

# When Do Students in Low-SES Schools Perform Better-Than-Expected on a High-Stakes Test? Analyzing School, Teacher, Teaching, and Professional Development Characteristics

Urban Education

1–35

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0042085916668953

ue.sagepub.com



**Christian Fischer<sup>1</sup>, Barry Fishman<sup>1</sup>,  
Abigail Jurist Levy<sup>2</sup>, Arthur Eisenkraft<sup>3</sup>,  
Christopher Dede<sup>4</sup>, Frances Lawrenz<sup>5</sup>,  
Yueming Jia<sup>2</sup>, Janna Fuccillo Kook<sup>2</sup>,  
Kim Frumin<sup>4</sup>, and Ayana McCoy<sup>3</sup>**

## **Abstract**

This empirical study analyzed data from 638 teachers and 11,800 students in low-socioeconomic status (SES) urban schools (and schools with urban characteristics) exploring associations of school, teacher, teaching, and professional development characteristics toward student performance on the revised Advanced Placement (AP) Biology and AP Chemistry examinations. The analyses indicated that districts per-student funding allocations, the

---

<sup>1</sup>University of Michigan, Ann Arbor, USA

<sup>2</sup>Education Development Center, Inc., Waltham, MA, USA

<sup>3</sup>University of Massachusetts Boston, USA

<sup>4</sup>Harvard University, Cambridge, MA, USA

<sup>5</sup>University of Minnesota, Minneapolis, USA

## **Corresponding Author:**

Christian Fischer, School of Education, University of Michigan, 610 East University Avenue, Suite 2211, Ann Arbor, MI 48109-1259, USA.

Email: [chrisfi@umich.edu](mailto:chrisfi@umich.edu)

days of instruction, teachers' knowledge and experience, and some aspects of teachers' professional development participation were significantly associated with student performance on AP science examinations that was better than predicted by students' Preliminary Scholastic Aptitude Test (PSAT) scores.

**Keywords**

science education, high-stakes testing, school context, professional development

As we strive for increased educational equity, a focus on narrowing achievement and opportunity gaps is important (Darling-Hammond, 2010). This opportunity gap is especially problematic for students in urban and high-poverty schools (Milner, 2012a; Tate, 2008). A recurring theme in urban education research is the aspiration of providing all students with equitable opportunities to succeed. This often involves investigations of how to increase access for disadvantaged students to high-quality learning opportunities, attempts to identify factors that enhance student achievement and college enrollment rates, and explorations of the far-reaching influences of students' socioeconomic status (SES) on achievement and outcomes (e.g., Achinstein, Curry, Ogawa, & Athanases, 2016; Archer-Banks & Behar-Horenstein, 2012; Burks & Hochbein, 2015; Cilesiz & Drotos, 2016; Hébert & Reis, 1999; Thompson, 2004; Ward, 2006).

At the high school level, the College Board's Advanced Placement (AP) programs in the sciences and other subject areas are viewed as high-quality opportunities for students to engage in rigorous learning experiences. Research indicates that participation in AP courses and success in AP examinations are associated with greater academic success in higher education, such as higher enrollment rates in 4-year colleges (Chajewski, Mattern, & Shaw, 2011), higher college graduation rates (Dougherty, Mellor, & Jian, 2006; Mattern, Marini, & Shaw, 2013), and higher college grade point averages (Hargrove, Godin, & Dodd, 2008; Patterson, Packman, & Kobrin, 2011; T. P. Scott, Tolson, & Lee, 2010). Historically, urban and economically disadvantaged students had less access to AP programs than their better-off peers (Schneider, 2009). Although extensive efforts to increase access for students in urban and high-poverty schools to AP programs have been undertaken (The College Board, 2014; Conger, Long, & Iatarola, 2009; Lichten, 2010; Roegman & Hatch, 2016; Wyatt & Mattern, 2011), tracking systems and the quantity of offerings are often barriers to enrollment in AP courses (Klopfenstein, 2004;

Klugman, 2013; Schneider, 2009; Zarate & Pachon, 2006). Nevertheless, simply increasing access to AP examinations does not increase the percentage of economically disadvantaged students passing AP examinations (Hallett & Venegas, 2011; Lichten, 2010). While AP participation of low-SES students increased from 11.4% ( $N=58,489$ ) in the class of 2003 to 27.5% ( $N=275,864$ ) in the class of 2013, only 21.7% of low-SES students in the class of 2013 scored a 3 or higher (passing grade), compared with 75.3% of non-low-SES students (The College Board, 2014). These performance discrepancies indicate that low-SES students are still less likely to obtain equitable learning opportunities despite the increased access to AP courses.

Milner's (2012b) classification of "urban" school settings emphasizes poverty, lack of resources, and high percentages of English-language learners. These are called "urban characteristics" (p. 559), and their presence may be related to student outcomes even if schools are geographically located outside of urban districts. Within urban districts, Milner (2012b) distinguishes between "urban intensive" and "urban emergent" schools based on city density. We employ Milner's (2012b) definitions in this study to explore the AP science performance of students in schools that are either urban or have urban characteristics. In these schools, where students might be expected to suffer from opportunity or achievement gaps, why do some students perform better-than-expected on the AP science examinations?

## Background

### *The AP Program*

The College Board's AP examinations and corresponding courses provide rigorous, college-level curricula for high school students in a broad variety of subjects. The summative nationwide high-stakes assessments are graded on a 1-5 scale using criterion-based rubrics. Students receiving a passing score (3 or higher) may be able to count their AP grade toward their college degree completion, depending on the policies of their institution of higher education.

The recent redesign of the AP science curriculum emerged from recommendations of the National Research Council suggesting de-emphasis of algorithmic-centered instruction and rote memorization (National Research Council, 2002). Responding to these recommendations, the College Board redesigned the AP science curriculum framework, increasing the emphasis on scientific practices, critical thinking, inquiry, and reasoning to deepen students' understanding of relevant science concepts (e.g., Magrogan, 2014; Yaron, 2014). The redesigned AP Biology examination was first administered in May 2013, followed by AP Chemistry in May 2014. Items focusing

on factual knowledge or purely algorithmic procedures were reduced on the redesigned AP science examinations to include more items accessing deeper conceptual understanding and higher-order cognitive skills (Domyancich, 2014; Magrogan, 2014). Many of these changes are in line with nationwide science standards described in the Framework for K-12 Science Education (National Research Council, 2012) and the Next Generation Science Standards (NGSS; NGSS Lead States, 2013).

These changes introduce new challenges for teachers who need to adopt to the new curricular frameworks and modify their science instruction. Therefore, teachers might be more inclined to participate in professional development (PD) activities due to the high-stakes nature of the AP examinations. Thus, this study provides a unique opportunity to explore how schools and teachers respond to this large-scale top-down mandated educational reform.

### *Achievement Gap Trends*

Integrating data from nationally representative studies, Reardon (2011, 2013) describes how the income achievement gap for students in the top and bottom 10th income percentile increased from the mid-1940s to the turn of the century by about 0.5 standard deviations. The influence of SES on student achievement is also documented in large-scale international comparative studies. For instance, the 2012 Program for International Student Assessment (PISA) study indicates that 15% of U.S. students' performance variation is attributable to students' SES (Organisation for Economic Co-Operation and Development [OECD], 2013a, 2013b). On the contrary, Reardon (2011, 2013) finds that achievement gaps due to race/ethnicity narrowed from the 1950s to the turn of the century with a decrease in the African American/White achievement gap of about 0.6 standard deviations, about 0.5 standard deviations smaller compared with the income achievement gap. Nevertheless, an SES-based effect on academic performance persists (e.g., Bohrnstedt, Kitmitto, Ogut, Sherman, & Chan, 2015; Milner, 2012c). For instance, the 2011 National Assessment of Educational Progress (NAEP) program evaluation ascertained that the African American/White achievement gap further decreases when controlling for SES (Bohrnstedt et al., 2015). Given that SES-based performance discrepancies on the AP examinations mirror general trends of widening income achievement gaps, this study exclusively focuses on low-SES urban schools (and schools with urban characteristics as defined by Milner, 2012b). Accounting for the intersection of race and class on student achievement, racial/ethnic background variables were included in the analyses as student-level covariates.

## Theoretical Framework

Hundreds of thousands of students and tens of thousands of AP science teachers are affected by the mandated, nationwide, top-down implementation of the revised AP science curricula and examinations. Although students and teachers share responsibility for student learning (Patrick, Mantzicopoulos, & Sears, 2010), teachers and teacher learning are instrumental for improving student learning and achievement (e.g., Ball & Cohen, 1999; Cohen & Ball, 1999; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Hattie, 2009). Thus, exploring urban students' performance on the redesigned AP science examinations is framed by an examination of how teachers navigate this change within their specific school contexts, and how schools support teachers in their AP science teaching.

