Teaching and Teacher Education 72 (2018) 107–121

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform

Christian Fischer [a, *, 1], Barry Fishman [a, 2], Chris Dede [b, 3], Arthur Eisenkraft [c, 4], Kim Frumin [b, 5], Brandon Foster [d, 6], Frances Lawrenz [e, 7], Abigail Jurist Levy [d, 6], Ayana McCoy [c, 4]

[a] *University of Michigan, Ann Arbor, USA*
[b] *Harvard University, USA*
[c] *University of Massachusetts at Boston, USA*
[d] *Education Development Center, Inc., USA*
[e] *University of Minnesota, USA*

HIGHLIGHTS

- Evaluation of a large-scale, top-down, national curriculum and examination reform.
- Highly valid high-stakes student performance measure.
- Analysis of school-, teacher-, teaching-, and student-level variables.
- Teacher professional development can influence classroom instruction.
- Weak association of some aspects of classroom instruction with student performance.

ARTICLE INFO

Article history:
Received 28 August 2017
Received in revised form
11 December 2017
Accepted 22 February 2018

Keywords:
Science education
Curriculum reform
Professional development
Instructional practice
High-stakes testing
Advanced placement

ABSTRACT

Situated in the context of the Advanced Placement curriculum reform in the sciences, this quantitative study validates selected elements of Desimone's (2009) conceptual framework on teacher professional development. Using national data sets with data from 133 336 students and 7434 teachers, multi-level structural equation models indicate that professional development participation and contextual school- and teacher-level factors influence teachers' classroom practices. In turn, aspects of instructional enactments characteristics are significantly, but very weakly, associated with student performance. Thus, this study reinforces calls to provide teachers opportunities for high-quality professional development and suggests to advance research that identifies effective instructional practices.

* Corresponding author. University of Michigan, School of Education, 610 East University Avenue, Suite 2211 Ann Arbor, MI 48109-1259, USA.

E-mail addresses: chrisfi@umich.edu (C. Fischer), fishman@umich.edu (B. Fishman), chris_dede@gse.harvard.edu (C. Dede), arthur.eisenkraft@umb.edu (A. Eisenkraft), kmf886@mail.harvard.edu (K. Frumin), bfoster@edc.org (B. Foster), lawrenz@umn.edu (F. Lawrenz), alevy@edc.org (A.J. Levy), ayana.mccoy@umb.edu (A. McCoy).

[1] Current affiliation: University of California, Irvine, UCI Teaching and Learning Research Center, 3000 Anteater Instruction and Research Building Irvine, CA 92697-4150, United States of America.

[2] Current affiliation: University of Michigan, School of Information, 105 South State Street, Room 4435 Ann Arbor, MI 48109-1285, United States of America.

[3] Current affiliation: Harvard University, Graduate School of Education, 323 Longfellow Hall, 13 Appian Way, Cambridge, MA 02138, United States of America.

[4] Current affiliation: University of Massachusetts Boston, Center of Science and Math in Context, Wheatley Hall, 4th Floor, Room 181, 100 Morrissey Boulevard, Boston, MA 02125–3393, United States of America.

[5] Current affiliation: Harvard University, Graduate School of Education, 424 Longfellow Hall, 13 Appian Way, Cambridge, MA 02138, United States of America.

[6] Current affiliation: Education Development Center, Inc., 43 Foundry Avenue, Waltham, MA 02453–8313, United States of America.

[7] Current affiliation: University of Minnesota, Department of Educational Psychology, 174 Education Science Building, 56 East River Road, Minneapolis, MN 55455, United States of America.

https://doi.org/10.1016/j.tate.2018.02.011

1. Introduction

In times of changing curricular standards induced through large-scale curricular reforms such as the Common Core State Standards Initiative (2010a, 2010b) or the Next Generation Science Standards (NGSS; NGSS Lead States, 2013), it is critical to prepare teachers for the challenge to adequately align their teaching to new educational landscapes. Desimone's (2009) logic model for studying the effects of professional development (PD) suggests that teachers' PD participation is associated with knowledge and skill gains that relate to changes in instructional practice, which in turn lead to increased student learning and achievement. While this conceptual framework is widely accepted and adopted in the field, validation studies indicate mixed empirical evidence and call for more research to better understand how teacher PD translates into effective practice (Desimone & Garet, 2015). This study responds to this call for research by examining how teachers adapt to the redesign of the Advanced Placement (AP) science program from a perspective of Desimone's (2009) framework.

College Board, the provider of the AP examinations, responded to the recommendations of the National Research Council (2002) and revised the AP program in an attempt to increase student learning and preparation for study beyond high school. The AP program provides opportunities for high school students to engage in rigorous, college-level courses in a broad range of subject areas. Students often regard AP examinations as high-stakes because of perceived benefits for college admission and the potential to count passing scores toward college credit or placement in more advanced disciplinary courses. The revised AP curriculum reduces its former emphasis on broad content coverage and prescribed algorithmic procedures. In turn, the emphasis on scientific practices, critical thinking, inquiry, and depth of understanding of science concepts is increased. These changes are in line with the Framework for K-12 Science Education (National Research Council, 2012a) and the NGSS (NGSS Lead States, 2013). Teachers have strong incentives to engage in PD activities to align their instruction with the new AP program in order to properly prepare their students for the revised AP examinations. Hence, this large-scale, top-down, nationwide curriculum reform constitutes an excellent opportunity to contribute to the in-service secondary science teacher education research base and to validate selected elements of Desimone's (2009) framework for studying the effects of PD. In particular, this study analyzes associations of teachers' PD participation with teachers' instruction, as well as associations of teachers' instruction with student achievement, situated in the corresponding local contexts.

2. Conceptual framework

2.1. Importance and impact of PD participation

As described in Desimone's (2009) framework, the most direct outcomes of teachers' participation in effective PD activities are increases in teacher knowledge and changes in teachers' beliefs which might indirectly enable teachers to modify their classroom instruction.

Characteristics of effective PD activities. In past decades many studies evaluated the impact of professional learning activities to discern characteristics of effective PD for teachers. Desimone (2009) summarizes this research base and identifies active learning, coherence, content focus, collective participation, and duration as core features of high-quality PD. Active learning refers to PD that affords opportunities for teachers to actively contribute to the knowledge and skills building process through activities such as interactive feedback on teaching demonstrations or review of student work. Coherence refers to PD that is connected to existing curriculum implementations, standards, and policies, as well as teachers' prior knowledge, skills, and beliefs. Content focus refers to PD that increases teachers' expertise related to different knowledge domains of teaching. Collective participation refers to affordances of PD activities that enable participation from teachers in similar local contexts such as teachers from the same grade-level, disciplinary concentration, or school. Duration refers to both the total contact time and frequency of teachers' interactions with the PD environment. Notably, this list of design features is similar to other lists of characteristics that constitute high-quality PD. For instance, Borko, Jacobs, and Koellner (2010) emphasized the importance for PD design to situate content in practice, focus on student learning, model teaching practices, afford active learning, help create collaborative professional learning communities, align goals to school settings, and provide on-going and sustainable learning opportunities. Similarly, Darling-Hammond, Hyler, and Gardner (2017) highlighted that the design of effective PD includes a focus on content, incorporation of active learning, support of collaboration, use of models of effective practice, opportunities for coaching and expert support, offers for feedback and reflection, and a sustained duration. Nevertheless, design features of PD activities only represent one aspect that might contribute to effective PD participation. For instance, Kennedy's (2016) review of 28 studies on the influence of PD on instructional practices concluded that PD effectiveness highly varies, even for PD with similar design characteristics. Kennedy (2016) indicated that PD effectiveness also depends on factors such as the PD program's underlying pedagogy to promote teacher learning. Other influences on PD effectiveness might include teachers' attitudes and beliefs with respect to PD, teachers' micro-level interactions during their PD engagement, and contextual school-level factors (Desimone & Garet, 2015; Fore, Feldhaus, Sorge, Agarwal, & Varahramyan, 2015; Kennedy, 2016).

Influence of PD participation on teachers' knowledge and instruction. Numerous research studies indicate that participation in PD that has a focus on content, provides coherent learning experiences, models instructional enactments, affords collective participation, or has high duration are associated with increases in teacher knowledge (e.g., Allen & Penuel, 2015; Fishman et al., 2013; Garet, Porter, Desimone, Birman, & Yoon, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Roth et al., 2011). Besides more formal PD activities, teacher participation in collaborative learning activities that include coaching or peer-mentoring components also possess potential to increase teachers' knowledge and skills (e.g., Bowe & Gore, 2017; Kyndt, Gijbels, Grosemans, & Donche, 2016).

Studies that explored direct associations of teachers' PD participation on the enactment of instructional practices found that PD that focuses on content, provides opportunities for collaborative or collective participation, ensures coherence with local contexts, includes active learning, or offers sustained and frequent exposure to professional learning lead to changes in teachers' classroom instruction (e.g., ; D. K. Cohen & Hill, 2000; Correnti, 2007; Fishman et al., 2013; Garet et al., 2001; Jeanpierre, Oberhauser, & Freeman, 2005; Matsumura, Garnier, & Resnick, 2010; Penuel et al., 2007; Roth et al., 2011).

