certified_i \times URG_i) + \mathbf{Z}_i' \beta + \epsilon_i$$

and

$$D_i = \beta_0 + \beta_1 Certified_i + \beta_2 URG_i + \beta_3 (Certified_i \times URG_i) + \epsilon_i$$

where Y_i represents mathematics achievement, β_0 represents an intercept term, $\beta_1 Certified_i$ represents estimates of certification status, $\beta_2 URG_i$ represent estimates of underrepresented racial group status (American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander) or not, $\beta_3 (Certified_i \times URG_i)$ represents the interaction between certification status and underrepresented racial group status, $\mathbf{Z}_i' \beta$ represents a vector of student-level covariates, and ϵ_i represents a residual term. A similar model is repeated for teacher dispositions to assess whether teacher dispositions between certification condition differed relative to underrepresented racial groups, where D_i represents teacher-reported dispositions regarding teaching practices. In the second case, only interactions between certification condition and underrepresented racial group status are of interest. Thus, the vector of covariates is excluded.

Model 1 generated population estimates by including all covariates using the balanced repeated replication recommended by the HSLS:09 team (Ingels et al., 2011). Models 2 through 6, employed the marginal mean weighting through stratification (MMW-S) procedure to achieve balance between the TPP conditions using the pretreatment covariates described in the previous section. However, various forms of the model were analyzed. Model 2 employed a nonparametric approach whereby balance was obtained using pretreatment covariates, but no additional covariates were modeled. Model 3 included scores of guiding functions covariates to capture changes in guiding functions that may have occurred after the algebra course. In addition, the math pretest covariate was also included as covariate in the model to increase robustness. Model 4 employed a different approach to capture guiding functions by including gain scores for the guiding functions variables between ninth and eleventh grade to account for changes that may have occurred after completing the algebra course. Model 5 included all covariates (similar to the population estimates in Model 1), which included two measures of guiding functions, once in ninth grade and once in eleventh grade. Finally, Model 6 included polynomials (quadratic terms) of all continuous covariates in attempt to model possible violations of linearity assumptions.

Because nonexperimental data were used in this study, the MMW-S method was the primary method of interest in this study. MMW-S is an innovative nonparametric strategy that incorporates strengths of propensity score stratification (PSS) and inverse-probability-of-treatment weighting (IPTW), but overcomes limitations of PSS and IPTW (Hong, 2010, 2012). Specifically, MMW-S requires (a) estimation of a propensity score based on pretreatment covariates to stratify a sample and (b) computation of a weight for each treated unit based on its representation within the associated stratum (Hong, 2012). Accordingly, the MMW-S weighting procedure equates the composition of pretreatment variables across all treatment groups, simulating a randomized experiment (Hong, 2012). MMW-S is represented in mathematical form by the following equation:

$$MMW\text{-}S = \frac{n_s \times pr\left\{Z = z\right\}}{n_z, s}$$

where n_s is the number of units in stratum s, $pr\left\{Z = z\right\}$ is the proportion of the entire sample to the treatment group z, and n_z, s is the actual number of units assigned to treatment z in stratum s. The numerator represents the expected number of units assigned to treatment z in stratum s in a completely randomized experiment; whereas the denominator represents a nonparametric way to compute the propensity score.

For binary treatments, a six-step analytic process is required to employ the MMW-S procedure. The steps include (a) estimating a propensity score based on collected pretreatment covariates, (b) assuring that each unit has a counterpoint in the treatment and control groups (excluding if false), (c) stratifying the sample based on the propensity score (5-6 strata usually reduces 90% of the variance), (d) computing the MMW for each unit in each group, (e) checking to ensure that the weighted sample is balanced across groups, and (f) estimating treatment effects (Hong, 2010, 2012). Treatment effects of MMW-S are aligned with Rubin's causal model to assess differences between two population average potential outcomes (Rubin, 1978). Appendices A and B describe additional information that informed the MMW-S procedures that were used to balance the data. For within-study comparison purposes, longitudinal analytic weights and 200 balanced repeated replication weights provided by the HSLS:09 team were used to generate population statistics and appropriate variance estimates that account for the multistage sampling design. These population estimates were used to assess similarities and differences between sample estimates generated using the MMW-S procedures. Pretreatment variables for the MMW-S procedures included guiding functions variables at early enrollment in ninth grade Algebra I, math pretest scores, eighth-grade math rigor, eighth-grade math letter grade (A-D and below D), whether the student was in the underrepresented racial group category, SES, and gender.