This study employed a modified version of Opfer and Pedder's (2011) "Dynamic Model of Teacher Learning and Change." Employing a complexity theory perspective, Opfer and Pedder (2011) describe how three recursive and autopoietic subsystems, the school-level system, the individual teacher-level system, and the PD-level system affect teacher learning and changes in classroom practices. This study modified Opfer and Pedder's (2011) framework in three ways: First, emphases on specific elements within each subsystem are slightly shifted. For instance, instead of foregrounding collective norms, structures, and belief systems about learning on the school-level system, this study highlighted the availability/scarcity of resources, given the study's focus on low-SES school settings. Second, Opfer and Pedder (2011) emphasize the recurrence, interdependence, and overlap of elements within and across subsystems. Conceptually, this study concurs with these notions but the data sources with their underlying variable structures posed some challenges on modeling such relationships. Third, Opfer and Pedder (2011) limit their framework to teacher- and school-level elements. This study extended this approach by connecting teacher learning and classroom practices to student achievement in accordance with other conceptualizations of teacher learning (Borko, 2004; Darling-Hammond et al., 2009; Desimone, 2009).

### *The Challenges of Urban Contexts*

In addition to the demands of acclimating to the AP redesign, the context of low-SES urban schools (and schools with urban characteristics) poses additional challenges for students and teachers that might widen opportunity gaps. High-poverty schools might suffer from substantially lower district expenditures, poorly equipped classrooms, higher student-teacher ratios,

more out-of-field teaching, difficulties to recruit and retain highly qualified teachers, and infrequent implementations of effective teaching (Biddle & Berliner, 2003; Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff, 2011; Goldhaber, Lavery, & Theobald, 2015; Hill, Guin, & Celio, 2003; Ingersoll, 1999; Isenberg et al., 2013) which illustrates underlying conditions that contribute to existing opportunity gaps.

### *Teacher and Teaching Characteristics*

On the teacher level, individual teacher characteristics and the quality of instruction are widely regarded as important preconditions for students' success on the AP science examinations (Hallett & Venegas, 2011; Klopfenstein, 2004; Lichten, 2010). Although teachers' knowledge and expertise is related to teaching quality, science content knowledge alone is insufficient for high-quality science teaching (e.g., Abell, 2007; Magnusson, Krajcik, & Borko, 1999). To better describe the different knowledge domains necessary for high-quality instruction, Ball, Thames, and Phelps (2008) extend Shulman's (1986) triad of "subject matter content knowledge," "pedagogical content knowledge," and "curricular knowledge" with the more nuanced multidimensional "Content Knowledge for Teaching" framework. Ball et al. (2008) describe the six knowledge domains as "common content knowledge" ("knowledge and skill[s] used in settings other than teaching" [p. 399]), "specialized content knowledge" ("knowledge and skill[s] unique to teaching" [p. 400]), "horizon content knowledge" ("awareness of how [disciplinary] topics are related over the span of [the discipline] included in the curriculum" [p. 403]), "knowledge of content and students" ("knowledge that combines knowing about students and knowing about [the discipline]" [p. 401]), "knowledge of content and teaching" ("combines knowing about teaching and knowing about [the discipline]" [p. 401]), and "knowledge of content and curriculum" (which is identical to Shulman's [1986] "curricular knowledge"). The greater teachers' expertise in each of these domains, the more likely they are to engage in high-quality instruction using "high-leverage practices," which Ball and Forzani (2011) define as "those activities of teaching which are essential; . . . competent engagement in them would mean that teachers are well-equipped to develop other parts of their practice and become highly effective professionals" (p. 19). Examples of such high-leverage practices include "explaining and modeling content, practices, and strategies"; "diagnosing particular common patterns of student thinking and development in a subject matter domain"; and "setting up and managing small group work" (TeachingWorks, 2016).

## Teacher PD

Due to the high-stakes nature of the AP examinations and the major curriculum changes of the AP redesign, we believe that AP science teachers have a strong incentive for engaging in PD. The ultimate goal of PD is to increase student learning and achievement (Darling-Hammond et al., 2009; Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2010). An accepted theory of change asserts that teacher participation in “high-quality” PD results in increases in teacher’s knowledge and experience leading to instructional changes that eventually affect student learning and achievement (Desimone, 2009; Fishman, Marx, Best, & Tal, 2003; Fishman et al., 2013; Loucks-Horsley & Matsumoto, 1999). A decade of systematically conducted empirical research studies on best practices of PD activities (e.g., Fishman et al., 2003, 2013; Banilower, Heck, & Weiss, 2007; Borko, 2004; Garet, Porter, Desimone, Birman, & Yoon, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Roth et al., 2011) led to a consensus of core PD characteristics constituting “high-quality” PD—content focus, active learning, coherence, duration, and collective participation (Desimone, 2009). Content focus refers to PD that enhances teachers’ expertise in knowledge domains. For example, PD might provide examples of how to support students’ scientific inquiry processes during laboratory investigations. Active learning refers to PD that emphasizes teachers’ active engagement in thinking processes to self-construct knowledge. For example, PD might provide opportunities to review student work, observe expert teaching, or being observed during own classroom teaching. Coherence refers to PD that is aligned with existing curriculum frameworks, assessments, and school/district/state/nationwide reforms and policies, as well as with teachers’ prior PD experiences, instructional practices, knowledge, and beliefs. For instance, first-year teachers might participate in very different PD activities compared with veteran AP teachers. Duration refers to both the total contact time and the time span in which the PD takes place. For example, the total contact time and time span of College Board’s 4- to 5-day summer institutes are predefined, whereas participation in online teacher communities might vary greatly in both total time and time span. Collective participation refers to PD that is attended by multiple teachers from the same school, department, or grade facilitating collegial and supportive relationship building among colleagues. For example, teachers who collectively participate in the same PD activity might communicate about PD content after the official end of the PD activity, which might foster sustainable changes of classroom practices. Although prior research established these “high-quality” PD characteristics, systematic empirical explorations relating teachers’ exposure to each of the “high-quality” PD features toward student achievement are still needed.

## Research Questions

This study is framed by the following two research questions focusing on the identification of factors that might narrow opportunity gaps in urban school (and schools with urban characteristics):

**Research Question 1:** How do school, teacher, teaching, and PD participation characteristics compare across three AP science teacher subgroups; teachers whose students perform on average *lower-than-expected*, *as-expected*, and *better-than-expected*?

**Research Question 2:** What are associations between school, teacher, teaching, and PD participation characteristics on students' AP performance gains controlling for student demographics?

## Method

### Data Sources

This study is part of a larger longitudinal research project that explores how student outcomes in response to changes introduced by the AP redesign are related to teachers' PD patterns. The data used in this study were gathered from web-based surveys sent to AP Biology and AP Chemistry teachers in May 2014 inquiring about teacher demographics (e.g., age, gender), teaching background (e.g., teaching experience, university education), PD participation (e.g., "high-quality" PD features), general attitudes toward PD (e.g., perceived PD effectiveness, belonging to professional organizations), AP science course characteristics (e.g., length of instruction, number of students/sections/preps), AP science instruction and school context (e.g., teaching practices, administrative support), and concerns (e.g., challenges with the AP redesign). Prior to the first administration in 2013, the surveys were piloted with selected AP teachers and critiqued by an advisory board with expertise in science education, PD, and measurement. Survey items were validated using a cognitive interview methodology (Desimone & Le Floch, 2004).

The College Board provided student- and school-level data for all students taking AP science examinations, which included student demographics (e.g., racial/ethnic background, parental educational attainment, English-language learner status), students' PSAT and AP science scores, school characteristics (e.g., enrollment in free- and reduced-price lunch programs, school neighborhood), and district-level information (e.g., per-student funding allocations).



## Population and Sample

The overall student population consisted of all students taking the AP Biology ( $N_{\text{Bio,S}} = 203,304$ ) and AP Chemistry ( $N_{\text{Chem,S}} = 133,323$ ) examination in May 2014. Web-based surveys were sent to every AP Biology ( $N_{\text{Biol,T}} = 9,511$ ) and AP Chemistry ( $N_{\text{Chem,T}} = 7,098$ ) teacher in the nation, unless they were placed (by personal request) on College Board's Do Not Contact List. The survey was completed by 2,482 AP Biology (response rate = 26.10%) and 2,563 AP Chemistry (response rate = 36.11%) teachers, which are considered good response rates for web-based surveys with this population size (Shih & Fan, 2009). Non-response analyses using non-parametric Mann-Whitney tests indicated that survey responders taught slightly higher achieving students on PSAT (Biology:  $z = -9.35$ ,  $p < .001$ ,  $d = -0.052$ ; Chemistry:  $z = -5.60$ ,  $p < .001$ ,  $d = -0.039$ ) and AP examinations (Biology:  $z = -17.46$ ,  $p < .001$ ,  $d = -0.095$ ; Chemistry:  $z = -24.71$ ,  $p < .001$ ,  $d = -0.143$ ). Furthermore, schools with survey respondents enrolled slightly lower percentages of students eligible for free- or reduced-price lunch programs (Biology,  $z = 15.89$ ,  $p < .001$ ,  $d = 0.094$ ; Chemistry,  $z = 18.28$ ,  $p < .001$ ,  $d = 0.112$ ). However, the effect sizes (using Cohen's  $d$ ) were very small, such that this analysis might be generalizable to the AP science teacher population.