2.2. Factors related to student learning

At the heart of every curriculum reform and PD activity is the desire to ultimately advance student learning. However, as indicated in Desimone's (2009) framework, relationships of PD participation and student achievement are indirect and mediated by teachers' knowledge and instructional practices. Teachers' classroom instruction can be seen as the most direct teacher-level influence on student learning. Besides teacher-level factors, student background and the local school characteristics also influence student learning.

Influence of PD participation. Although the influence of teachers' PD participation on student learning is mediated by numerous factors, several research studies were able to detect direct effects of PD on student achievement (e.g., D. K. Cohen & Hill, 2000; Fishman, Marx, Best, & Tal, 2003; Lai & McNaughton, 2016; Meissel, Parr, & Timperley, 2016; Penuel, Gallagher, & Moorthy, 2011; Roth et al., 2011).

Influence of teacher and teaching characteristics. A large array of research studies indicated that variations in teacher quality are associated with differences in student achievement (e.g., Aaronson, Barrow, & Sander, 2007; Jackson, Rockoff, & Staiger, 2014; Kane & Staiger, 2008; Rivkin, Hanushek, & Kain, 2005). Notably, the focus of these studies was not to analyze direct effects of classroom instruction with student performance but to identify teacher characteristics associated with increased student performance. Such teacher characteristics are likely to moderate the effectiveness of teachers' instruction (D. K. Cohen & Ball, 1999; Supovitz & Turner, 2000). For instance, teachers' knowledge is often viewed as an important predictor for student performance (Hill, Rowan, & Ball, 2005; Ma, 2010). Dimensions of teacher knowledge considered necessary for high-quality instruction are often characterized by the triad of "subject matter knowledge," "pedagogical content knowledge," and "curricular knowledge" (Shulman, 1986) or the "content knowledge for teaching" framework (Ball, Thames, & Phelps, 2008). Furthermore, teacher knowledge gains can be related to the years of teaching experience. For instance, studies detected direct effects of teachers' years of teaching experience with student achievement (e.g., Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008; Nye, Konstantopoulos, & Hedges, 2004; Papay & Kraft, 2015; Wiswall, 2013). Other important teacher-level influences include teachers' attitudes and beliefs (Klassen & Tze, 2014; Varghese, Garwood, Bratsch-Hines, & Vernon-Feagans, 2016). The impact of teachers' classroom instruction on student learning and achievement is validated in many research studies and research syntheses (e.g., Hattie, 2009, 2012; National Research Council, 2005, 2012b). For instance, research studies in the context of mathematics and science curriculum reforms indicate that teachers' enactments of reform- or inquiry-oriented instructional elements are significantly associated with increases in students' performance (Hamilton et al., 2003; Secker, 2002). Similarly, Desimone, Smith, and Phillips' (2013) study found stronger student achievement gains for teachers who emphasize more advanced topics compared to more procedural skills.

Student characteristics. Students' individual background traits have a substantial influence on student learning and performance. Prior knowledge is often viewed as an important predictor of student achievement. In the context of the AP program, prior studies validated associations between students' performance on the Preliminary Scholastic Aptitude Test (PSAT) and the AP examinations (Ewing, Camara, & Millsap, 2006; Ewing, Huff, & Kaliski, 2010; Zhang, Patel, & Ewing, 2014). However, this relationship can be partly explained by students' socioeconomic status (SES) (Atkinson & Geiser, 2009; Rothstein, 2004). For example, an analysis of student performance on the Program for International Student Assessment (PISA) benchmark assessments estimates that 15% of the variance in student scores is explained by student-level socioeconomic factors. Such factors include family wealth and income, parental educational attainment and occupation, and neighborhood and school resources, among others (National Center for Education Statistics, 2012).

Local school contexts. While it is tempting to attribute teaching quality and student achievement gains in large parts to teacher quality, Kennedy (2010) cautioned that contextual factors outside of teachers' control can influence instruction. Supportive educational leadership is related to participation in professional learning opportunities, self-efficacy, changes in instructional practices, and increases in teacher effectiveness (e.g., Fackler & Malmberg, 2016; Johnson, Kraft, & Papay, 2012; Kraft & Papay, 2014; Ladd, 2009; May & Supovitz, 2011). For instance, a study by Supovitz, Sirinides, and May (2010) in the context of a midsized urban district indicated that both peer influence (e.g., conversations with peers about instruction, seeking and providing assistance regarding instructional topics) and principal leadership (e.g., trusted teacher-principal relationships, principal focuses leadership on instruction) were indirectly associated with increased student learning. Also, Coburn, Russell, Kaufman, and Steins' (2012) indicated that teachers with a strong social network that includes teachers with deep content expertise and teachers who they frequently interact with demonstrate more sustained instructional improvement related to curriculum reforms. Additionally, time allotted for course preparation and course instruction, continuous assignments to teach courses in similar grade levels, and collaboration with and support from other teachers in the school are associated with teachers' instruction and student achievement (Fitchett & Heafner, 2017; Kennedy, 2010; Ost, 2014; Reeves, Pun, & Chung, 2017). Furthermore, school affluence, which is often estimated with measures that describe the availability of resources for classroom instruction, school or district funding, and crime rates, is related to teaching quality and student performance (Sass, Hannaway, Xu, Figlio, & Feng, 2012; Steinberg, Allensworth, & Johnson, 2011; Supovitz & Turner, 2000).

3. Research questions

This study responds to the call for research by Desimone and Garet (2015) to validate Desimone's (2009) framework by analyzing how PD can translate into changes in instructional practice that relates to increased student performance. The research questions of this study are aligned with the study of Desimone, Smith, and Phillips (2013), which is similar in scope and also uses Desimone's (2009) framework. The research questions are as follows:

Research Question 1: What are the relationships among teacher professional development, teacher characteristics, and school characteristics on teachers' self-reported instructional practices?

Research Question 2: What are the relationships among teachers' self-reported instructional practices, school context, and student characteristics on students' performance on the AP science examinations?

From a PD perspective, the first research question assumes an indirect (and not measured) effect of increases in teachers' knowledge and skills induced through teachers' PD participation. The second research question can be viewed as an implicit analysis of the distal effects of teachers' PD participation on student achievement on the AP exams mediated by teachers' classroom instruction. Both research questions are answered using the same statistical modeling framework to account for such implicit relationships.

4. Methodology

4.1. Data sources and sample

This study is connected to a longitudinal National Science Foundation-funded research project. The goals of the larger project are to better understand teachers' PD adoption patterns and their relations to student achievement in response to the AP examination and curriculum reform in the sciences. The data in this study comes from two sources. First, student- and school-level data for all students taking redesigned AP science examinations is provided from the College Board. Student-level data includes student achievement data (i.e., AP and PSAT scores), as well as student-reported family background characteristics (i.e., parental educational attainment). School-level data includes information on the enrollment in the school and in free- or reduced-priced lunch programs. Second, teacher-level information is collected through web-based surveys to all AP Biology, AP Chemistry, and AP Physics in the United States, unless teachers opted out of College Board's official communication. Web-based surveys were administrated instead of other potential methods to survey AP teachers to reduce financial and time costs and increase response rates (Cobanoglu, Warde, & Moreo, 2001). Approximately 30% of the AP teacher population responded to the web-based surveys, which is considered a good response rate for populations of this size (Shih & Fan, 2009). The surveys inquired about PD participation (e.g., quantity and quality of PD), teaching background (e.g., years of teaching experience), school context (e.g., principal support, length of instruction), classroom instruction (e.g., enactment of labs, enactment of AP practices), and concerns (e.g., perceived challenges with AP redesign). Prior to the first administration of the surveys a national panel of experts in PD, AP exams, science content, science education, and measurement expertise critiqued survey pilots. Additionally, the internal validity of survey items was validated using a cognitive interview approach (Desimone & Le Floch, 2004), in which AP teachers were invited to work through survey items using a talk-aloud methodology in order to verify that teacher interpretation of each item matched researchers' intended meaning. This process was pursued iteratively to reduce ambiguity. Subsequent survey reliability was tested through comparisons of survey response distributions across survey disciplines and years.

This study uses data related to the 2014 (AP Biology and AP Chemistry) and 2015 (AP Biology, AP Chemistry, and AP Physics 1) AP science examinations. The first redesigned AP Biology examination was administered in 2013, AP Chemistry in 2014, and AP Physics in 2015, respectively. Therefore, this study describes the second and third year of the AP Biology redesign, the first two years of the AP Chemistry redesign, and the first year of the AP Physics redesign. Table 1 describes the samples sizes after list-wise deletion of observations with missing data. These samples are called "analytical samples." Across all years and disciplines, this study examines data from 133 336 students and 7434 teachers. Mann-Whitney U tests compare the analytical samples with comparison samples that include all other schools/students in the schools to estimate the generalizability of the analytical samples (Table 2). The analyses indicate that students in the analytical samples perform slightly better than the comparison group on the AP examinations (Biology Year 2: $z = 31.55$, $p < 0.001$, $r = 0.071$; Biology Year 3: $z = 27.70$, $p < 0.001$, $r = 0.061$; Chemistry Year 1: $z = 32.42$, $p < 0.001$, $r = 0.091$; Chemistry Year 2: $z = 20.86$, $p < 0.001$, $r = 0.056$; Physics Year 1: $z = 14.95$, $p < 0.001$, $r = 0.038$) and the PSAT examinations (Biology Year 2: $z = 10.48$, $p < 0.001$, $r = 0.026$; Biology Year 3: $z = 8.06$, $p < 0.001$, $r = 0.020$; Chemistry Year 1: $z = 6.56$, $p < 0.001$, $r = 0.020$; Chemistry Year 2: $z = 20.86$, $p < 0.001$, $r = 0.056$; Physics Year 1: $z = 2.25$, $p < 0.05$, $r = 0.006$). Also, the schools of students in the analytical samples have a slightly lower percentage of enrollment in free- or reduced-priced lunch programs than schools in the comparison group (Biology Year 2: $z = 5.93$, $p < 0.001$, $r = 0.065$; Biology Year 3: $z = 5.23$, $p < 0.001$, $r = 0.057$; Chemistry Year 1: $z = 5.61$, $p < 0.001$, $r = 0.071$; Chemistry Year 2: $z = 5.69$, $p < 0.001$, $r = 0.069$; Physics Year 1: $z = 2.80$, $p < 0.01$, $r = 0.042$). All differences between analytical samples and comparison groups are below a 0.1 effect size threshold, constituting a small effect (J. Cohen, 1992; Ferguson, 2009). Therefore, the analytical samples can be viewed as a good representation of the overall AP science population in the United States in the corresponding years and disciplines.