Results

Results show that the main effects for the standard certification condition were significant across all six models. All estimates are shared in Tables 1 to 3. Unstandardized regression estimates ranged from +1.25 to +1.73 points on the algebra examination, which is normed with a mean of 50 and standard deviation of 10. Thus, the model predicts that, holding all other factors constant in the model, students who were taught algebra by certified novice teachers in ninth grade would earn higher scores of approximately 1.25 points on the algebra test in eleventh grade, relative to students who were taught algebra by teachers who held nonstandard certification in ninth grade. Using the most conservative estimate, 1.25 points is roughly .13 of the normed standard deviation and .15 of the current sample standard deviation (M = 48.75, SD = 8.29). Based on published effect-size benchmarks for nationally normed tests, the conservative estimate of .13 is approximately equivalent to 8 months of expected growth from tenth to eleventh grade (Hill, Bloom, Black, & Lipsey, 2008). When examining interactions between underrepresented racial groups and certification condition, there were no significant differences noted in any of the models (Tables 1-3).

In terms of dispositions, there were several differences in teacher-reported instructional practices between teachers in the standard versus nonstandard certification condition. Specifically, teachers in the standard certification condition (coded as 1) reported higher use of algorithms $b = 0.21$, $t(1{,}966) = 4.56$, $p < .01$ and connecting mathematics ideas, $b = 0.10$, $t(1{,}961) = 2.38$, $p < .05$. Contrarily, teachers in the nonstandard certification condition (coded as 0) reported higher emphasis on increasing mathematics interests, $b = -0.17$, $t(1{,}967) = -3.41$, $p < .01$, standardized test preparation, $b = -0.24$, $t(1{,}958) = -4.49$, $p < .01$, business or industry applications of mathematics, $b = -0.13$, $t(1{,}960) = -2.12$, $p < .05$, and speedy and accuracy of computations, $b = -0.17$, $t(1{,}962) = -3.09$, $p < .01$. There were no differences between the teachers regarding emphasis on teaching mathematics concepts, developing computational skills, developing problem-solving skills, mathematical reasoning, preparation for future mathematics study, logical structure of mathematics, history and nature of mathematics, or effectively explaining mathematics ideas. Finally, there was only one significant interaction for dispositions, whereby teachers in the standard certification condition reported focusing less on increasing mathematics interests for underrepresented students, $b = -0.23$, $t(1{,}967) = -2.47$, $p < .05$.

Discussion

The conceptual framework used in the study, known as the research-based scheme for promoting enhanced classroom

Table 1. Multiple Linear Regression Predicting Differences in Teacher Traits of Teachers With Standard Versus Nonstandard Certification (Population Estimates = 3,84,266; Sample n = 3,104).

	Model 1		
	B	BRR *SE B*	*t*
Standard certification	1.73	0.84	2.08*
URG	−0.09	1.65	−0.05
Std. Cert. × URG	−1.19	1.85	−0.64
Guiding functions			
11th grade math efficacy	0.48	0.30	1.60
11th grade math interest	0.31	0.35	0.88
11th grade math utility	0.04	0.34	0.11
11th grade math identity	0.87	0.39	2.23*
9th grade math efficacy	0.52	0.29	1.76
9th grade math interest	−0.60	0.32	−1.88
9th grade math utility	−0.15	0.29	−0.52
9th grade math identity	−0.13	0.36	−0.37
Prior math experiences			
Math pretest	0.52	0.04	14.79**
8th grade math letter grade			
B	−1.19	0.58	−2.04*
C	−1.68	0.72	−2.31*
D	−0.90	1.73	−0.52
Below D	−2.54	1.45	−1.75
8th grade math rigor			
Advanced/honors math 8	3.04	1.33	2.28*
Algebra or higher	0.36	0.83	0.43
11th grade math rigor	−0.10	0.36	−0.28
Demographics			
SES	1.38	0.45	3.07**
Male	0.19	0.52	0.37
Constant	22.82	1.90	11.98*