To focus on factors related to improved student learning and achievement in low-SES urban schools (and schools with urban characteristics), the research questions were explored using a reduced sample. This reduced sample included all observations of teachers who responded to the survey and taught in schools with at least 50% of their student body enrolled in free- or reduced-price lunch programs yielding a sample size of 11,800 AP students (Biology: 6,410 students; Chemistry: 5,390 students) and 638 AP teachers (Biology: 318 teachers, Chemistry: 320 teachers).

Of the 11,000 students, 43.4% students were taught in schools that Milner (2012b) would consider "urban intensive" or "urban emergent." The remaining 56.6% students were taught in schools that College Board did not classify as urban schools based on National Center for Education Statistics (NCES) Local Code classification and ZIP code information. However, these schools exhibited features that Milner (2012b) describes as "urban characteristic"—high levels of poverty, scarcity of resources, and increased numbers of English-language learners. High levels of poverty are related to low SES which is often measured with students' eligibility for free or reduced-price lunches (National Center for Education Statistics, 2011) and/or parental educational attainment (National Center for Education Statistics, 2012). Given the subgroup sampling strategy, at least 50% of students were eligible for free or reduced-price lunches in the selected

schools. Parental median education levels were similarly low for students in schools with urban characteristics (mother: some college; father: business/trade school) compared with “urban intensive” and “urban emergent” schools (mother, father: some college) and considerably lower compared with students not included in the low-SES sample (mother, father: bachelor’s or 4-year college degree). Regarding the scarcity of resources, overall district funding for schools with urban characteristics in the low-SES sample averaged about US\$8,500 per student. In contrast, “urban intensive” and “urban emergent” schools overall district expenditures were slightly higher averaging about US\$9,000 per student. Similarly, overall district expenditures for schools not included in the low-SES sample averaged about US\$9,000 per student. Regarding the number of English-language learners in the community, 17.5% of students in schools with urban characteristics in the low-SES sample did not report English as their first language compared with 11.0% of students not included in the low-SES sample. Thus, the low-SES sample can be considered as a good representation of students and teachers in “urban” settings.

### *Analytical Methods*

Before conducting statistical analyses, data preparation strategies were applied using the full sample, separated by science discipline to reduce sampling biases. *Missing data* were imputed using Markov Chain Monte Carlo multiple imputation methods with 150 iterations and 40 imputations yielding power falloffs less than 1% compared with full-information maximum-likelihood approaches (Graham, 2009; Graham, Olchowski, & Gilreath, 2007). For both student- and school-level imputation models, auxiliary variables were used to improve the imputed estimates. The percentage of missing data was below 5% for almost all variables.

*Composite variables* were computed using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on two randomly sampled equal-sized independent data sets, separated by science discipline. EFA was conducted using the Guttman–Kaiser criterion and scree plot analyses to determine the number of retained factors. Items were gradually excluded from composite variables for factor loadings below 0.25 thresholds, which is conservative compared with conventionally used thresholds of 0.3 to 0.4 (Grice, 2001). Assuming that factors were correlated to each other, parameters were extracted using normalized oblimin oblique rotation methods. CFA used the maximum-likelihood estimation method. Model fits were compared based on the EFA, goodness-of-fit statistics, and likelihood-ratio tests. Bartlett factor scores were computed to create standardized factor scores

(DiStefano, Zhu, & Mindrila, 2009). Cronbach's  $\alpha$  was computed to estimate the reliability of each composite variable.

Exploring the first research question, school-, teacher-, and teaching-level variables were compared between three groups of AP science teachers: teachers whose students performed on average *lower-than-expected*, *as-expected*, and *better-than-expected* on the AP science examination than predicted by students' PSAT scores. To independently test differences across the three groups, parametric one-way ANOVA or non-parametric Kruskal–Wallis *H* tests were conducted. Observations were independent because teachers were uniformly distributed across all three groups. Normality was tested through graphing plots of each variable because ANOVAs are fairly stable against non-normal distributions. Homogeneity of variance was tested using Levene's test based on mean values if the data were normally distributed, Brown-Forsythe's test based on the median if the data were heavily skewed, or Brown-Forsythe's test based on a trimmed mean if the data were heavily tailed. Multiple-group comparisons were conducted using Tukey–Kramer or post hoc Mann–Whitney tests with Bonferroni corrections. Effect sizes were measured using eta-squared; 0.04 (recommended minimum effect size), 0.25 (moderate effect), 0.64 (strong effect; Ferguson, 2009).

For the second research question, direct associations of school, teacher, teaching, and PD participation characteristics with students' AP performance gains were explored using two-level fixed-effect hierarchical linear models (HLMs) with robust standard errors (Raudenbusch & Bryk, 2002), controlling for student-level covariates. Due to missing student–teacher identifiers, schools with more than one AP science teacher in the corresponding discipline were removed from the sample. Therefore, a two-level approach nesting students within teachers/schools was sufficient. Prior to the HLM analyses, the underlying HLM assumptions (Raudenbusch & Bryk, 2002) were tested, and the intraclass correlation coefficient (ICC) was computed. For instance, the observations were independent because student–teacher combinations were uniformly distributed in the data. Multicollinearity of independent variables was tested calculating variance inflation factors on both levels. Homoskedasticity of residuals was tested similarly to Research Question 1.

## Measures

The *dependent variable* used for the HLM analyses was a continuous variable comparing students' actual performance on the AP science examination with their predicted performance based on their PSAT examination scores. Students' PSAT performance was used as an academic achievement measure

prior to students' enrollment in AP science courses. This difference between students' actual AP science scores and students' predicted AP science scores was called "AP performance gain" (Biology:  $n = 6,410$ ,  $M = -0.110$ ,  $SD = 0.650$ ; Chemistry:  $n = 5,390$ ,  $M = -0.167$ ,  $SD = 0.834$ ). Positive AP performance gains indicated that students performed better-than-expected on the AP examination than predicted by the PSAT examination and vice versa. The rationale for using students' AP performance gains instead of students' AP science scores is twofold: First, teachers were classified into groups based on their students' AP performance gains. As prior knowledge often strongly predicts current knowledge, teacher-level effects on student learning would be more difficult to detect if such teacher groupings were not controlling for students' prior knowledge. Second, this study attempted to identify factors related to improved student performance beyond students' predicted AP scores by the PSAT examination attempting to generate more intuitive implications for educational policy makers and practitioners.

The data suggest that PSAT scores strongly correlate with AP science scores,  $r = .672$ ,  $p < .001$ , which is consistent with prior research (Ewing, Camara, & Millsap, 2006; Lichten, 2010; Lichten & Wainer, 2000), such that students' PSAT scores can be viewed as predictors of AP science performance. Students' AP performance gains were computed separate for each science discipline applying linear regressions using every student's PSAT ( $x$ -axis) and AP score ( $y$ -axis). The distance (on the  $y$ -axis) between students' actual AP score and students' projected AP score represented students' AP performance gain. A positive difference indicated that a student was performing better-than-expected on the AP examination and vice versa. To identify teachers whose students performed on average better-than-expected, a continuous variable averaging students' performance gains for all students taught by one teacher ( $n = 638$ ,  $M = -0.179$ ,  $SD = 0.427$ ) was computed.

Single-indicator *independent variables* were included in the analyses on the student-, school-, teacher-, and teaching level (Table 1) as covariates to reduce confounding effects. Student-level variables included students' English-language learner status and dichotomous variables capturing students' racial/ethnic background; the latter were included to account for the intersectionality of race and class on student achievement. School-level variables included districts' per-student funding allocations, the length of the school year, and whether enrollment criteria for AP science courses existed. Teacher- and teaching-level variables included teachers' gender, major, and the number of completed laboratory investigations from the AP laboratory guide.

Similarly, composite *independent variables* were included on the student-, teacher-, and school level (Table 2). Student-level composite independent

**Table 1.** Single-Indicator Independent Variables.

	Description	Range	M (SD)
<b>Level 1 (student characteristics)</b>			
English language <sup>a</sup>	Students' first language (0 = other than English, 1 = English or English and another language)	0, 1	0.808
White <sup>a</sup>	Students' ethnicity: White	0, 1	0.373
Black <sup>a</sup>	Students' ethnicity: Black/African American	0, 1	0.151
Asian <sup>a</sup>	Students' ethnicity: Asian/Asian American or Pacific Islander	0, 1	0.183
Hispanic <sup>a</sup>	Students' ethnicity: Mexican/Mexican American/Puerto Rican, or other Hispanic/Latino/Latin American	0, 1	0.279
Native <sup>a</sup>	Students' ethnicity: American Indian/Alaska Native	0, 1	0.010
<b>Level 2 (school characteristics)</b>			
District funding <sup>b</sup>	Total per-student expenditures in US\$1,000	[3.25, 13.00]	9.00 (2.34)
Length of school year <sup>b</sup>	Length of school year (days)	[1, 351]	275.78 (33.08)
Charter <sup>a</sup>	School is charter school	0, 1	0.204
Enrollment criteria <sup>a</sup>	Enrollment criteria for AP course	0, 1	0.549
<b>Level 2 (teacher and teaching characteristics)</b>			
Female <sup>a</sup>	Teachers' sex is female	0, 1	0.651
Disciplinary major <sup>a</sup>	Major in corresponding discipline, life sciences (Biology)/physical sciences (Chemistry)	0, 1	0.674
Labs <sup>b</sup>	Number of completed laboratory investigations from AP laboratory guide	[0, 16]	6.27 (3.62)

Note. AP = Advanced Placement.