4.2. Measures

Student-level measures. The student-level variables in the analysis include students' examination scores on the redesigned AP Biology, AP Chemistry, and AP Physics 1 examinations. These variables are used as continuous dependent variables in the corresponding models and represent the main student-level outcome of interest. Students' prior achievement is treated as a continuous variable and measured through students' PSAT scores as previous research indicates strong correlation of PSAT scores with students' performance on AP examinations (e.g., Ewing et al., 2006). Mothers' educational attainment is included in the models to describe students' family background, which is assumed to be related to student learning (e.g., Davis-Kean, 2005; Desforges & Abouchaar, 2003; Reeves et al., 2017; Woessmann, 2004). This ordinal variable distinguishes educational attainment in the categories *no post-secondary education*, *some post-secondary education* (including Associate degrees), *Bachelor's degree*, and *graduate degree* (including doctoral and professional degrees). Table 3 summarizes descriptive statistics of the included student-level variables. The appendix includes the question wording and potential answer choices on the web-based surveys.

School-context measures. School-context variables include a continuous variable that describes SES as measured by the school-level percentage of students enrolled in free- or reduced-priced lunch programs. Similar to parental education attainment, lunch program enrollment is often used to describe poverty levels (National Center for Education Statistics, 2011, 2012) and the relationship of SES with student achievement is well documented (OECD, 2013; Sass et al., 2012; Supovitz & Turner, 2000). Additionally, a continuous composite variable that describes teachers' perceived administrative support (i.e., principal understands the challenges for AP science students, principal understand challenges for AP science teachers, principal support PD participations, lighter teaching loads for AP teachers, fewer out-of-class responsibilities for AP teachers, additional funding for AP science, availability of equipment to perform labs, and availability of expendable supplies to perform labs) is included in the models, as supportive school environments are often found to improve teachers' educational effectiveness (Kraft & Papay, 2014; Ladd, 2009; Waters, Marzano, & McNulty, 2003). Table 3 summarizes descriptive statistics of the included school context variables. The appendix includes the question wording and potential answers choices on the web-based surveys.

Teacher characteristics measures. The teacher characteristics variables in the models include teachers' years of AP teaching experience and years of AP redesign experience in the corresponding science discipline. Previous studies indicated that teaching experience is an important factor for increasing teacher effectiveness and student learning (Boyd et al., 2008; Kraft & Papay, 2014; Papay & Kraft, 2015; Wiswall, 2013), especially if the accumulated teaching experience is closely related to current instructional assignments (Ost, 2014). Additionally, also included is a continuous composite variable that describes teachers' self-reported challenges with the AP redesign (i.e., teachers feel challenged with science content, organization of science content, laboratory investigations, inquiry laboratory investigations, format of questions/problems/AP examination, application of science

Table 1
Descriptions of analytical samples.

	Biology Year 2	Biology Year 3	Chemistry Year 1	Chemistry Year 2	Physics Year 1
N – Student-level	29 632	25 195	30 740	24 993	22 776
N – Teacher/School-level	1544	1530	1770	1518	1072

Table 2
Descriptive information of non-response analysis.

	AP scores			PSAT scores		
	N	Mean	SD	N	Mean	SD
Biology Year 2 (in sample)	29 632	3.09	1.02	29 632	166.64	26.82
Biology Year 2 (comparison)	168 966	2.88	1.05	132 080	164.88	27.55
Biology Year 3 (in sample)	25 195	3.08	1.00	25 195	165.47	26.58
Biology Year 3 (comparison)	183 785	2.88	1.03	145 137	163.99	27.61
Chemistry Year 1 (in sample)	30 740	2.86	1.23	30 740	174.59	26.43
Chemistry Year 1 (comparison)	97 350	2.60	1.25	78 278	173.32	27.08
Chemistry Year 2 (in sample)	24 993	2.78	1.21	24 993	173.78	26.63
Chemistry Year 2 (comparison)	111 940	2.60	1.23	92 395	173.20	27.69
Physics Year 1 (in sample)	22 776	2.37	1.18	22 776	168.86	26.27
Physics Year 1 (comparison)	129 489	2.25	1.15	108 353	168.37	27.16

	Percentage free- or reduced-priced lunch program		
	N	Mean	SD
Biology Year 2 (in sample)	1544	26.16%	24.20%
Biology Year 2 (comparison)	6689	30.66%	26.16%
Biology Year 3 (in sample)	1530	29.09%	25.64%
Biology Year 3 (comparison)	6959	33.12%	26.93%
Chemistry Year 1 (in sample)	1770	24.73%	23.37%
Chemistry Year 1 (comparison)	4502	28.78%	25.19%
Chemistry Year 2 (in sample)	1518	26.77%	24.19%
Chemistry Year 2 (comparison)	5198	31.15%	25.84%
Physics Year 1 (in sample)	1072	29.30%	24.58%
Physics Year 1 (comparison)	3349	32.03%	25.77%

Table 3
Descriptive statistics of variables in analysis.

	Biology Year 2		Biology Year 3		Chemistry Year 1		Chemistry Year 2		Physics Year 1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Student-Level										
AP scores	3.09	1.02	3.08	1.00	2.86	1.23	2.78	1.21	2.37	1.18
PSAT scores	166.64	26.82	165.47	26.58	174.59	26.43	173.78	26.63	168.86	26.27
Mother's educational attainment (in N, %)										
No post-secondary	5003	16.88	4376	17.37	4554	14.81	3817	15.27	4093	17.97
Some post-secondary	6338	21.39	5245	20.82	5955	19.37	4784	19.14	4650	20.42
Bachelor's	10 107	34.11	8409	33.38	10 957	35.64	8756	35.03	7873	34.57
Graduate degree	8184	27.62	7165	28.44	9274	30.17	7636	30.55	6160	27.05
Teacher/School-Level										
Hours of AP instruction (in 10 h)	17.22	5.30	17.16	5.30	17.13	5.31	16.78	5.18	15.68	4.36
Number of laboratory investigations	13.59	5.58	13.97	5.53	15.28	5.43	15.18	5.58	16.94	6.26
Enactment of AP practices†	0.01	1.17	−0.01	1.14	−0.03	1.16	−0.02	1.13	−0.03	1.23
Administrative support†	0.01	1.09	−0.01	1.11	−0.03	1.10	−0.03	1.10	−0.02	1.09
Percent free- or reduced lunch program	0.26	0.24	0.29	0.26	0.25	0.23	0.27	0.24	0.29	0.25
Years AP redesign experience	1.89	0.31	2.63	0.71	1.00	0.00	1.88	0.33	1.00	0.00
Years AP teaching experience	6.75	5.99	6.68	6.38	7.02	6.73	7.03	6.74	5.83	6.38
Challenges with AP redesign†	0.00	1.05	0.00	1.04	0.00	1.05	0.02	1.03	0.00	1.04
Active learning	3.88	3.32	3.61	3.35	3.99	2.73	3.36	2.86	3.81	3.05
Responsive agenda	5.46	4.17	5.25	4.06	5.56	3.68	5.05	3.88	4.83	3.80
Modeling teaching	4.38	3.97	4.47	3.99	4.51	3.43	4.29	3.79	4.07	3.42
Focus on student work	4.37	3.91	4.37	3.92	4.29	3.26	4.20	3.65	3.65	3.16
Relationship building	5.71	4.38	5.54	4.47	5.70	3.81	5.05	3.99	5.10	4.02
Effective support	5.87	4.46	5.75	4.55	6.20	4.08	5.68	4.45	5.01	3.93
Number of conventional PDs	1.67	1.17	1.61	1.18	1.81	1.12	1.63	1.17	1.50	1.07
Number of supplementary PDs	4.97	1.35	6.62	1.72	4.09	1.41	5.80	1.72	6.21	2.25

Notes. The provided descriptive statistics represent the variables after the list-wise deletion process but prior to z-score transformations and grand-mean centering; †: Composite variable. Table 4 lists the variables included in the computation of composite variables.

practices, development of new syllabi, “exclusion statements,” design of new student assessments, use of the textbook, pacing of the course, and facilitation of conceptual understandings of science). Table 3 summarizes descriptive statistics of the included teacher characteristics variables. The appendix includes the question wording and potential answers choices on the web-based surveys.