Note. R^2 = .45. BRR *SE* = balance repeated replication standard errors; standard certification = teacher has standard certification (1) vs. teacher has not met all requirements for standard certification (0); underrepresented racial groups (URGs) = American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander students (1) vs. Asian, more than one race, and White (0); Std. Cert. × URG = interaction between standard certification and underrepresented racial group; 8th grade math letter grade = reference group is letter grade of "A"; math pretest = standardized algebra assessment taken early upon enrollment in 9th grade algebra; 8th grade math rigor (reference group is math 8 and pre-algebra); SES = socioeconomic status; 11th grade math rigor = mathematics course taken in 11th grade.
*p < .05. **p < .01

teaching and learning (Boykin & Noguera, 2011), emphasized three major domains: classroom-based asset focused factors, guiding functions, and student engagement. Findings in this study are aligned to this framework in that these traits were either related to student outcomes or teaching practices. Teachers from both pathways reported using various asset-focused strategies, albeit at differing amounts of emphasis, such as enhancing computational skills or making mathematics connections to business and industry. Moreover, guiding functions, especially mathematics efficacy and identity were strongly related to student achievement. As demonstrated by Steiner et al. (2010), additional covariates, such as mathematics pretests and prior mathematics experiences were strongly related to student outcomes as well. When accounting for these factors, students who were taught by teachers in the standard certification condition generally earned higher achievement scores than those who were taught by teachers from the nonstandard condition. However, achievement for underrepresented racial groups was underwhelming for students who were taught by teachers in both certification conditions.

Figure 1 provides a graphical representation of Model 6, the model that used a nonlinear approach and explained the most variation (R^2 = .49) out of the six models. Figure 1 is presented to describe some of the nuances associated with the interaction term between teachers who held standard certificates versus teachers who held nonstandard certificates, relative to mathematics achievement. Although statistically insignificant, Figure 1 reveals some interesting nuances. First, as shown by the left side of the graph, clear distinctions between teacher pathways are evident between students who were not included in the underrepresented racial group (Asian, More than one Race, and White) as compared with the students who were included in the underrepresented racial group (right side of graph). Specifically, students who were not included in the underrepresented racial category achieved consistently higher results on the mathematics achievement measure. The standard error bars are also narrower in the certified condition, indicating more consistent outcomes.

The standard error bars shown in Figure 1 relative to underrepresented racial group status (American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander) are wider, indicating less consistent outcomes. This may be related to concepts noted in the literature that teachers of underrepresented students may serve as advocates or resisters (Hill-Jackson & Lewis, 2010). For example, teachers who held standard certificates reported focusing less on increasing interests for underrepresented racial groups relative to students who were not included in the underrepresented category. This finding begs the question of why? These findings could highlight systemic bias that has permeated teacher training programs (Feagin, 2000) or may denote inappropriate dispositions noted by Milner and Laughter (2015). Thus, future research that investigates dispositional differences between teachers, connected to student outcomes, may provide needed insight for better schooling experiences for all students, especially for underrepresented racial groups.

Teachers with nonstandard certification indicated that they emphasized several key dispositions above the level of emphasis used by teachers with standard certification. When examining the relationships between teacher dispositions and underrepresented groups, teachers with nonstandard certification indicated that the emphasized increasing interests more so than the teachers with standard certification. However,

Table 2. Multiple Linear Regression Predicting Differences in Student Outcomes Between Teachers With Standard Versus Nonstandard Certification.