<sup>a</sup>Dichotomous variable ("0"—no and "1"—yes, unless otherwise indicated).

<sup>b</sup>Continuous variable.

[x1, x2]: Every value between x1 and x2 is possible

x1, x2: Only x1 and x2 are possible values

**Table 2.** Composite Independent Variables, Excluding Teachers' PD Participation.

	Description	$\alpha$	Range	$M$ ( $SD$ )
Level 1 (student characteristics)				
Parents' education <sup>a</sup>	Average parental educational attainment (1 = Grade school, 2 = Some high school, 3 = High school diploma, 4 = Business/trade school, 5 = Some college, 6 = Associate's degree, 7 = Bachelor's degree, 8 = Some graduate or professional school, 9 = Graduate/professional degree)	—	[1, 9]	4.73 (2.22)
Level 2 (school characteristics)				
Administrative support <sup>b</sup>	Composite: (a) principal understands challenges for AP science students, <sup>c</sup> (b) principal understands challenges for AP science teachers, <sup>c</sup> (c) principal supports PD, <sup>c</sup> (d) lighter teaching load for AP science teachers, <sup>c</sup> (e) fewer out-of-class responsibilities for AP science teachers, <sup>c</sup> (f) AP science is given additional funding, <sup>c</sup> (g) availability of equipment to perform labs, <sup>d</sup> (h) availability of expendable (consumable) supplies to perform labs <sup>d</sup>	.73	[−2.99, 2.29]	−0.212 (1.125)
AP workload <sup>b</sup>	Composite: (a) number of students across all AP Biology/Chemistry sections, <sup>b</sup> (b) number of AP Biology/Chemistry sections, <sup>b</sup> (c) weekly number of preps <sup>b</sup>	.65	[−1.61, 4.97]	−0.301 (0.865)
Level 2 (teacher and teaching characteristics)				
Teachers' knowledge and experience <sup>b</sup>	Composite: (a) years teaching high school science, <sup>b</sup> (b) years teaching AP Biology/Chemistry, <sup>b</sup> (c) number of science-teaching-related professional organizations, <sup>b</sup> (d) number of conference attendances within the past 3 years, <sup>b</sup> (e) years serving as AP reader, (f) years serving as AP consultant, <sup>b</sup> (g) time of assignment to teach AP science <sup>a</sup>	.55	[−1.94, 3.77]	−0.330 (0.857)

*(continued)*

**Table 2. (continued)**

	Description	$\alpha$	Range	$M$ ( $SD$ )
PD inclination <sup>b</sup>	Composite: (a) importance of PD in instructional performance, <sup>c</sup> (b) importance of PD in student performance, <sup>c</sup> (c) effectiveness of self-teaching compared with formal PD participation, <sup>c</sup> (d) efficacy of PD participation, <sup>c</sup> (e) enjoyment of participation in face-to-face PDs <sup>c</sup>	.81	[-4.52, 1.81]	0.168 (1.055)
Challenges with the AP redesign <sup>b</sup>	Composite: Challenges with (a) Biology/Chemistry content, <sup>c</sup> (b) organization of Biology/Chemistry content, <sup>c</sup> (c) labs, <sup>c</sup> (d) inquiry labs, <sup>c</sup> (e) format of questions/problems/exams, <sup>c</sup> (f) application of science practices, <sup>c</sup> (g) developing new syllabi, <sup>c</sup> (h) understanding the “exclusion statements,” <sup>c</sup> (i) designing new student assessments, <sup>c</sup> (j) using the textbook, <sup>c</sup> (k) working with new/different textbooks, <sup>c</sup> (l) pacing of course, <sup>c</sup> (m) moving students to conceptual understandings of Biology/Chemistry <sup>c</sup>	.87	[-2.91, 3.17]	0.179 (1.068)
Enactment: AP practices <sup>b</sup>	Composite: (a) students work on laboratory investigations, <sup>c</sup> (b) provide guidance on integrated content questions, <sup>c</sup> (c) provide guidance on open/free response questions, <sup>c</sup> (d) students report laboratory findings to another, <sup>c</sup> (e) students perform inquiry laboratory investigations <sup>c</sup>	.65	[-4.88, 3.19]	0.051 (1.220)
Enactment: AP curriculum <sup>b</sup>	Composite: (a) refer to the “Big Ideas” of Biology/Chemistry, <sup>c</sup> (b) use science practices outside of the classroom, <sup>c</sup> (c) refer how enduring understandings relate to the “Big Ideas,” <sup>c</sup> (d) refer to learning objectives from AP curriculum, <sup>c</sup> (e) refer to the curriculum framework <sup>c</sup>	.83	[-2.53, 2.40]	0.250 (1.088)

Note. PD = professional development; AP = Advanced Placement.

<sup>a</sup>Ordinal variable, treated as continuous in subsequent analyses.

<sup>b</sup>Continuous variable.

<sup>c</sup>5-point Likert-type scale item.

<sup>d</sup>4-point Likert-type scale item.

[x1, x2]: Every value between x1 and x2 is possible

x1, x2: Only x1 and x2 are possible values

variables included parents' educational level. School-level composite independent variables included teachers' perceived administrative support and AP workload. Teacher-level composite independent variables included teachers' knowledge and experience, PD inclination, enactment of AP redesign practices, enactment of AP redesign curricular elements, and challenges with the AP redesign.

Teachers' PD participation was measured for conventional and unconventional PD activities (Table 3). Conventional PD activities were described through 5-point Likert-type scales describing the "high-quality" PD features active learning experiences, responsiveness to teachers' needs and interests, focus on student work, modeling teaching, and opportunities to build relationships with colleagues. An additional variable inquired whether teachers felt effectively supported for teaching AP by their PD participation. The duration of PD activities was classified as 1 = low duration ( $\leq 8$  hr), 2 = moderate duration (8-40 hr), and 3 = long duration ( $>40$  hr).

Composite variables of conventional PD activities for each PD feature were based on total "exposure," summing up the Likert-type scale scores (0-4) for all PD teachers participated in. Accounting for the dosage of PD exposure, each Likert-type scale score was multiplied by the corresponding PD duration score. These scalar products were added across all PD teachers participated in to generate composite variables for each PD feature. For unconventional PD activities, the composite variables described the total number of unconventional PD activities teachers engage in, separated by face to face and materials (Table 4).

## Findings

### *Key Characteristics of the AP Science Teacher Population*

The first research question attempted to identify distinctive features of the AP science teacher population in low-SES urban schools (and schools with urban characteristics). Teacher characteristics were compared among three AP science teacher groupings: teachers whose students perform on average more than one third of an AP science score lower (*lower-than-expected*,  $n = 232$ ), within a range of one third below and above their predicted score (*as-expected*,  $n = 339$ ), and more than one third of an AP science score higher than students' predicted score (*better-than-expected*,  $n = 67$ ). Table 5 describes omnibus between-groups effects between teacher groupings.

The analysis indicated significant differences for some school-, teacher-, teaching-, and PD-related characteristics across the student performance-based teacher groupings (Table 5). This suggested that the composition of the three



**Table 3.** Description of Teachers' PD Participation Rates.

	Duration	<i>n</i>
<b>Conventional PD activities</b>		
F2F: AP summer institute <sup>a</sup>	2	360
F2F: AP fall workshop <sup>a</sup>	1	125
F2F: Transition to inquiry-based labs workshop <sup>a</sup>	1	35
F2F: Day with AP reader <sup>a,b</sup>	1	14
F2F: Laying the foundation, by NMSI <sup>b</sup>	2	16
F2F: BSCS Leadership Academy, by BSCS and NABT <sup>b</sup>	2	3
Online: Transition to inquiry-based labs <sup>a</sup>	1	18
Online: Introduction to AP Biology/Chemistry <sup>a</sup>	1	19
Online: AP Central Webcast: Exploring atomic structure using photoelectron spectroscopy <sup>a,c</sup>	1	26
Online community: AP online teacher community <sup>a</sup>	<sup>d</sup>	299
Online community: NSTA online community	<sup>d</sup>	49
<b>Unconventional PD activities</b>		
F2F: District/regional/local college/teacher-initiated meetings	—	123
F2F: Mentoring/coaching one-on-one or with other teachers	—	92
F2F: Conferences or conference sessions	—	55
Materials: AP course and exam description <sup>a</sup>	—	609
Materials: AP lab manual <sup>a</sup>	—	543
Materials: Textbook teacher guide and related materials	—	457
Materials: Instructional materials developed by colleagues	—	506
Materials: Articles from magazines or journals	—	315
Materials: Video resources	—	385

Note. PD = professional development; AP = Advanced Placement; NMSI = National Math + Science Initiative; BSCS = Biological Sciences Curriculum Study; NABT = National Association of Biology Teachers; NSTA = National Science Teachers Association.