Instructional practice measures. Teachers' classroom teaching is measured with a continuous variable describing teachers' self-reported number of laboratory investigations and a continuous composite variable that consists of teachers' enactment of practice elements related to the AP redesign (i.e., provide guidance on integrated content, provide guidance on open and free response questions, enable students to report laboratory findings to one another, have students perform laboratory investigations, and have students perform inquiry laboratory investigations). Laboratory investigations are often viewed as important for high school science courses and the AP science curriculum redesign further emphasizes the importance of labs to promote inquiry learning (e.g., Magrogan, 2014; National Research Council, 2006; Price & Kugel, 2014). Similarly, research indicates that changes in instructional enactments aligned with more ambitious curricular goals such as inquiry or reform-based instruction are related to improved student performance (e.g., Hamilton et al., 2003; Secker, 2002). Furthermore, the models included a continuous variable that describes the total hours of AP science course instruction, as exposure to instruction is often assumed to be associated with student performance (Fitchett & Heafner, 2017; Marcotte & Hansen, 2010). Table 3 summarizes descriptive statistics of the included instructional practices variables. The appendix includes the question wording and potential answer choices on the web-based surveys.

Professional development measures. Teachers' PD participation is measured with continuous variables that evaluate both quantity and quality of teachers' PD engagement. The variables that measure quantitative aspects of teachers' PD participation describe the number of teachers' self-reported participations in conventional and supplementary PD activities. The variables that measure qualitative aspects of teachers' PD participation are inspired by frameworks of design features for high-quality PD activities. Also, these particular variables describe the degree in which teachers' overall PD exposure includes elements of active learning, has an agenda responsive to teachers' needs and interests, models teaching, has a focus on student work, offers opportunities to build relationships with colleagues, and effectively supports teaching redesigned AP science courses. Numerous research studies relate teachers' PD participation with increases in teachers' knowledge and skills and changes in teaching practices (e.g., Banilower, Heck, & Weiss, 2007; Fishman et al., 2013; Penuel et al., 2011; Roth et al., 2011). Table 3 summarizes descriptive statistics of the included teacher PD variables. The appendix includes the question wording and potential answer choices on the web-based surveys.

4.3. Analytical methods

Prior to the exploration of the research questions, data preparation strategies are applied separately for each discipline and year. The composite variables that describe teachers' perceived administrative support, challenges with the AP redesign, and enactment of AP practice elements use the full sample of teachers responding to the web-based surveys. These composite variables are computed with Bartlett factor scores derived from initial exploratory and confirmatory factor analysis approaches as described in Fischer et al. (2016). PD participation patterns composite variables are computed through summation and scalar multiplication operations. First, the number of conventional and supplementary PD activities are based on teachers' self-reported indication of PD participation on the pre-defined lists of PD activities. Second, teachers self-reported the quality of each conventional PD activity they participated in based on each of these PD features using a 5-point Likert scale (0–4). This rating is multiplied by a duration factor (1 = low duration [$\leq$ 8 h]; 2 = moderate duration [8–40 h]; 3 = high duration [> 40 h]) and summed up across all conventional PD activities a teacher participated in to create overall "exposure" measures as described in Fischer et al. (2016). Table 4 lists all variables included in the computation of all composite variables. The appendix includes the question wording and potential answers choice on the web-based surveys.

The analysis applies multi-level structural equation modeling with students (level 1) nested within teachers/schools (level 2) (Hoyle, 2012; Rabe-Hesketh, Skrondal, & Pickles, 2004; Rabe-Hesketh, Skrondal, & Zheng, 2007). The analysis was conducted in M*plus* 7.4 (Muthén & Muthén, 2015). Table 5 lists all variables included in the analysis and describes whether variables are grand-mean centered and z-score transformed. Teacher- and school-level variables are included on the same level due to the absence of student-teacher identifiers. Students can only be linked to their school. Therefore, schools with more than one AP science teacher in the corresponding discipline are removed from the analytical sample in order to uniquely match students with teachers.

The nature of Desimone's (2009) "Conceptual Framework for Studying the Effects of PD" (p. 185), which is at the heart of this study, suggests a sequential incorporation of core constructs (i.e., PD participation, teacher knowledge, instructional changes, student learning). This is in contrast to hierarchical linear models that examine direct relationships with specific constructs. Furthermore, this study acknowledges that the effectiveness of PD participation patterns is complex. In order to focus on the more general patterns, this study does not attempt to identify particular elements that might distinguish effective from less-effective teacher PD participation patterns by introducing a latent PD variable measurement construct. Fig. 1 describes the resulting multi-level latent variable structural equation models.

Overall, model building of the latent variable structural equation models is guided by both conceptual and statistical considerations (Table 6-7). From a conceptual perspective, variables are selected with respect to the literature base. Also, the models are built to be consistent across disciplines and years. From a statistical perspective, the model optimization processes utilize modification indices, and other strategies, to improve the model fit. Model fit is assessed by goodness-of-fit indices including normed chi-square (χ^2/df), the Tucker-Lewis index (TLI), comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Normed chi-square evaluations indicate that the models slightly exceed some thresholds (Table 6), $\chi^2/\mathrm{df} < 3$ (e.g., Schreiber, Nora, Stage, Barlow, & King, 2006). However, chi-square tests are sensitive to sample size (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003) and the samples of each model can be considered as large (N > 20 000). Following Hu and Bentler (1999), alternative fit indices were computed to provide alternative goodness-of-fit indicators which demonstrate substantially better fit than commonly described standards of sufficient model fit, TLI $\geq$ 0.95, CFI $\geq$ 0.95, RMSEA $\leq$ 0.06, SRMR $\leq$ 0.08 (e.g., Schreiber et al., 2006), across all models.

4.4. Limitations

Limitations of this study relate to the nature of the data sources. The major threat to external validity is the absence of student-teacher identifiers such that student-level data is tied to school-level data. In order to uniquely match students to teachers, only schools with one teacher in the corresponding discipline are included in the analytical samples. However, the non-response analysis indicates that the influence of this threat is minimal. Therefore, the results of the analysis can be interpreted as

Table 4
Variables included in the computations of composite variables.

Administrative support	(a) Principal understands challenges for AP science students[×], (b) principal understands challenges for AP science teachers[×], (c) principal supports PD[×], (d) lighter teaching load for AP science teachers[×], (e) fewer out-of-class responsibilities for AP science teachers[×], (f) AP science is given additional funding[×], (g) availability of equipment to perform labs[◊], (h) availability of expendable (consumable) supplies to perform labs[◊]
Enactment: AP practices	(a) Students work on laboratory investigations[×], (b) provide guidance on integrated content questions[×], (c) provide guidance on open/free response questions[×], (d) students report laboratory findings to another[×], (e) students perform inquiry laboratory investigations[×]
Challenges	Challenges with (a) science content[×], (b) organization of science content[×], (c) labs[×], (d) inquiry labs[×], (e) format of questions/problems/exams[×], (f) application of science practices[×], (g) developing new syllabi[×], (h) understanding the "exclusion statements[×]," (i) designing new student assessments[×], (j) using the textbook[×], (k) working with new/different textbooks[×], (l) pacing of course[×], (m) moving students to conceptual understandings of science[×]
Conventional PD activities[□]	*Face-to-face*: AP Summer Institute[CB], AP Fall Workshop[CB], Transition to inquiry-based labs workshop[CB], Day with AP reader[CB], Laying the foundation by NMSI, BSCS Leadership Academy by BSCS and NABT, Reasoning skills workshop[CB] *Online courses*: Transition to inquiry labs[CB], Introduction to AP Biology/Chemistry/Physics[CB], AP Central webcast – Exploring atomic structure using photoelectron spectroscopy[CB], AP insight[CB] *Online communities*: AP online teacher community[CB], NSTA online teacher community
Supplementary PD activities[□]	*Face-to-face*: District/regional/local college/teacher-initiated meetings, mentoring/coaching one-on-one or with other teachers, conference or conference sessions, Serving as AP exam reader, Serving as AP consultant *Materials*: AP course and exam description[CB], AP lab manual[CB], teacher textbook guide and related materials, student guide – data analysis[CB], teacher guide – quantitative skills and analysis[CB], AP practice exams, materials developed from colleagues, articles from magazines or journals, video resources, computer-based simulations

Note. [□]: Teachers were not asked to indicate their participation on all listed PD activities for each discipline and year, [×]: treated as 5-point Likert scale item, [◊]: treated as 4-point Likert scale item, [CB]: PD provided by the College Board.

Table 5
List of variables included in analysis.