	Model 2			Model 3			Model 4			Model 5		
	B	*SE*	β	*B*	*SE*	β	*B*	*SE*	β	*B*	*SE*	β
Standard certification	1.25	0.59	0.05*	1.39	0.52	0.06**	1.40	0.53	0.06**	1.53	0.51	0.06**
URG	−2.19	1.01	−0.12*	0.15	0.99	0.01	1.31	1.00	0.07	0.87	0.98	0.05
Std. Cert. × URG	0.11	1.09	<0.00	−0.94	1.05	−0.05	−1.47	1.06	−0.08	−1.21	1.03	−0.07
Guiding functions												
11th grade math efficacy				0.59	0.20	0.07**				0.50	0.20	0.06*
11th grade math interest				0.01	0.20	<0.01				0.07	0.20	0.01
11th grade math utility				0.15	0.17	0.02				0.25	0.17	0.03
11th grade math identity				1.11	0.21	0.14**				1.00	0.23	0.13**
9th grade math efficacy										0.19	0.21	0.23
9th grade math interest										−0.37	0.20	−0.05
9th grade math utility										−0.10	0.18	−0.01
9th grade math identity										0.05	0.22	0.01
Efficacy gain scores							0.04	0.16	0.01			
Interest gain scores							0.31	0.16	0.04*			
Math utility gain scores							0.29	0.14	0.04*			
Math identity gain scores							0.39	0.19	0.04*			
Prior math experiences												
Math pretest				0.58	0.02	0.57**	0.62	0.02	0.61**	0.53	0.02	0.52**
8th grade math letter grade												
B										−1.35	0.37	−0.08**
C										−1.64	0.49	−0.08**
D										−2.10	0.79	−0.05**
Below D										−2.41	1.03	−0.05*
8th grade math rigor												
Advanced/honors math 8										1.44	1.05	0.03
Algebra or higher										0.83	0.40	0.04*
11th grade math rigor				0.10	0.20	0.01	0.24	0.20	0.02	0.03	0.20	<0.01
Demographics												
SES							1.23	0.23	0.11**	1.10	0.23	0.09**
Male							−0.26	0.30	−0.02	−0.40	0.30	−0.02
Constant	48.29	0.54**	—	19.99	1.12**	—	17.25	1.10**	—	22.68	1.20**	—
$*p < .05$ $**p < .01$	$R^2 = .02$; $n = 2{,}243$			$R^2 = .45$; $n = 1{,}608$			$R^2 = .43$; $n = 1{,}608$			$R^2 = .47$; $n = 1{,}608$		

Note. β = standardized beta coefficients; standard certification = teacher has standard certification (1) vs. teacher has not met all requirements for standard certification (0); underrepresented racial group (URG) = American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander (1) vs. Asian, more than one race, and White (0); Std. Cert. × URG = interaction between standard certification and URG; gain scores = difference from 9th to 11th grade; 8th grade math letter grade = reference group is letter grade of "A"; math pretest = standardized algebra assessment taken early upon enrollment in 9th grade algebra; 8th grade math rigor (reference group is math 8 and pre-algebra); SES = socioeconomic status; 11th grade math rigor = mathematics course taken in 11th grade.

student outcomes were not higher for students taught by teachers with nonstandard certification. Thus, it is not clear if teachers were indeed emphasizing the instructional practices with greater emphasis, but with reduced effectiveness, or if the teachers' perspectives regarding emphasis on the various instructional practices were not representative of what occurred. Thus, additional dispositional studies for teachers with nonstandard certificates are also necessary.

Although achievement results were lower and less consistent for underrepresented racial groups, Figure 1 highlights a key nuance for teachers with nonstandard certificates. As shown by the dashed error bars on the right-hand side of the graph, student achievement is wildly inconsistent for underrepresented students who were taught by teachers with nonstandard certification. These large errors may indicate that some nonstandard teachers are highly successful with underrepresented racial groups, even more so than teachers with standard certification. In addition, nonstandard programs may attract more diverse candidates, which may also enhance guiding functions, such as mathematics identity. Conversely, the bottom portion of the dashed error bar may also indicate that teachers with nonstandard certificates interject much

Table 3. Multiple Linear Regression Predicting Differences in Teacher Traits of Teachers With Standard Versus Nonstandard Certification.

	Model 6		
	B	*SE*	β
Standard certification	1.54	0.51	.06**
URG	0.91	0.98	.05
Std. Cert. × URG	−1.23	1.03	−.07
Guiding functions			
11th grade math efficacy	0.52	0.20	.07**
11th grade math efficacy^2	0.09	0.12	.02
11th grade math interest	−0.01	0.20	<.00
11th grade math interest^2	0.05	0.14	.01
11th grade math utility	−0.05	0.18	−.01
11th grade math utility^2	−0.36	0.10	−.08**
11th grade math identity	1.11	0.23	.14**
11th grade math identity^2	0.18	0.16	.02
9th grade math efficacy	0.14	0.21	.02
9th grade math efficacy^2	0.08	0.12	.01
9th grade math interest	−0.35	0.21	−.04
9th grade math interest^2	−0.13	0.13	−.02
9th grade math utility	−0.07	0.19	−.01
9th grade math utility^2	−0.03	0.11	−.01
9th grade math identity	0.04	0.22	.01
9th grade math identity^2	−0.06	0.15	−.01
Prior math experiences			
Math pretest	0.51	0.02	.51**
Math pretest^2	<0.01	<0.01	.03
8th grade math letter grade			
B	−1.24	0.37	−.08**
C	−1.54	0.49	−.07**
D	−1.81	0.79	−.05*
Below D	−2.43	1.04	−.05*
8th grade math rigor			
Advanced/honors math 8	1.47	1.05	.03
Algebra or higher	0.80	0.39	.04*
11th grade math rigor	−3.58	0.81	−.33**
11th grade math rigor^2	1.93	0.43	.34**
Demographics			
SES	0.67	0.39	.06
SES^2	0.28	0.23	.04
Male	−0.28	0.30	−.02
Constant	48.49	0.64**	—