<sup>a</sup>Provided by the College Board.

<sup>b</sup>Biology only.

<sup>c</sup>Chemistry only.

<sup>d</sup>Teacher self-reports.

[x1, x2]: Every value between x1 and x2 is possible

x1, x2: Only x1 and x2 are possible values

teacher groups was based on different profiles. Differences in student participation in low-SES urban schools (and schools with urban characteristics) did not

**Table 4.** Teachers' PD Participation Patterns.

	Range	<i>M</i>	<i>SD</i>
Conventional PD characteristics			
Active learning	[0, 12]	2.18	2.06
Responsive agenda	[0, 16]	3.24	2.65
Focus on student work	[0, 16]	2.39	2.55
Modeling teaching	[0, 17]	2.62	2.67
Building relationships	[0, 17]	3.54	2.84
Effective support	[0, 18]	3.86	3.12
Unconventional PD activities			
Face to face	[0, 3]	0.42	0.73
Materials	[0, 6]	4.41	1.34

Note. PD = professional development.

[*x*1, *x*2]: Every value between *x*1 and *x*2 is possible

*x*1, *x*2: Only *x*1 and *x*2 are possible values

seem to occur at random or only with respect to inherent student characteristics. Further analyses on the significant differences of the omnibus tests using multigroup comparisons yielded interesting insights (Table 6).

*School-level variables.* ANOVA indicated significant differences regarding schools' overall district funding allocations between the three teacher groups,  $F(2, 635) = 4.58, p < .05, \eta^2 = .014$ . Tukey–Kramer multiple-comparison tests indicated significantly lower district-level per-student funding allocations to schools of teachers in the *lower-than-expected* group ( $M = \text{US\$8,652}, SD = \text{US\$2,397}$ ) compared with the *as-expected* (AE) group ( $M = \text{US\$9,144}, SD = \text{US\$2,245}$ ),  $TK = 3.51, p < .05$ , and the *better-than-expected* group ( $M = \text{US\$9,463}, SD = \text{US\$2,457}$ ),  $TK = 3.56, p < .05$ . Kruskal–Wallis  $H$  tests indicated small significant differences in the days of the school year across the three teacher groups,  $\chi^2(2, 635) = 9.40, p < .01, \eta^2 = .139$ . Post hoc Whitney–Mann  $U$  tests with Bonferroni corrections indicated that the number of days in the school year was significantly lower for teachers in the *lower-than-expected* group ( $M = 270.90, SD = 39.68$ ) compared with the AE teacher group ( $M = 278.43, SD = 28.95$ ),  $U = -2.96, p < .01$ . These findings suggested that contextual features for teachers in the *lower-than-expected* group were substantially less favorable for providing equitable learning opportunities to students because schools of these teachers were given considerable less district funding and teachers needed to prepare students for the AP examinations in considerably fewer days of instruction.

**Table 5.** Level 2 Omnibus Group Comparisons Using ANOVA and Kruskal–Wallis  $H$  tests.

	Test	$F$ or $\chi^2$	$\eta^2$
School characteristics			
District funding	ANOVA	4.58*	.014
Days of school year	Kruskal–Wallis	9.40**	.139
Administrative support	ANOVA	.29	.001
AP workload	Kruskal–Wallis	2.92	.013
Teacher and teaching characteristics			
Knowledge and experience	Kruskal–Wallis	14.20**	.317
PD inclination	Kruskal–Wallis	1.94	.006
Labs	Kruskal–Wallis	4.40	.030
Challenges with AP redesign	ANOVA	3.31*	.010
Enactment: AP practices	Kruskal–Wallis	4.08	.026
Enactment: AP curriculum	ANOVA	1.27	.004
PD characteristics			
Active learning	Kruskal–Wallis	4.64	.034
Responsive agenda	Kruskal–Wallis	8.54*	.114
Focus on student work	Kruskal–Wallis	11.06**	.192
Modeling teaching	Kruskal–Wallis	5.13	.041
Building relationships	Kruskal–Wallis	6.00	.057
Effective support	Kruskal–Wallis	6.79*	.072
Unconventional PD: F2F	Kruskal–Wallis	3.15	.016
Unconventional PD: Materials	Kruskal–Wallis	9.54**	.143

Note. AP = Advanced Placement; PD = professional development.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

*Teacher-level variables.* Kruskal–Wallis  $H$  tests indicated moderate significant differences across the three groups regarding teachers' knowledge and experience,  $\chi^2(2, 635) = 14.20, p < .01, \eta^2 = .317$ . Post hoc Whitney–Mann  $U$  tests with Bonferroni corrections indicated that teachers' knowledge and experience in the *better-than-expected* group ( $M = -0.080, SD = 0.849$ ) were significantly higher compared with teachers in the AE group ( $M = -0.289, SD = 0.872$ ),  $U = -2.20, p < .05$ , and the *lower-than-expected* (LTE) group ( $M = -0.462, SD = 0.819$ ),  $U = -3.44, p < .01$ ; the difference between the AE and LTE groups was also significant,  $U = -2.49, p < .05$ . Note that all mean values were negative because the composite variables were computed using the “full sample” of all AP science teachers responding to the web-based surveys (and not the low-SES sample). Regarding teachers' perceived challenges with the AP redesign, ANOVA indicated significant differences across the

**Table 6.** Level 2 Post Hoc Multiple-Group Comparisons.

	Lower-than-expected [0]		As-expected [1]		Better-than-expected [2]		0→1	1→2	0→2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>p</i>	<i>p</i>
School characteristics									
District funding	8.65	2.40	9.14	2.25	9.46	2.56	*	<i>ns</i>	*
Days of school year	270.9	39.7	278.4	29.0	279.3	25.0	**	<i>ns</i>	<i>ns</i>
Teacher characteristics									
Knowledge and experience	-0.462	0.819	-0.289	0.872	-0.080	0.849	*	*	**
Challenges with AP redesign	0.316	1.096	0.118	1.054	0.011	1.000	<i>ns</i>	<i>ns</i>	<i>ns</i>
PD characteristics									
Responsive agenda	2.96	2.60	3.53	2.72	2.77	2.31	**	<i>ns</i>	<i>ns</i>
Student work	2.26	2.59	2.60	2.55	1.81	2.39	*	**	<i>ns</i>
Effective support	3.52	2.99	4.17	3.19	3.49	3.06	*	<i>ns</i>	<i>ns</i>
Unconventional PD: Materials	4.26	1.34	4.55	1.37	4.25	1.17	**	*	<i>ns</i>

Note. AP = Advanced Placement; PD = professional development.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

three teacher groups below the recommended minimum effect size,  $F(2, 635) = 3.31$ ,  $p < .05$ ,  $\eta^2 = .010$ . Consequently, Tukey–Kramer multiple-comparison tests did not indicate significant differences across the three teacher groups. These findings suggest that the profiles of teachers in the three groups are similar regarding most teacher and teaching characteristics. The exception was that teachers in the groups with higher average student achievement gains were more knowledgeable and experienced. This raises concerns that students whose AP performance was considerably lower than anticipated and who might have needed guidance from highly qualified teachers were not taught by the most able teachers.

**PD characteristics.** Kruskal–Wallis  $H$  tests indicated small significant differences across the three groups regarding teachers' combined ratings of the responsiveness of the agenda of the PD to teachers' interests and needs,  $\chi^2(2, 635) = 8.54$ ,  $p < .05$ ,  $\eta^2 = .014$ , the focus of the PD on student work,  $\chi^2(2, 635) = 11.06$ ,  $p < .01$ ,  $\eta^2 = .0192$ , and how effective teachers felt supported for

teaching the AP redesign,  $\chi^2(2, 635) = 6.79, p < .05, \eta^2 = .072$ . In addition, teachers' unconventional PD participation through materials significantly differed across the three groups,  $\chi^2(2, 635) = 9.54, p < .01, \eta^2 = .143$ . Post hoc Whitney–Mann  $U$  tests with Bonferroni corrections indicated that teachers in the AE group had significantly higher ratings, compared with the LTE group, of their PD experience being responsive to their interests and needs (AE:  $M = 3.53, SD = 2.72$ ; LTE:  $M = 2.96, SD = 2.60$ ),  $U = -2.67, p < .01$ , focusing on student work (AE:  $M = 2.60, SD = 2.55$ ; LTE:  $M = 2.26, SD = 2.59$ ),  $U = -2.09, p < .05$ , and effectively supporting teaching for the redesigned AP course (AE:  $M = 4.17, SD = 3.19$ ; LTE:  $M = 3.52, SD = 2.99$ ),  $U = -2.46, p < .05$ , as well as using significantly more unconventional PD materials (AE:  $M = 4.55, SD = 1.37$ ; LTE:  $M = 4.26, SD = 1.34$ ),  $U = -2.79, p < .01$ . However, surprisingly teachers in the *better-than-expected* group rated their PD experiences regarding focus on student work ( $M = 1.81, SD = 2.39$ ) significantly lower than teachers in the AE group ( $M = 2.60, SD = 2.55$ ),  $U = -2.46, p < .05$ . Also, teachers in *better-than-expected* group used significantly less unconventional PD materials ( $M = 4.25, SD = 1.17$ ) compared with teachers in the AE group ( $M = 4.55, SD = 1.37$ ),  $U = 2.19, p < .05$ . These findings suggested that PD participation patterns varied across the three teacher groups, and they were particularly dissimilar comparing teachers with the AE group whose PD experiences exposed them with the highest dosage of “high-quality” PD characteristics. This indicated that additional factors beyond teachers' PD participation seem vital for elevating student achievement beyond their predictions, contrary to commonly held beliefs of “the more PD engagement, the better student performance.”