Level 1
Student characteristics
AP scores[†]
PSAT scores[†,‡]
Mothers' educational attainment [D]
Level 2
School characteristics
Percentage of students enrolled in free- or reduced-priced lunch programs[†]
Administrative support[†,‡]
Teaching characteristics
Hours of AP science instruction[†]
Number of laboratory investigations[†]
Enactment of AP science practice elements[†,‡]
Teacher characteristics
Years of AP redesign experience[†]
Years of AP teaching experience[†]
Challenges with the AP redesign[†,‡]
PD characteristics
PD includes active learning[†,‡]
PD has responsive agenda[†,‡]
PD models teaching[†,‡]
PD focuses on student work[†,‡]
PD helps relationship building[†,‡]
PD effectively supports instruction[†,‡]
Number of conventional PD participations[†]
Number of supplementary PD participations[†]

Note. [†]: Grand-mean centered, [‡]: z-score transformed, [D]: Series of dummy variables.

representative for the AP science teacher population. Notably, AP teachers and students are often considered high achievers, which might limit inferences to the overall student and teacher populations in the United States. Threats to internal validity include that data that more explicitly assesses teachers' knowledge was not collected beyond teachers' years of AP teaching experience and teachers' years of experience with the redesigned curriculum. The construct, teachers' perceived challenges with the AP redesign, might bring forth a slightly different concept. The major threat to objectivity is that instructional practice measures are based on teachers' self-reports. The major treat to reliability is that the use of web-based surveys might have introduced measurement error from both survey respondents and the survey instrument itself (Couper, 2000; Dillman, Tortora, & Bowker, 1998). While similar studies also rely on self-reported data (Banilower et al., 2007; Garet et al., 2001; Supovitz & Turner, 2000), its validity and reliability remain unclear (e.g., Desimone, Smith, & Frisvold, 2010). However, given the national scope of this project, the collection of additional data such as classroom observations was not feasible.

From a methodological perspective, limitations of the multilevel structural equation models include that only linear relationships are modeled, which could in part explain the normed chi-square values. Some relationships might be better described with polynomial, exponential, or other relationships. For instance, previous research indicates that the influences of the years of teaching experience are stronger in the first years of teaching compared to

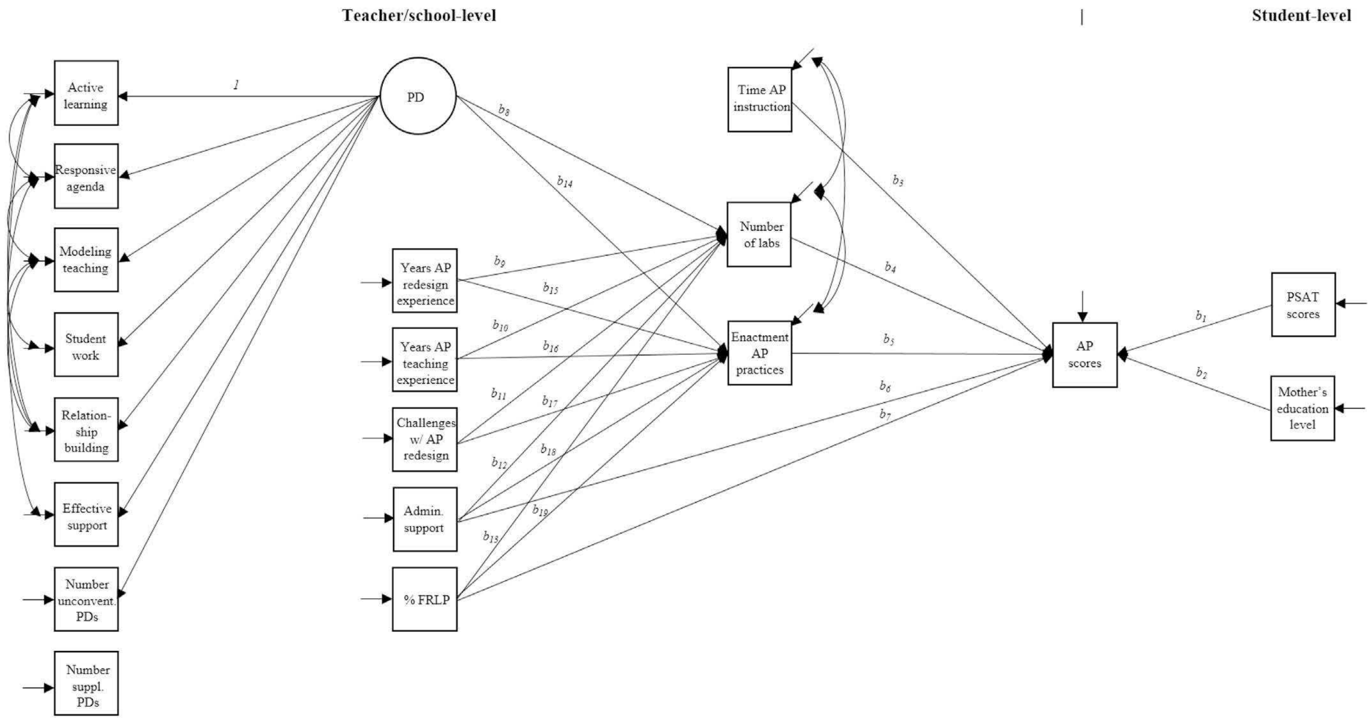

Fig. 1. Model of the relations of teachers' school context, PD participation, instructional enactments, and student performance.

later years in a teaching career (Boyd et al., 2008; Wiswall, 2013). Additionally, in order to follow the sequential logic of Desimone's (2009) framework for studying the effects of PD, the analysis follows what Opfer and Pedder (2011) describe as "process-product logic" (p. 384). From a complexity theory perspective, processes in the educational system are more likely to constitute interdependent, dynamic, and multidimensional relationships (Cochran-Smith, Ell, Ludlow, Grudnoff, & Aitken, 2014) such that the used methodology might oversimplify existing real-life processes. Furthermore, in the attempt to increase consistency across models for all disciplines and years, the model fit for individual models is slightly lower compared to hypothetical models that do not adhere to this consistency principle. Releasing this restriction could increase normed chi-square model fit indicators. Nonetheless, all models in the analysis fulfill recommended model fit thresholds for most fit indices.

5. Results

5.1. Influence on teachers' instructional enactments

The first research question seeks to identify factors that relate to teachers' instructional practices (i.e., the number of laboratory investigations, teachers' enactment of AP science practice elements). The multi-level structural equation models indicate significant associations for teacher PD, teacher, and school characteristics across all disciplines and years (Table 6).

PD participation. Teachers' PD participation is significantly positive associated with the number of enacted laboratory investigations (Biology Year 2: $b = 0.345$, $p < 0.05$; Biology Year 3: $b = 0.391$, $p < 0.05$; Chemistry Year 1: $b = 0.542$, $p < 0.001$; Chemistry Year 2: $b = 0.325$, $p < 0.05$; Physics Year 1: $b = 0.854$, $p < 0.001$) and teachers' enactment of AP science practices (Biology Year 2: $b = 0.187$, $p < 0.001$; Biology Year 3: $b = 0.192$, $p < 0.001$; Chemistry Year 1: $b = 0.179$, $p < 0.001$; Chemistry Year 2: $b = 0.216$, $p < 0.001$; Physics Year 1: $b = 0.225$, $p < 0.001$) across all disciplines and years. These findings indicate that teachers' PD participation can directly influence the enactment of instructional practices in the classroom.

Teacher characteristics. Regarding teacher characteristics, teachers' challenges with the AP redesign and teachers' AP teaching experience are significantly associated with instructional practices. A standard deviation increase in teachers' perceived challenges with the AP design is significantly related to teachers' enactment of 0.50–0.80 fewer laboratory investigations (Biology Year 2: $b = -0.500$, $p < 0.01$; Biology Year 3: $b = -0.664$, $p < 0.001$; Chemistry Year 1: $b = -0.768$, $p < 0.001$; Chemistry Year 2: $b = -0.799$, $p < 0.001$; Physics Year 1: $b = -0.593$, $p < 0.01$) and up to 0.13 standard deviations fewer AP science practices elements in their instruction (Biology Year 2: $b = -0.128$, $p < 0.001$; Biology Year 3: $b = -0.132$, $p < 0.001$; Chemistry Year 1: $b = -0.069$, $p < 0.05$; Chemistry Year 2: $b = 0.027$, n.s.; Physics Year 1: $b = -0.116$, $p < 0.001$). This indicates that teachers who feel more challenged by the AP redesign enact fewer AP redesign related instructional elements.

Notably, teachers' years of AP science teaching experience is only significantly associated with increases in the number of laboratory investigations but not with the enactment of AP science practice elements. A one-year increase in teachers' AP science teaching experience is significantly associated with teachers' enactment of 0.06–0.17 more laboratory investigations (Biology Year 2: $b = 0.172$, $p < 0.001$; Biology Year 3: $b = 0.138$, $p < 0.001$; Chemistry Year 1: $b = 0.096$, $p < 0.001$; Chemistry Year 2: $b = 0.064$, $p < 0.01$; Physics Year 1: $b = 0.127$, $p < 0.001$). This indicates that more experienced AP science teachers enact more laboratory investigations in their classrooms.

School context. The enrollment percentage of students in free- or reduced-priced lunch programs is associated with the number of laboratory investigations across all years and disciplines. A ten percent increase of student enrollment in free- or reduced-priced

Table 6
Multi-level structural equation models.