Note. β = standardized beta coefficients; standard certification = teacher has standard certification (1) vs. teacher has not met all requirements for standard certification (0); underrepresented racial group (URG) = American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander Students (1) vs. Asian, more than one race, and White (0); Std. Cert. × URG = interaction between standard certification and URG;^{2} = quadratic term; 8th grade math letter grade = reference group is letter grade of "A"; math pretest = standardized algebra assessment taken early upon enrollment in 9th grade algebra; 8th grade math rigor (reference group is math 8 and pre-algebra); SES = socioeconomic status; 11th grade math rigor = mathematics course taken in 11th grade.
R^2 = .49; n = 1,608.
*$p < .05$. **$p < .01$.

more risk for underrepresented racial groups (and all students), considering the depressed achievement levels. It is not likely that one approach to TPPs will prevail in the United States. However, no matter the entry pathway, a healthy balance between minimizing risk and maximizing outcomes for students are essential for effective teacher education policy.

Conclusion

TPPs have historically been essential components of maintaining inequity in schools (J. D. Anderson, 1998; Milam, 2010). This study supports this historical phenomenon. In this study, regardless of certification status, results are underwhelming for underrepresented racial groups. These findings highlight a need to redefine notions of teacher effectiveness and rethink teacher preparation for underrepresented racial groups. Additional research that examines variables beyond the "usual suspects" and considers the salience of race in education and teacher preparation should be considered in the future. When discussing race and schooling in the 1930s, W. E. B. Du Bois (1935) indicated that a key factor for the proper induction of students into life required that teachers develop compassionate relationships with students, based on the premise of social equality, by obtaining knowledge of individual students and understanding the students' surroundings and collective histories. Du Bois's assertion for teacher preparation is still relevant today. However, large-scale research does not adequately address the collective histories of students who are members of underrepresented racial groups.

In this study, I used a within-study comparison approach, employing a propensity score framework to generate sample estimates and a replication approach to generate national population estimates. These approaches diversify the methodology commonly used in large-scale TPP research, allowing for relaxed parametric assumptions and accounting for nonrandomized assignment to teacher certification conditions. One of the critiques of using national data sets for assessing the relationships between student test scores and teachers is that tests are often not aligned with the subjects in which the teachers teach (Clotfelter, Ladd, & Vigdor, 2010). I address this gap in the literature by limiting the analysis to students who (a) were taught Algebra I in ninth grade by novice teachers with known certification statuses and (b) completed a pre- and post-algebra assessment.

In addition, I addressed gaps in the literature by assessing student outcomes and teacher dispositions. These outcomes were interpreted by accounting for the historical significance of race in education and teacher preparation. I designate this type of research as race-conscious research, or research that attempts to explain social phenomena or improve outcomes, practices, or policies for a specific racial group or groups by interrogating assumptions, accounting for current or historical experiences of specific racial groups, or examining practices and policies that result in differential implementation or outcomes for racial