### *Associations to Students' AP Science Performance*

The explorations of the teacher grouping profiles identified several distinguishing features providing some indications of what characteristics might relate to better-than-expected student performance. HLMs were applied to detect direct associations on students' performance gains (Table 7). Student-level variables (Level 1) accounted for 75% of the variance in students' performance gains, whereas 25% of the total variance in students' performance gains occurred between schools/teachers (Level 2; ICC = .25). Given that common ICC values in the social sciences range from .05 to .20 (Peugh, 2010), this ICC value justified the added value of multilevel modeling approaches compared with nested ordinary least squares multiple regressions. Most notably, each group of variables (school, teacher and teaching, and PD participation) included in the analysis significantly contributed to explain variance in students' AP performance gains (PD participation

**Table 7.** Fixed-Effect HLMs With Robust Standard Errors.

	Model 1		Model 2		Model 3		Model 4	
AP performance gain	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Level 1 (student characteristics)								
Chemistry (vs. Biology)	-0.088**	0.033	-0.090**	0.032	-0.094*	0.037	-0.104**	0.037
Parents' education level	-0.005	0.003	-0.005	0.003	-0.005	0.003	-0.006	0.003
English language proficiency	-0.138***	0.021	-0.136***	0.021	-0.137***	0.021	-0.137***	0.021
Black/African American	-0.002	0.023	-0.010	0.023	-0.003	0.023	-0.005	0.023
Asian/Asian American	0.031	0.024	0.027	0.024	0.028	0.024	0.029	0.024
Hispanic	-0.068**	0.023	-0.069**	0.023	-0.067**	0.022	-0.067**	0.022
Native	0.036	0.087	0.037	0.087	0.043	0.088	0.041	0.086
Level 2								
Intercept	0.007	0.030	-0.039	0.033	-0.043	0.058	-0.034	0.057
School characteristics								
District funding (US\$1,000 increments)			0.025***	0.007	0.023**	0.007	0.023**	0.007
Days of school year (10-day increments)			0.015***	0.004	0.014***	0.004	0.013**	0.004
Charter school			-0.061	0.077	-0.064	0.074	-0.080	0.084
Administrative support			-0.001	0.015	-0.006	0.015	-0.010	0.014
AP workload			0.022	0.017	0.016	0.017	0.010	0.017
Criteria for AP enrollment			0.084**	0.032	0.088**	0.031	0.094**	0.031
Teacher and teaching characteristics								
Female (vs. male)					0.024	0.036	0.019	0.037
Disciplinary major					-0.017	0.042	-0.025	0.042

(continued)

**Table 7. (continued)**

	Model 1		Model 2		Model 3		Model 4	
AP performance gain	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE
Knowledge and experience					0.076***	0.019	0.075***	0.019
PD inclination					-0.023	0.016	-0.023	0.016
Number of labs					-0.004	0.004	-0.003	0.004
Challenges with the AP redesign					-0.026	0.015	-0.026	0.015
Enactment: AP practices					0.020	0.014	0.019	0.014
Enactment: AP curriculum					-0.040*	0.016	-0.042*	0.016
PD characteristics								
Active learning							-0.018	0.010
Responsive agenda							0.014	0.011
Focus on student work							-0.007	0.011
Modeling teaching							-0.015	0.013
Building relationships							-0.015	0.010
Effective support							0.022*	0.011
Unconventional PD: F2F							0.041*	0.021
Unconventional PD: Materials							-0.007	0.012
Level 2 variance	0.1336		0.1248		0.1157		0.1125	
Explained variance (Level 2; %)	3.38		9.78		16.33		18.66	
$\chi^2$	87.19		32.47		32.88		14.02	
<i>p</i> value	<.001		<.001		<.001		.081	

Note. White was the reference for the race/ethnicity variables; continuous variables were grand mean centered. HLM = hierarchical linear model; AP = Advanced Placement; PD = professional development.

\**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

variables group was approaching significance). School context variables explained 6.40% of the variance,  $\chi^2(8) = 32.47, p < .001$ , teacher and teaching variables explained additional 6.55%,  $\chi^2(8) = 32.88, p < .001$ , and the PD characteristics explained additional 2.33% of the variance in students' AP performance gains,  $\chi^2(8) = 14.02, p = 0.081$ . Analyzing associations on the item level, several patterns emerged, as described below.

*School-level variables.* Validating findings from Research Question 1, districts' total funding allocations were significantly associated with increases in student performance gains,  $b = 0.023, t(615) = 3.29, p < .01$ , indicating that for every additional US\$1,000 per student, students' AP performance increased by 0.023 beyond their PSAT score prediction. This finding suggested that the more financial resources were available to school, the greater the potential for students to perform better-than-expected on the AP science examinations. Also, this finding underlines the importance of sufficient funding for low-SES urban schools (and schools with urban characteristics; Biddle & Berliner, 2003). Increasing the number of days in the school year was significantly associated with an 0.013 AP performance gain for every additional 10 days of the school year,  $b = 0.013, t(615) = 3.30, p < .01$ , which was consistent with findings of Research Question 1 and prior research examining associations of the length of schools with student performance (Marcotte & Hansen, 2010). The lengthier the school year in low-SES urban settings, which assumes that the total hours of instructional time teaching for the AP examination increases, the greater the potential for students to perform better-than-expected on the AP science examinations. Also, this finding alluded that teachers' classroom instruction per se might influence student learning and achievement. Enforcing criteria for student enrollment in AP courses was significantly associated with an 0.094 AP performance gain,  $b = 0.094, t(615) = 3.02, p < .001$ . This finding suggested that restricting access to AP courses, for instance, by increasing selectivity in AP course admission and presumably creating more homogeneous structures enrolling higher percentages of more able students, improved student performance. However, enacting this practice would be contrary to current efforts to increase AP participation of all students striving to narrow opportunity gaps and increase educational equity (The College Board, 2014; Conger et al., 2009; Lichten, 2010; Wyatt & Mattern, 2011).

*Teacher-level variables.* Increased knowledge and experience was significantly associated with student achievement,  $b = 0.075, t(615) = 3.95, p < .001$ , which validated findings of Research Question 1. Roughly a 1-standard-deviation increase in teachers' knowledge and experience composite



corresponded with an 0.075 AP performance gain. This finding suggested that the higher the teachers' expertise, the greater the potential for students to perform better-than-expected on the AP science examinations. In addition, this finding underscores the importance to counteract challenges for low-SES schools to recruit highly qualified and effective teachers (Goldhaber et al., 2015; Isenberg et al., 2013).

Regarding teachers' classroom instruction, self-reported enactment of curricular elements of the AP redesign had a significant negative association with students' AP performance gain,  $b = -0.042$ ,  $t(615) = -2.52$ ,  $p < .05$ , with an 0.042 AP score penalty for about a standard deviation increase of teachers' rating on curricular enactments of the AP redesign. This counterintuitive finding suggested that the higher the teachers' self-reported enactment of AP redesign curriculum elements, the smaller the potential for students to perform better-than-expected on the AP science examinations. Potential explanations might be measurement related. Teachers' perceptions of curricular elements of the AP redesign might differ from College Board's intentions. For instance, teachers might only enact curricular elements on a surface level, thus, self-reporting high enactment while ratings by external classroom observers might be considerably lower.

**PD characteristics.** Each point increase in teachers' rating of a single PD activity as being supportive for teaching redesigned AP courses increased students' AP performance gains by 0.022 of an AP score,  $b = 0.022$ ,  $t(615) = 1.99$ ,  $p < .05$ . Although this PD characteristic was not explicitly included in the Desimone (2009) list of "high-quality" PD features, it is implicitly underlying all PD-related research and might be seen as a meta-PD characteristic. If teachers had not perceived PD experiences as worthy of their time and valuable for their instruction, lacking associations toward changes in teaching practice or improvements on student outcome measures would not have seemed surprising. Similarly, this finding suggested that the more teachers felt effectively supported for their AP teaching as a result of their PD experiences, the greater the potential for students to perform better-than-expected on the AP science examinations.

Each participation in unconventional face-to-face PD activities was associated with a 0.041 AP performance gain,  $b = 0.041$ ,  $t(615) = 1.98$ ,  $p < .05$ . This finding suggested that the more often teachers participated in teacher-initiated meetings, mentoring activities, or conferences, the greater the potential for students to perform better-than-expected on the AP science examinations. Commonalities of these PD activities include its highly collaborative and informal character in which teachers might broaden and deepen their professional networks.