	Biology Year 2		Biology Year 3		Chemistry Year 1		Chemistry Year 2		Physics Year 1	
	b	SE	*b*	SE	*b*	SE	*b*	SE	*b*	SE
Student-Level										
AP score										
PSAT score, b_1	0.647***	0.005	0.642***	0.005	0.648***	0.007	0.650***	0.007	0.629***	0.010
Mother's educational attainment (vs. no post-secondary), b_2										
Some post-secondary	0.010	0.012	0.022	0.013	−0.037*	0.016	−0.008	0.017	−0.006	0.017
Bachelor's	0.069***	0.013	0.070***	0.014	0.017	0.016	0.031~	0.016	0.040*	0.016
Graduate degree	0.084***	0.014	0.100***	0.015	0.031~	0.016	0.043*	0.018	0.053**	0.019
Teacher/School-Level										
AP score										
Hours of AP instruction (in 10 h), b_3	0.012***	0.002	0.012***	0.002	0.017***	0.002	0.013***	0.003	0.011**	0.003
Number of laboratory investigations, b_4	0.008***	0.002	0.006**	0.002	0.025***	0.003	0.023***	0.003	0.011***	0.003
Enactment of AP practices, b_5	−0.004	0.010	−0.001	0.010	−0.029*	0.013	−0.008	0.014	−0.014	0.018
Administrative support, b_6	0.011	0.010	0.011	0.009	0.008	0.013	0.012	0.014	0.000	0.015
Percent free- or reduced lunch program, b_7	−0.369***	0.045	−0.309***	0.040	−0.734***	0.059	−0.675***	0.061	−0.482***	0.065
Number of laboratory investigations										
PD, b_8	0.345*	0.160	0.391*	0.161	0.542***	0.154	0.325*	0.164	0.854***	0.202
Years AP redesign experience, b_9	0.847*	0.415	0.391~	0.217	–	–	0.644	0.431	–	–
Years AP teaching experience, b_{10}	0.172***	0.025	0.138***	0.026	0.096***	0.020	0.064**	0.022	0.127***	0.029
Challenges with AP redesign, b_{11}	−0.500**	0.145	−0.664***	0.148	−0.768***	0.125	−0.799***	0.142	−0.593**	0.188
Administrative support, b_{12}	0.293*	0.144	−0.047	0.138	0.041	0.128	0.121	0.138	0.071	0.190
Percent free- or reduced lunch program, b_{13}	−1.341*	0.578	−1.468**	0.537	−2.588***	0.561	−3.155***	0.592	−1.258~	0.762
Enactment of AP practices										
PD, b_{14}	0.187***	0.028	0.192***	0.029	0.179***	0.030	0.216***	0.030	0.225***	0.033
Years AP redesign experience, b_{15}	−0.041	0.090	−0.051	0.043	–	–	0.122	0.084	–	–
Years AP teaching experience, b_{16}	0.007	0.005	0.013**	0.005	0.003	0.004	0.004	0.004	0.004	0.005
Challenges with AP redesign, b_{17}	−0.128***	0.027	−0.132***	0.030	−0.069*	0.027	−0.040	0.029	−0.116***	0.032
Administrative support, b_{18}	0.023	0.026	0.051~	0.027	0.007	0.025	0.005	0.026	0.060~	0.032
Percent free- or reduced lunch program, b_{19}	0.092	0.107	−0.023	0.103	0.014	0.108	0.329**	0.109	0.103	0.124
Comparative Fit Index (CFI)	0.990		0.986		0.984		0.984		0.989	
Tucker-Lewis Index (TLI)	0.985		0.980		0.977		0.977		0.984	
Root Mean Square Error of Approx. (RMSEA)	0.011		0.013		0.013		0.013		0.010	
Standard Root Mean Square Residual (SRMR)										
Student-level	0.000		0.000		0.000		0.000		0.000	
Teacher/school-level	0.042		0.053		0.040		0.045		0.043	
χ^2	405.708		470.300		491.355		483.095		277.487	
χ^2/df	4.458		5.168		6.066		5.309		3.426	

Note. ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001; latent variable has all letters capitalized.

Table 7
Description of the latent variable construct in the structural equation models.

	Biology Year 2		Biology Year 3		Chemistry Year 1		Chemistry Year 2		Physics Year 1	
	b	SE	*b*	SE	*b*	SE	*b*	SE	*b*	SE
PD										
Active learning	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
Responsive agenda	1.081***	0.019	1.060***	0.021	1.064***	0.020	1.095***	0.019	1.042***	0.023
Modeling teaching	1.055***	0.028	1.042***	0.023	1.050***	0.026	1.075***	0.026	1.041***	0.027
Focus on student work	1.046***	0.027	1.026***	0.025	1.001***	0.026	1.048***	0.027	0.983***	0.028
Relationship building	1.083***	0.022	1.042***	0.020	1.056***	0.021	1.076***	0.019	1.033***	0.024
Effective support	1.139***	0.023	1.108***	0.021	1.116***	0.023	1.133***	0.024	1.091***	0.025
Number of conventional PDs	1.306***	0.026	1.279***	0.025	1.283***	0.028	1.355***	0.030	1.186***	0.036
Number of supplementary PDs	0.412***	0.039	0.458***	0.049	0.509***	0.041	0.650***	0.055	0.645***	0.077

Note. ~ p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

lunch programs is significantly associated with 0.13–0.32 fewer enacted laboratory investigations in teachers' instruction (Biology Year 2: $b = -1.341$, $p < 0.05$; Biology Year 3: $b = -1.468$, $p < 0.01$; Chemistry Year 1: $b = -2.588$, $p < 0.001$; Chemistry Year 2: $b = -3.155$, $p < 0.001$; Physics Year 1: $b = -1.258$, $p < 0.10$). This indicates that teachers in schools that are economically challenged enact fewer instructional elements related to the AP redesign.

5.2. Influences on student performance

The second research question addresses how instructional enactments relate to students' performance on the AP science examinations. The multi-level structural equation models indicate significant associations for teaching elements, as well as student- and school-context characteristics across all disciplines and years (Table 6).

Classroom instruction. The hours of AP instruction and the number of enacted laboratory investigations have very weak, but significant, associations with students' AP scores across all disciplines and years. However, significance of these weak relationships could be viewed as an artifact of the large sample size. A ten-hour increase in AP science instruction is significantly associated with a

0.01–0.02 AP score increase (Biology Year 2: $b = 0.012$, $p < 0.001$; Biology Year 3: $b = 0.012$, $p < 0.001$; Chemistry Year 1: $b = 0.017$, $p < 0.001$; Chemistry Year 2: $b = 0.013$, $p < 0.001$; Physics Year 1: $b = 0.011$, $p < 0.01$). This indicates that students performed on average marginally better on the AP science examinations, the more AP science instruction exposure they receive. Enactment of one additional laboratory investigation is significantly associated with a 0.01–0.03 AP score increase (Biology Year 2: $b = 0.008$, $p < 0.001$; Biology Year 3: $b = 0.006$, $p < 0.01$; Chemistry Year 1: $b = 0.025$, $p < 0.001$; Chemistry Year 2: $b = 0.023$, $p < 0.001$; Physics Year 1: $b = 0.011$, $p < 0.001$). This indicates that enacting more laboratory investigation slightly increases student performance.

Student characteristics. Students' prior academic achievement and students' family background are significantly associated with students' AP science examination scores across all disciplines and years. A standard deviation increase in students' PSAT scores is significantly associated with a 0.63–0.65 AP score increase (Biology Year 2: $b = 0.647$, $p < 0.001$; Biology Year 3: $b = 0.642$, $p < 0.001$; Chemistry Year 1: $b = 0.648$, $p < 0.001$; Chemistry Year 2: $b = 0.650$, $p < 0.001$; Physics Year 1: $b = 0.629$, $p < 0.001$). This indicates that students' prior mathematics and reading achievement levels helps predict AP performance.

Higher maternal educational attainment is significantly associated with increased student performance. For instance, students whose mothers hold graduate degrees have up to 0.10 higher AP scores compared to students with mothers without postsecondary education (Biology Year 2: $b = 0.084$, $p < 0.001$; Biology Year 3: $b = 0.100$, $p < 0.001$; Chemistry Year 1: $b = 0.031$, $p < 0.10$; Chemistry Year 2: $b = 0.043$, $p < 0.05$; Physics Year 1: $b = 0.053$, $p < 0.01$). Similarly, students whose mothers hold bachelor's degrees have up to 0.07 higher AP scores compared to students with mothers without postsecondary education (Biology Year 2: $b = 0.069$, $p < 0.001$; Biology Year 3: $b = 0.070$, $p < 0.001$; Chemistry Year 1: $b = 0.017$, n.s.; Chemistry Year 2: $b = 0.031$, $p < 0.10$; Physics Year 1: $b = 0.040$, $p < 0.05$). This indicates that students with more educated parents are performing slightly better on the AP examinations.

School context. The school-level enrollment percentage in free- or reduced-priced lunch program is significantly associated with students' AP scores across all disciplines and years. A ten percent increase in students enrolled in free- or reduced-priced lunch programs is associated with a 0.03–0.07 AP score decrease (Biology Year 2: $b = -0.369$, $p < 0.001$; Biology Year 3: $b = -0.309$, $p < 0.001$; Chemistry Year 1: $b = -0.734$, $p < 0.001$; Chemistry Year 2: $b = -0.675$, $p < 0.001$; Physics Year 1: $b = -0.482$, $p < 0.001$). This indicates that school-level socioeconomic factors can help predict students' AP performance.