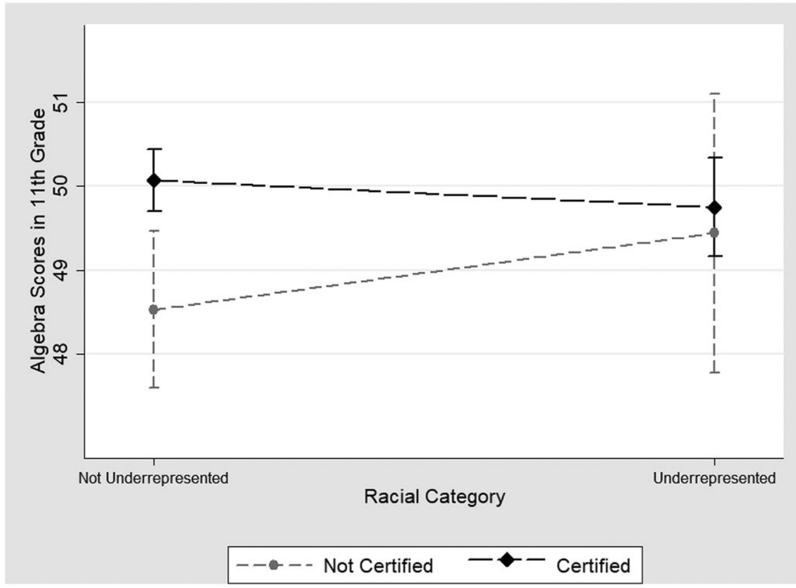

Figure 1. Plot of interactions between racial groups and certification status of algebra scores 2 years after taking Algebra I (95% CIs).
Note. Interactions were not statistically significant, but reflect key nuances described in the narrative. CI = confidence interval.

groups. Where appropriate, custodians of knowledge (e.g., editors, reviewers, and publishers) should consider encouraging large-scale TPP researchers to interrogate race-based outcomes, teacher dispositions, and practices as well as overall or average outcomes, rather than just solely focusing on the latter.

Large-scale, race-conscious teacher preparation research that is connected to an array of student outcomes is sorely needed and this study is one example. Race-conscious, large-scale teacher preparation research has the potential to inform ideological debates about how teachers should be prepared and may lead to broader, more critical empirical perspectives for setting teacher preparation policy, especially for underrepresented groups. This study demonstrates that effective teachers and their associated dispositions matter, but there is much room for improvement. Future research should examine long-term trajectories of underrepresented racial groups based on the certification tracks and dispositions of teachers in critical subjects. In sum, effective teachers and teacher dispositions are linked to student outcomes; nonetheless effectiveness is nuanced and requires structural as well teacher preparation reform to enhance outcomes for underrepresented racial groups.

Appendix A

Summary of Pairwise Correlations.

	Math score	Certification status	Math pretest	8th grade rigor	8th grade marks	URG	Male	SES	Math efficacy	Math interests	Math utility	Math identity
Math score	—											
Certification status	.09*	—										
Math pretest	.64*	−.05*	—									
8th grade rigor	−.06*	−.02	.08*	—								
8th grade marks	−.33*	−.02	−.34*	.06*	—							
URG	−.11*	−.04*	−.18*	<.01	.09*	—						
Male	−.04*	.03	−.01	.02	.10*	.01	—					
SES	.30*	.07*	.32*	.05*	−.21*	−.27*	.01	—				
Math efficacy	−.27*	.01	.29*	.08*	−.35*	.01	.07*	.13*	—			
Math interests	.15*	−.02	.17*	<.01	−.26*	.09*	−.04*	.04*	.53*	—		
Math utility	−.07*	.01	.08*	−.02	−.11*	.09*	.00	−.01	.37*	.46*	—	
Math identity	−.31*	−.00	.33*	.05*	−.39*	.00	.04*	.08*	.61*	.54*	.34*	—

Note. Math score = standardized algebra assessment taken in 11th grade; certification status = teacher has standard certification (1) versus teacher has not met all requirements for standard certification (0); math pretest = standardized algebra assessment taken early upon enrollment in 9th-grade algebra; 8th grade rigor = rigor of mathematics course in 8th grade; underrepresented racial group (URG) = American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander students; SES = socioeconomic status.
*$p < .05$.

Appendix B

Computation of MMW-S for Mathematics Achievement.