In general, finding direct associations of teachers' PD participation on student achievement is somewhat unexpected because this relationship is mediated by changes in teachers' knowledge and skills and shifts in instructional practices (Desimone, 2009). Also, the strength of these associations is stronger than expected. For example, teachers' participation in two unconventional face-to-face PD activities and in two 1-day conventional PD activities, which were self-reported as maximally effective for supporting AP teaching, corresponded with an average 0.258 student AP performance gain. Being able to detect such direct associations for teachers in low-SES urban schools (or schools with urban characteristics) emphasizes the potential of purposefully selected PD activities to narrow opportunity gaps and to improve student learning and achievement.

### *Limitations and Future Work*

The main limitations of this study were related to the nature of the data source and the applied statistical methods. The major threat to internal validity was that teacher-level data were limited to teachers' self-reports to the web-based surveys. Given the nationwide scope and scale of this project, collecting additional triangulation data, such as classroom observations of teachers' instruction, was not feasible. Threats to external validity were that student identifiers were unique for AP Biology and AP Chemistry, creating the possibility that students taking the AP Biology and AP Chemistry examinations were treated as two separate cases yielding oversampling and a selection bias. However, this bias should be small, given typical science course-taking patterns in high school. Also, with the absence of student-teacher identifiers, student-level data were tied to school-level data. Hence, teachers associated with two or more schools and multiple AP teachers who taught the same AP subject in the same school were removed from the analysis. Future research will evaluate the relevance of this constraint by exploring similarities and differences of school, teacher, teaching, and PD participation characteristics when there are solo versus multiple AP science teachers in a subject.

Methodologically, HLM assumes linear relationships between independent and dependent variables detecting direct effects. However, some relations might be better described through polynomial, exponential, or other relationships. In addition, interaction, mediating, and moderating effects might occur, indicating that independent variables might have indirect, dynamic relationships toward student achievement. Therefore, future studies could extend this research through multilevel structural equation models and path analyses to explore such indirect effects.

## Discussion

### *Scholarly Significance*

As a large-scale, quantitative study (638 teachers, 11,800 students), with a good nationwide representation of the AP science teacher population in low-SES urban schools (and schools with urban characteristics), this study offers a unique contribution to the research base on student achievement. The mandated top-down curriculum and assessment changes to the revised AP science courses and examinations constituted a unique opportunity for research into student achievement related to these large-scale changes and how that achievement is shaped by associations with school, teacher, teaching, and PD participation characteristics. Insights into factors that increase student performance in urban schools (and schools with urban characteristics) beyond predicted scores may generalize to other nationwide educational assessment and curriculum reforms, such as the NGSS or the Common Core State Standards Initiative. To the best of the authors' knowledge, this is the first large-scale study that analyzes associations toward student achievement in low-SES urban schools (and schools with urban characteristics) at an early implementation stage of a nationwide science curriculum reform.

Also, the approach of evaluating students' actual achievement in correspondence with their predicted performance represents an advancement in existing research. This novel approach allows us to simultaneously account for both students' current and prior achievement, for instance, aiding classifications of student performance-based teacher groups. Thus, interpretations of student outcome measures can be shifted toward identifying "what works" to aid students to perform better-than-expected on the high-stakes AP science examinations.

### *Implications and Conclusion*

This study attempts to provide guidance to inform educational policy makers' and school leaders' decision-making processes for narrowing opportunity and income achievement gaps, and fostering educational equity, especially within low-SES urban schools (and schools with urban characteristics). The three main conclusions from this study are as follows:

First, school context matters. This has long been known, of course, but seeing how context matters in relation to a specific high-stakes exam with critical implications for college course-taking is an extension of prior literature in this area (e.g., Roegman & Hatch, 2016). Districts' per-student total funding allocations and the length of the school year have positive significant

associations with students' AP performance gains. Therefore, increasing district's total expenditures per student as well as the length of instruction for teaching AP science in low-SES urban schools (and schools with urban characteristics) could be further explored. Furthermore, to the best of the authors' knowledge, this is the first study that analyzed influences of districts' funding allocations and school year lengths on students' AP science performance.

Second, teachers make a difference. Teachers' knowledge and experience had positive significant associations with students' AP performance gains. Therefore, incentivizing experienced and skilled teachers to be recruited and retained within low-SES urban schools (and schools with urban characteristics) should be further explored. This is one of the few studies that directly relates teachers' knowledge and experience in low-SES schools to student achievement strengthening prior research that stated the need for disadvantaged students to have equitable access to highly qualified teachers (e.g., Isenberg et al., 2013).

Third, PD can help teachers improve student achievement but only in particular circumstances. Participation in PD activities that teachers rated as effective for helping them teach redesigned AP science courses and participation in unconventional face-to-face PD activities such as teacher-initiated meetings, mentoring or coaching activities, and conference participations were positively and significantly associated with students' AP performance gains. Therefore, guiding teachers in low-SES urban schools (and schools with urban characteristics) to purposefully select their PD participations could be further explored. Our data also reinforce findings that PD needs to be coherent with respect to what teachers are asked to do in the classroom (Penuel et al., 2007). When teachers indicated that PD was effective in helping them with core features of AP instruction, their students performed better.

The guiding vision of this study ultimately aims for changes in the educational landscape to narrow opportunity gaps and increase overall student learning and achievement. Our data suggest that teacher participation in purposefully selected PD activities, in alignment with proactive educational policies increasing school funding, days of instruction, and teacher quality, can make a difference in the challenge of assisting students in low-SES urban schools to succeed on their path through the U.S. education system.

## **Acknowledgments**

The authors thank the following people for their contributions to this work: Amy Wheelock and Ted Gardella of the College Board, Allison Scheff of the Massachusetts Board of Higher Education, and the thousands of AP teachers who helped shape and participated in this project.

## Authors' Note

A previous version of this article was presented at the 2015 annual meeting of the American Educational Research Association in Chicago, Illinois. The views contained in this article are those of the authors, and not their institutions, the College Board, or the National Science Foundation.

## Declaration of Conflicting Interests

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

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the National Science Foundation through the Discovery Research PreK-12 program (DRK-12), Award 1221861.

## References

- Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research in science education* (pp. 1105-1149). Mahwah, NJ: Lawrence Erlbaum.
- Achinstein, B., Curry, M. W., Ogawa, R. T., & Athanases, S. Z. (2016). Organizing high schools for Latina/o youth success: Boundary crossing to access and build community wealth. *Urban Education, 51*, 824-854.
- Archer-Banks, D. A., & Behar-Horenstein, L. S. (2012). Ogbu revisited: Unpacking high-achieving African American girls' high school experiences. *Urban Education, 47*, 198-223.
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass.
- Ball, D. L., & Forzani, F. M. (2011). Building a common core for learning to teach: And connecting professional learning to practice. *American Educator, 35*(2), 17-39.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389-407. doi:10.1177/0022487108324554
- Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the National Science Foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching, 44*, 375-395. doi:10.1002/tea.20145
- Biddle, B. J., & Berliner, D. C. (2003). *What research says about unequal funding for schools in America*. San Francisco, CA: WestEd.

- Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School composition and the Black-White achievement gap* (No. NCES 2015-018). Washington, DC: National Center for Education Statistics. Retrieved from <http://eric.ed.gov/?id=ED560723>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3-15.
- Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). How neighborhoods influence NYC teachers careers. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity* (pp. 377-396). New York, NY: Russell Sage.
- Burks, J., & Hochbein, C. (2015). The students in front of us: Form for the current generation of urban high school students. *Urban Education*, 50, 346-376.
- Chajewski, M., Mattern, K. D., & Shaw, E. J. (2011). Examining the role of Advanced Placement exam participation in 4-year college enrollment. *Educational Measurement: Issues and Practice*, 30(4), 16-27.
- Cilesiz, S., & Drotos, S. M. (2016). High-poverty urban high school students' plans for higher education: Weaving their own safety nets. *Urban Education*, 51(1), 3-31.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement* (CPRE Research Report Series No. RR-43). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- The College Board. (2014). *The 10th annual AP report to the nation*. New York, NY: Author.
- Conger, D., Long, M. C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management*, 28, 555-576. doi:10.1002/pam.20455
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession* (A status report on teacher development in the United States and abroad). Washington, DC: National Staff Development Council.
- Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38, 181-199. doi:10.3102/0013189X08331140
- Desimone, L., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, 26, 1-22.
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1-11.
- Domyancich, J. M. (2014). The development of multiple-choice items consistent with the AP Chemistry curriculum framework to more accurately assess deeper understanding. *Journal of Chemical Education*, 91, 1347-1351. doi:10.1021/ed5000185