6. Discussion

6.1. Scholarly significance

This large-scale, quantitative study contributes to the in-service secondary science teacher education research base by analyzing and validating relationships described in Desimone's (2009) framework for studying the effects of PD. The context of the AP science redesign as a nationwide, top-down curriculum reform connected to changes in high-stakes national examinations provides a unique setting for such educational research in the United States. This is the first project that has access to such a comprehensive national data base with student-, teacher-, and school-level variables across multiple science discipline to examine a curriculum and examination reform in the high school science context. Therefore, this study might also allow for generalizations to future or current nationwide curriculum reforms such as the Common Core State Standards Initiative (2010a, 2010b) or the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) for how teachers respond to such large-scale changes in the educational landscape. With respect to the applied methodology, it is one of the few studies that analyzes relationships between school context, PD participation, teacher and teaching characteristics, and student learning using multi-level structural equation modeling.

7. Conclusions, implications, and future work

The findings of this study provide support for some of the relationships described in Desimone's (2009) framework across multiple science disciplines and across different years of the science reform implementation. The main two contributions and their implications are as follows:

First and foremost, this study validates some relationships described in Desimone's (2009) framework for studying the effects of PD. Teachers' PD participation is positively associated with teachers' classroom practice. However, the observed measures portraying elements of instructional practice only have a very small influence on students' performance in the expected direction. This implies that PD participation can make a difference for teachers to change their classroom teaching. This supports perspectives that PD can help teachers to align their instruction with curriculum reforms (e.g., Correnti, 2007; Garet et al., 2001; Penuel et al., 2007). The weak link of instructional elements to student achievement, which corresponds with previous research that found positive but very small effects of reform-oriented instructional elements on student achievement (e.g., Hamilton et al., 2003), could also in part serve as an alternative explanation why several recent PD effectiveness studies did not find considerable direct effects of teachers' PD participation on student achievement (e.g., Arens et al., 2012; Bos et al., 2012; Garet et al., 2011, 2008; Jacob & McGovern, 2015). While PD participation might have produced growth in teachers' knowledge and skills that fostered changes in classroom teaching, instructional changes might not translate to large increases of student learning.

Second, classroom instruction and student learning are situated in their local contexts. On the teacher- and school-level, contextual features such as SES and teachers' years of teaching experience substantially influence teachers' classroom instruction. This mirrors previous studies that emphasize the importance of teacher knowledge, teaching experience, and other teacher-level influences, as well as local context characteristics such as school affluence for shaping classroom instruction (e.g., Garet et al., 2008; Ingvarson, Meiers, & Beavis, 2005; Kennedy, 2010; Supovitz & Turner, 2000). Similarly, contextual features on the student-level such as students' prior mathematics and reading achievement levels and parental educational attainment substantially influence student learning. These findings are in accordance with previous research that detected relationships of students' prior mathematics and reading achievement levels (i.e., PSAT scores) with students' current knowledge (i.e., AP scores) (e.g., Ewing et al., 2006; Zhang et al., 2014), as well as research that relates students' family background with student achievement (Davis-Kean, 2005; Desforges & Abouchaar, 2003; Woessmann, 2004). Thus, this study implies that the mission of advancing teachers' instruction and fostering student learning is multi-faceted and should be approached from several perspectives.

Overall, this study reinforces calls to provide teachers with high-quality professional learning opportunities, to retain experienced teachers in schools, and to guide teachers toward classroom practices that enhance student learning. Furthermore, this study also motivates and illustrates the importance for advancing research in at least two directions. The first set of future studies relates to Opfer and Pedder (2011) conceptualization that teacher professional learning is embedded in the complex system of schooling with its numerous dynamic, interdependent relationships. Motivated by the multitude of detected relationships on teachers' instruction and student learning, future research could go beyond what Opfer and Pedder (2011) describe as "process-product logic" (p. 384) and apply a complexity theory lens (Byrne & Callaghan, 2014; Cochran-Smith et al., 2014; Opfer & Pedder, 2011). The second set of studies is motivated by the detected weak relationship of instructional practices with students' AP scores, which suggests to further analyze immediate influences of specific teaching practices on student learning in more depth. In particular, further research should attempt to identify sets of instructional practices that relate to increased student learning, which in turn should inform future teacher PD activities.

Acknowledgements

The authors thank the following people for their contributions to this work: Amy Wheelock and Ted Gardella of the College Board, former members of the research team Yueming Jia and Janna Fuccillo Kook, and the thousands of AP teachers who helped shape and participated in this project. This work is supported by the National Science Foundation through the Discovery Research PreK-12 program (DRK-12), Award 1221861. The views contained in this article are those of the authors, and not their institutions, the College Board, or the National Science Foundation.

Appendix

The appendix includes questions of the web-based surveys sent to all AP science teachers. Specifically, these questions were taken from the surveys sent to all AP Biology teachers in 2015 who did not respond to previous web-based surveys of the NSF project. Please note that some survey questions and answer choices varied across disciplines and years. Only the questions that were used in the analysis are included in this appendix. Fischer (2017) includes a complete exemplary web-based survey. Descriptions of College Board provided data that was used to create variables used in the analysis (i.e., AP scores, PSAT scores, mothers' educational attainment, percentage of students enrolled in free- or reduced-priced lunch programs) is not included in this appendix.

Table A1
Survey questions and answer choices included in web-based surveys.

Professional development composite variables

There are many different ways that teachers might prepare for teaching AP science. Below, please indicate which of the following resources, informal professional development (PD) activities, or formal PD activities you used as part of your preparation to teach the revised AP curriculum within the past year (since the conclusion of the prior school year in June, 2014).

1. Face to Face or In-Person PD Options Did you participate in any of the following face-to-face or in-person PD activities as part of preparation or support for teaching your AP course during the past year? (Please check all that apply)
   - AP Summer Institute (4–5 days), from the College Board
   - AP Biology Workshop (1 day), from the College Board
   - AP Biology: Transitioning to Inquiry-based Labs workshop (1 day), from the College Board
   - Day with an AP Reader (1 day), from the College Board
   - Laying the Foundation (4–5 days) offered by NMSI
   - BSCS Leadership Academy (4–5 days) offered by NMSI and NABT
   - A district, regional, local college, or teacher-initiated meeting
   - Mentoring or coaching one-on-one or with other teachers
   - Conferences or conference sessions
   - Were you an AP Biology exam reader (in the past year)?
   - Were you an AP Biology Consultant (in the past year)?
   - Other ______________________

2. Self-Paced Online PD Options Did you participate in any self-paced online PD courses as part of preparation or support for teaching your AP course during the past year? (Please check all that apply)
   - Transitioning to Inquiry-Based Labs online PD (6 h, self-paced), from the College Board
   - Introduction to AP Biology online PD (6 h, self-paced), from the College Board
   - AP Insight, from the College Board
   - Other online PD courses? ______________________

3. Online Communities or Discussion Boards Did you participate in any online communities or discussion boards as part of preparation or support for teaching your AP course during the past year? (Please check all that apply)
   - AP Teacher Community (provided by the College Board)
   - National Science Teachers' Association (NSTA) online community
   - Other online communities? ______________________

4. Published or Print Materials Did you use any of the following published, print, or downloaded materials on your own as part of preparation or support for teaching your AP course in the past year? (Please check all that apply)
   - The AP Course and Exam Description from the College Board
   - The AP Lab Manual from the College Board
   - Practice AP Exams from the College Board
   - My textbook teacher guide and related materials

5. Other Kinds of Materials Did you use any of the following other materials as part of preparation or support for teaching your AP course in the past year? (Please check all that apply)
   - Instructional materials developed by colleagues, including handouts, pacing guides, labs, tests and quizzes, etc.
   - Articles from magazines or journals
   - Video resources, such as how-to videos for lab equipment or procedures or video guides to teaching techniques, etc. (NOT including the "Exploring Atomic Structure with PES Data" video from AP Central)

(*continued on next page*)

Table A1 (*continued*)

- Online or computer-based simulations (such as PhET)
- Other ______________________

6. No Professional Development If you did not engage in any PD activities during the past year to help you prepare for teaching the revised AP curriculum, please check the box below.
   - I did not participate in any PD activities related to the teaching of my AP course in the past year.

The following questions refer to the AP Summer Institute (4–5 days)[1] provided by the College Board as PD for your AP Science course. Please respond to the following questions by selecting the best choice from the offered options. Your answers should reflect your personal experience with the AP Summer Institute PD activity.

To what extent did the AP Summer Institute include passive and/or active learning experiences? Active learning might include hands-on activities or small-group activities. More passive forms of PD might consist of lectures or presentations.
- (1) Almost all passive learning
- (2)
- (3) Equal mix of passive and active learning
- (4)
- (5) Almost all active learning

Was the AP Summer Institute responsive to your needs as a participant? For example, was the agenda flexible or customizable to accommodate your (and others') varying interests or needs? Or was the agenda fixed and followed rigidly?
- (1) Almost completely fixed
- (2)
- (3) Equal mix of fixed and responsive
- (4)
- (5) Almost completely responsive

Was student work or materials a focus of the AP Summer Institute? For example, did you examine student lab reports or student test results as a means to understanding common student errors?
- (1) Almost no focus on student work
- (2)
- (3) Some focus on student work
- (4)
- (5) Major focus on student work

Was teaching modeled as part of the AP Summer Institute? Modeling teaching could include observing demonstrations of the type of teaching that would be seen in AP classes or watching videos from AP classes.
- (1) Almost no focus on modeling teaching
- (2)
- (3) Some focus on modeling teaching
- (4)
- (5) Major focus on modeling teaching

To what extent was the AP Summer Institute intentionally designed to provide opportunities to build collegial and/or supportive relationships with other teachers?
- (1) Almost no opportunities to build relationships
- (2)
- (3) Some opportunities to build relationships
- (4)
- (5) Ample opportunities to build relationships

Did the AP Summer Institute effectively support your needs with respect to teaching the revised AP course?
- (1) Not effective
- (2)
- (3) Somewhat effective
- (4)
- (5) Extremely effective

Teacher and school characteristics composite variables

In the current school year (2014-15), the AP redesign may have posed challenges to your instruction. Please indicate below how much of a challenge each of the following elements of the AP redesign was for you. (Reminder: Nobody from the College Board will have access to your individual responses to this or any other question in this survey.)