		Unweighted sample		MMW-S		Weighted sample	
Stratum	*n*	Nonstandard certification	Standard certification	Nonstandard certification	Standard certification	Nonstandard certification	Standard certification
1	520	98	422	0.782	1.051	76.64	443.52
2	520	93	427	0.824	1.038	76.63	443.23
3	520	76	444	1.001	0.999	76.08	443.56
4	520	70	450	1.094	0.985	76.58	443.25
5	519	46	473	1.663	0.936	76.50	442.78
Total	2,599	383	2,216			382.43	2,216.34

Note. Slight differences in totals are due to rounding. All *F* tests of covariates were insignificant between teacher pathways after balancing. MMW-S = marginal mean weight through Stratification.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. For classification purposes, racial groups that are underrepresented in high achievement categories are referred to as underrepresented racial groups in the article. High achievers exist across all racial groups, but when disaggregating students by race, several racial groups have lower representation, in high achievement categories, relative to their overall population. The term "underrepresented" racial group sometimes bears a negative connotation. In this article, it simply refers to the numerical representation of these students in higher versus lower achieving categories relative to their overall representation in public schools.
2. The term race in this study focuses primarily on socially determined race categories that were designated by the United States Office of Management and Budget's Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity Census Bureau (Office of Management and Budget, 1997). This revision required a minimum of six categories for reporting race: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and White. This is one perspective of race, but race is socially determined and members who are categorized into specific categories may reject such designations.

References

Anderson, J. D. (1998). *The education of Blacks in the South, 1860-1935*. Chapel Hill: The University of North Carolina Press.

Anderson, K. A. (2014). Equity in opportunities to learn mathematics: Policy and practice implications for high-achieving black students. In Y. Sealey-Ruiz, C. W. Lewis, & I. Toldson (Eds.), *Teacher education and black communities: Implications for access, equity, and achievement* (pp. 253-272). Charlotte, NC: Information Age.

Anderson, K. A. (2015). An introduction to optimal resource theory: A framework for enhancing student achievement. *Journal of Negro Education*, *84*(1), 25-39.

Anderson, K. A. (2016). Examining organizational practices that predict persistence among high-achieving black males in high school. *Teachers College Record*, *118*(6), 1-26.

Andersson, C., Johansson, P., & Waldenstrom, N. (2011). Do you want your child to have a certified teacher? *Economics of Education Review*, *30*, 65-78.

Blackwell, L. A., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246-263.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, *2*(1), 176-216.

Boykin, A. W., Lilja, A. J., & Tyler, K. M. (2004). The influence of communal vs. individual learning context on the academic performance in social studies of grade 4-5 African Americans. *Learning Environments Research*, *7*(3), 227-244.

Boykin, A. W., & Noguera, P. (2011). *Creating the opportunity to learn: Moving from research to practice to close the achievement gap*. Alexandria, VA: Association for Supervision and Curriculum Development.

Brown, A. L. (2011). "Same old stories": The black male in social science and educational literature, 1930s to the present. *Teachers College Record*, *113*(9), 2047-2079.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778-820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in High School: A cross-subject analysis with fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Wheeler, J. (2010). High poverty schools and the distribution of principals and teachers. *North Carolina Law Review*, *85*(5), 1345-1379.

Cochran-Smith, M. (1995). Color blindness and basket making are not the answers: Confronting the dilemmas of race, culture, and language diversity in teacher education. *American Educational Research Journal*, *32*(3), 493-522.

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification* (Final Report, NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

D'alessio, M. (2018). The effect of microteaching on science teaching self-efficacy beliefs in preservice elementary teachers. *Journal of Science Teacher Education*, *29*(6), 441-467. doi:10.1080/1046560X.2018.1456883

DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study-kindergarten cohort. *Journal of Educational Psychology*, *99*(2), 369-379.

Du Bois, W. E. B. (1935). Does the Negro need separate schools? *The Journal of Negro Education*, *4*(3), 328-335.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2014). Selecting growth measures for use in school evaluation systems: Should proportionality matter? *Educational Policy*, *30*(3), 465-500. doi:10.1177/0895904814557593

Feagin, J. (2000). *Racist America: Roots, current realities, and future reparations*. New York, NY: Routledge.

Goldhaber, D., Liddle, S., & Theobold, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *34*, 29-44. doi:10.1016/j.econedurev.2013.01.011

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Hill-Jackson, V., & Lewis, C. W. (2010). Dispositions matter: Advancing habits of the mind for social justice. In V. Hill-Jackson & C. W. Lewis (Eds.), *Transforming teacher education: What went wrong with teacher training, and how we can fix it* (pp. 61-92). Sterling, VA: Stylus.