- Dougherty, C., Mellor, L. T., & Jian, S. (2006). *The relationship between Advanced Placement and college graduation* (No. 2005 AP Study Series, Report 1). Austin, TX: The National Center for Educational Accountability.
- Ewing, M., Camara, W. J., & Millsap, R. E. (2006). *The relationship between PSAT/NMSQT scores and AP examination grades: A follow-up study*. New York, NY: The College Board.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40, 532-538. doi:10.1037/a0015808
- Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education*, 64(5), 426-438.
- Fishman, B., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 19(6), 643-658. doi:10.1016/S0742-051X(03)00059-3.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915-945.
- Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44, 293-307.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576. doi:10.1146/annurev.psych.58.110405.085530
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206-213. doi:10.1007/s11121-007-0070-9
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6, 430-450. doi:10.1037/1082-989X.6.4.430
- Hallett, R. E., & Venegas, K. M. (2011). Is increased access enough? Advanced Placement courses, quality, and success in low-income urban schools. *Journal for the Education of the Gifted*, 34, 468-487. doi:10.1177/016235321103400305
- Hargrove, L., Godin, D., & Dodd, B. (2008). *College outcomes comparisons by AP and non-AP high school experiences*. New York, NY: The College Board.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hébert, T. P., & Reis, S. M. (1999). Culturally diverse high-achieving students in an urban high school. *Urban Education*, 34, 428-457.
- Hill, P. T., Guin, K., & Celio, M. B. (2003). Minority children at risk. In P. E. Peterson (Ed.), *Our schools and our future: Are we still at risk?* (pp. 111-139). Palo Alto, CA: Hoover Institution Press.
- Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher*, 28(2), 26-37.



- Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to effective teaching for disadvantaged students* (No. NCEE 2014-4001). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://eric.ed.gov/?id=ED544345>
- Klopfenstein, K. (2004). Advanced placement: Do minorities have equal opportunity? *Economics of Education Review*, 23, 115-131.
- Klugman, J. (2013). The Advanced Placement arms race and the reproduction of educational inequality. *Teachers College Record*, 115(5), 1-34.
- Lichten, W. (2010). Whither Advanced Placement—Now? In P. M. Sadler, G. Sonnert, R. H. Tai, & K. Klopfenstein (Eds.), *AP: A critical examination of the Advanced Placement program* (pp. 233-243). Cambridge, MA: Harvard Education Press.
- Lichten, W., & Wainer, H. (2000). The aptitude-achievement function: An aid for allocating educational resources, with an Advanced Placement example. *Educational Psychology Review*, 12, 201-228.
- Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics*, 99, 258-271.
- Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (Eds.). (2010). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks, CA: Corwin.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-132). Dordrecht, The Netherlands: Kluwer Academic. Retrieved from [http://link.springer.com/content/pdf/10.1007/0-306-47217-1\\_4.pdf](http://link.springer.com/content/pdf/10.1007/0-306-47217-1_4.pdf)
- Magrogan, S. (2014). Past, present, and future of AP Chemistry: A brief history of course and exam alignment efforts. *Journal of Chemical Education*, 91, 1357-1361. doi:10.1021/ed500096f
- Marcotte, D. E., & Hansen, B. (2010). Time for school. *Education Next*, 10(1), 52-59.
- Mattern, K. D., Marini, J. P., & Shaw, E. J. (2013). *Are AP students more likely to graduate from college on time?* (No. Research Report 2013-5). New York, NY: The College Board.
- Milner, H. R. (2012a). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, 43, 693-718.
- Milner, H. R. (2012b). But what is urban education? *Urban Education*, 47, 556-561.
- Milner, H. R. (2012c). Understanding urban education from the outside in and inside out. *Urban Education*, 47, 1019-1024.
- National Center for Education Statistics. (2011). *The Nation's Report Card: Science 2009* (No. NCES 2011-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- National Center for Education Statistics. (2012). *Improving the measurement of socioeconomic status for the national assessment of educational progress: A theoretical foundation*. Washington, DC: National Center for Education Statistics.



- National Research Council. (2002). *Learning and understanding: Improving advanced study of mathematics and science in U.S. high schools*. Washington, DC: National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Next Generation Science Standards Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: Achieve, Inc. (On behalf of the twenty-six states and partners that collaborated on the NGSS)
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, 81, 376-407. doi:10.3102/0034654311413609
- Organisation for Economic Co-Operation and Development. (2013a). *Country note: Results from PISA 2012. United States*. Paris, France: Author.
- Organisation for Economic Co-Operation and Development. (2013b). *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science* (Vol. I). Paris, France: Author.
- Patrick, H., Mantzicopoulos, P., & Sears, D. (2010). Effective classrooms. In K. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook* (Vol. 2, pp. 443-469). Washington, DC: American Psychological Association Publications.
- Patterson, B. F., Packman, S., & Kobrin, J. L. (2011). *Advanced Placement exam-taking and performance: Relationships with first-year subject area college grades*. New York, NY: The College Board.
- Penuel, W. R., Fishman, B., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921-958.
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85-112. doi:10.1016/j.jsp.2009.09.002
- Raudenbusch, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity* (pp. 91-116). New York, NY: Russell Sage.
- Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership*, 70(8), 10-16.
- Roegman, R., & Hatch, T. (2016). The AP lever for boosting access, success, and equity. *Phi Delta Kappan*, 97(5), 20-25.
- Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, 48, 117-148.
- Schneider, J. (2009). Privilege, equity, and the Advanced Placement program: Tug of war. *Journal of Curriculum Studies*, 41, 813-831. doi:10.1080/00220270802713613
- Scott, T. P., Tolson, H., & Lee, Y.-H. (2010). Assessment of Advanced Placement participation and university academic success in the first semester: Controlling

- for selected high school academic abilities. *Journal of College Admission*, 208, 26-30.
- Shih, T.-H., & Fan, X. (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review*, 4, 26-40. doi:10.1016/j.edurev.2008.01.003
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Tate, W. F. (2008). "Geography of opportunity": Poverty, place, and educational outcomes. *Educational Researcher*, 37, 397-411. doi:10.3102/0013189X08326409
- TeachingWorks. (2016). *High-leverage practices*. Retrieved from <http://www.teachingworks.org/work-of-teaching/high-leverage-practices>
- Thompson, B. R. (2004). Equitable measurement of school effectiveness. *Urban Education*, 39, 200-229. doi:10.1177/0042085903261325
- Ward, N. L. (2006). Improving equity and access for low-income and minority youth into institutions of higher education. *Urban Education*, 41, 50-70. doi:10.1177/0042085905282253
- Wyatt, J. N., & Mattern, K. D. (2011). *Low-SES students and college outcomes: The role of AP fee reductions*. New York, NY: The College Board.
- Yaron, D. J. (2014). Reflections on the curriculum framework underpinning the redesigned Advanced Placement Chemistry course. *Journal of Chemical Education*, 91, 1276-1279. doi:10.1021/ed500103e
- Zarate, M. E., & Pachon, H. P. (2006). *Gaining or losing ground? Equity in offering Advanced Placement courses in California high schools 1997-2003*. Los Angeles, CA: Tomas Rivera Policy Institute.

## Author Biographies

**Christian Fischer** is a doctoral candidate at the University of Michigan School of Education. His current research focuses on empirical approaches to explore and evaluate ways to support science teachers in improving student learning.

**Barry Fishman** is Arthur F. Thurnau professor of Learning Technologies in the University of Michigan School of Information and School of Education. He studies gameful learning, teacher learning, and design-based implementation research.

**Abigail Jurist Levy** is co-director of Science and Math Programs, Education Development Center, Waltham, MA. Her research focuses on the impact of science instruction on student outcomes, and policies and practices that advance the effectiveness of the teaching workforce.

**Arthur Eisenkraft** is distinguished professor of Science Education, professor of Physics, and director of the Center of Science and Math in Context (COSMIC) at UMass Boston. His current scholarly work is focused on curricula innovation, professional development, and sustainable change in school districts.

**Christopher Dede** is the Timothy E. Wirth professor in Learning Technologies at Harvard University. His fields of scholarship include emerging technologies, policy, and leadership. In 2007, he was honored by Harvard University as an outstanding teacher, and in 2011, he was named a Fellow of the American Educational Research Association.

**Frances Lawrenz** is the associate vice president for Research at the University of Minnesota and also the Wallace Professor of Teaching and Learning in the Educational Psychology Department where she specializes in STEM program evaluation.

**Yueming Jia**, PhD, is founder of ABA Research & Educational Consulting in Cambridge, MA. She is an educational researcher and developmental psychologist who has more than 10 years of experience in program evaluation, educational policy research, and child development.

**Janna Fuccillo Kook** is a senior research associate at Education Development Center, Waltham, MA. Her research focuses on effective early childhood education and early STEM learning and teaching, with a focus on the link between teacher professional development and child outcomes.

**Kim Frumin** is a doctoral student at the Harvard Graduate School of Education, focusing on educational technology and research-practice partnerships.

**Ayana McCoy** is an associate project director in the Center of Science and Mathematics in Context (COSMIC) at the University of Massachusetts Boston. Her research interests includes exploring the nexus of social justice and science education by addressing the disparity of students of color in the STEM K-20 pathway and the intersectionality of race, class, and gender within a science education context.