Hollins, E. R. (2011). Teacher preparation for quality teaching. *Journal of Teacher Education*, *62*(4), 395-407. doi:10.1177/0022487111409415

Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *Journal of Educational and Behavioral Statistics*, *35*(5), 499-531. doi:10.3102/1076998609359785

Hong, G. (2012). Marginal mean weighting through stratification: A generalized method for evaluating multivalued and multiple treatments with nonexperimental data. *Psychological Methods*, *17*(1), 44-60. doi:10.1037/a0024918

Hurley, E. A., Boykin, A. W., & Allen, B. A. (2005). Communal versus individual learning of a math estimation task: African American children and the culture of learning contexts. *Journal of Psychology*, *139*(6), 513-527.

Hurley, E. A., Allen, B. A., & Boykin, A. W. (2009). Culture and the interaction of student ethnicity with reward structure in group learning. *Cognition and Instruction*, *27*(2), 121-146.

Ingels, S. J., Dalton, B., Holder, T. E., Lauff, E., & Burns, L. J. (2011). *High school longitudinal study of 2009 (HSLS:09): A first look at fall 2009 ninth-graders* (NCES No. 2011327). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011327

Lin, Z., Li, Q., & Sun, Y. (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics*, *178*(1), 167-179. doi:10.1016/j.jeconom.2013.08.014

Lubienski, S. T., Lubienski, C., & Crane, C. C. (2008). Achievement differences and school type: The role of school climate, teacher certification, and instruction. *American Journal of Education*, *115*(1), 97-138.

Milam, J. (2010). (Re) envisioning teacher education: A critical exploration of missed historical moments and promising possibilities. In V. Hill-Jackson & C. W. Lewis (Eds.), *Transforming teacher education: What went wrong with teacher training, and how we can fix it* (pp. 3-36). Sterling, VA: Stylus.

Milner, H. R. (2015).). *Rac(e)ing to class: Confronting poverty and race in schools and classrooms*. Cambridge, MA: Harvard Education Press.

Milner, H. R., & Laughter, J. C. (2015). But good intentions are not enough: Preparing teachers to center race and poverty. *Urban Review*, *47*, 341-363. doi:10.1007/s11256-014-0295-4

Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity*. Retrieved from https://obamawhitehouse.archives.gov/omb/fedreg_1997standards/

Price, G. N., Spriggs, W., & Swinton, O. H. (2011). Propensity score matching estimates from the National Survey of Black Americans. *Review of Black Political Economy*, *38*, 103-130. doi:10.1007/s12114-011-9088-0

Raudenbush, S. W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher*, *44*(2), 138-141. doi:10.3102/0013189X15575345

Reiman, A. J., & Johnson, L. E. (2007). Beginning teacher dispositions: Examining the moral/ethical domain. *Teaching and Teacher Education*, *23*, 676-687. doi:10.1016/j.tate.2006.12.006

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Education Evaluation and Policy Analysis*, *34*(1), 3-26. doi:10.3102/0162373711420865

Ronfeldt, M. (2015). Field placement schools and instructional effectiveness. *Journal of Teacher Education*, *66*(4), 304-320. doi:10.1177/0022487115592463

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching? *Teaching and Teacher Education*, *28*(8), 1091-1106.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, *6*, 34-58. doi:10.1214/aos/1176344064

Sleeter, C. E. (2014). Toward teacher education research that informs policy. *Educational Researcher*, *43*(3), 146-153. doi:10.3102/0013189X14528752

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*(3), 250-267. doi:10.1037/a0018719

Timmermans, A. C., Doolard, S., & de Wolf, I. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement*, *4*(22), 393-413. doi:10.1080/09243453.2011.590704

Wiseman, A. W., & Al-baker, F. (2013). The elusiveness of teacher quality: A comparative analysis of teacher certification and student achievement in Gulf Cooperation Council (GCC) countries. *Prospects*, *43*, 289-309. doi:10.1007/s11125-013-9272-z

Author Biography

Kenneth Alonzo Anderson, a former middle school teacher, is associate dean for Research and Sponsored Programs at Howard University. His primary research areas include education policy, curriculum and teacher effectiveness, and computational literacy. His publications have appeared in journals such as the *Journal of Negro Education, Teachers College Record, Urban Review*, and *Teaching and Teacher Education*.