[5-point Likert scale item: 1 – No challenge at all, 3 – A moderate challenge, 5 – A large challenge]
- Biology content
- The organization of Biology content
- Labs
- Inquiry Labs
- Format of questions/problems/exam
- Application of science practices to the content
- Development of a new syllabus
- Understanding the "boundary statements"
- Designing new student assessments
- Using the textbook for the Biology AP redesign
- Working with a new or different textbook
- The pacing of my course
- Moving my students to a conceptual understanding of Biology

In the current (2014-15) school year, how often did you do each of the following in your AP Biology class?

Table A1 (*continued*)

[1 – Never or only once/year, 2 – Once/quarter, 3 – Once/month, 4 – Once/week, 5 – Nearly every day] ◦ Have students work on laboratory investigations ◦ Have students perform guided inquiry laboratory investigations ◦ Provide guidance on test questions which integrate content and process (e.g., essential knowledge and science practices) ◦ Provide guidance on test questions that are open/free response ◦ Have students report laboratory findings to other students
How much equipment do you have available to perform all the labs you would like to complete? ◦ None or little of the equipment ◦ Some of the equipment ◦ Most of the equipment ◦ All the equipment How much expendable (consumable) supplies do you have available to perform all the labs you would like to complete? ◦ None or little of the expendable supplies ◦ Some of the expendable supplies ◦ Most of the expendable supplies ◦ All of the expendable supplies
These items are about your perceived support from your administrator or principal during the current (2014-15) school year. (Reminder: Nobody from neither the College Board or your school will have access to your individual responses to this or any other question in this survey.)
[5-point Likert scale item: 1 – Strongly disagree, 2 – Disagree, 3 – Neither disagree not agree, 4 – Agree, 5 – Strongly agree] ◦ My principal has a good understanding of how challenging AP science courses are for students ◦ My principal has a good understanding of the challenges of teaching an AP science course ◦ My principal is supportive of teacher participation in PD ◦ I am given a lighter teaching load because I teach an AP science course ◦ In comparison to non-AP teachers, I have fewer out of class responsibilities (e.g., hall duty) ◦ AP science is given additional funding by my administrator exclusively for the course
Other variables
Approximately how many years have you taught AP Biology (not including this year)? [Dropdown menu] ◦ 0 years (This is my first year as a AP Biology teacher) ◦ 1 year ◦ … ◦ 50 years ◦ More than 50 years
Approximately how many lab investigations in total did your students complete in the current (2014-15) school year? [Dropdown menu] ◦ None ◦ 1 lab investigation ◦ … ◦ 30 lab investigations ◦ More than 30 lab investigations
What was the approximate start-date of your AP course this school year? [Dropdown menu] *Month* ◦ January ◦ … ◦ December *Day* ◦ 1 ◦ … ◦ 31
What is the approximate end-date of your AP course this school year? [Dropdown menus] *Month* ◦ January ◦ … ◦ December *Day* ◦ 1 ◦ … ◦ 31
Roughly how many minutes total per week (on average) does each section of your AP Biology class meet? ◦ Less than 150 min ◦ 151 min to 200 min ◦ 201 min to 250 min ◦ 251 min to 300 min ◦ 301 min to 350 min ◦ 351 min to 400 min ◦ 401 min to 450 min ◦ 451 min to 500 min ◦ More than 500 min

[1] These "in-depth" PD question were dynamically displayed to teachers. For each PD activity that teachers checked such questions were displayed. Please note: These questions were displayed for most face-to-face, online, and online community PD activities. For some PD activities additional questions were displayed or selected questions were removed. A full list of questions for each PD activity can be found in Fischer (2017).

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135. https://doi.org/10.1086/508733.

Allen, C. D., & Penuel, W. R. (2015). Studying teachers' sensemaking to investigate teachers' responses to professional development focused on new standards. *Journal of Teacher Education, 66*(2), 136–149. https://doi.org/10.1177/0022487114560646.

Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., et al. (2012). *Effects of curriculum and teacher professional development on the language proficiency of elementary English language learner students in the Central Region (No. NCEE 2012-4013)*. Denver, CO: Mid-Continent Research for Education and Learning.

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher, 38*(9), 665–676. https://doi.org/10.3102/0013189X09351981.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/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*(3), 375–395. https://doi.org/10.1002/tea.20145.

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (pp. 548–556). Oxford, UK: Elsevier.

Bos, J. M., Sanchez, R. C., Tseng, F., Rayyes, N., Ortiz, L., & Sinicrope, C. (2012). *Evaluation of quality teaching for English learners (QTEL) professional development* (No. NCEE 2012-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bowe, J., & Gore, J. (2017). Reassembling teacher professional development: The case for Quality Teaching Rounds. *Teachers and Teaching, 23*(3), 352–366. https://doi.org/10.1080/13540602.2016.1206522.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27*(4), 793–818. https://doi.org/10.1002/pam.20377.

Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art*. New York, NY: Routledge.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research, 43*(4), 441.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education, 119*(1), 137–182. https://doi.org/10.1086/667699.

Cochran-Smith, M., Ell, F., Ludlow, L., Grudnoff, L., & Aitken, G. (2014). The challenge and promise of complexity theory for teacher education research. *Teachers College Record, 116*(5), 1–38.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement* (CPRE Research Report Series No. RR-43). Philadelphia, PA: University of Pennsylvania. Consortium for Policy Research in Education.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294–343.

Common Core State Standards Initiative. (2010a). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Common Core State Standards Initiative. (2010b). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis, 29*(4), 262–295. https://doi.org/10.3102/0162373707309074.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464–494. https://doi.org/10.1086/318641.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19*(2), 294–304. https://doi.org/10.1037/0893-3200.19.2.294.

Desforges, C., & Abouchaar, A. (2003). *The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review* (No. Research Report RR433). Nottingham, UK: Department for Education and Skills.

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140.

Desimone, L., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7*(3), 252–263. https://doi.org/10.25115/psye.v7i3.515.

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), 1–22. https://doi.org/10.3102/01623737026001001.

Desimone, L., Smith, T. M., & Frisvold, D. E. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24*(2), 267–329. https://doi.org/10.1177/0895904808330173.

Desimone, L., Smith, T. M., & Phillips, K. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in title I schools. *Teachers College Record, 115*(5), 1–46.

Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). *Principles for constructing web surveys* (Technical Report No. 90–50). Pullmann, WA.

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.

Ewing, M., Huff, K., & Kaliski, P. (2010). Validating AP exam scores. Current research and new directions. In P. M. Sadler, G. Sonnert, R. H. Tai, & K. Klopfenstein (Eds.), *AP: A critical examination of the advanced placement program* (pp. 85–105). Cambridge, MA: Harvard Education Press.

Fackler, S., & Malmberg, L.-E. (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects. *Teaching and Teacher Education, 56*, 185–195. https://doi.org/10.1016/j.tate.2016.03.002.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. https://doi.org/10.1037/a0015808.

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. https://doi.org/10.1016/S0742-051X(03)00059-3.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., et al. (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.

Fischer, C., Fishman, B., Levy, A., Eisenkraft, A., Dede, C., Lawrenz, F., … McCoy, A. (2016). 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*. OnlineFirst. https://doi.org/10.1177/0042085916668953

Fischer, C. (2017). *Examining formsand frames for science teacher learning related to large-scale reforms: A multi-manuscript dissertation*. Ann Arbor, MI: University of Michigan (Doctoral Thesis).

Fitchett, P. G., & Heafner, T. L. (2017). Student demographics and teacher characteristics as predictors of elementary-age students' history knowledge: Implications for teacher education and practice. *Teaching and Teacher Education, 67*, 79–92. https://doi.org/10.1016/j.tate.2017.05.012.

Fore, G. A., Feldhaus, C. R., Sorge, B. H., Agarwal, M., & Varahramyan, K. (2015). Learning at the nano-level: Accounting for complexity in the internalization of secondary STEM teacher professional development. *Teaching and Teacher Education, 51*, 101–112. https://doi.org/10.1016/j.tate.2015.06.008.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., … Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading Instruction and achievement* (No. NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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*(4), 915–945. https://doi.org/10.3102/00028312038004915.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., … Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (No. NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1–29. https://doi.org/10.3102/01623737025001001.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371.

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. New York, NY: The Guilford Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives, 13*(10), 1–26. https://doi.

org/10.14507/epaa.v13n10.2005.
Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics, 6*(1), 801–825. https://doi.org/10.1146/annurev-economics-080213-040845.
Jacob, A., & McGovern, K. (2015). *The Mirage: Confronting the hard truth about our quest for teacher development*. New York, NY: The New Teacher Project.
Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. *Journal of Research in Science Teaching, 42*(6), 668–690. https://doi.org/10.1002/tea.20069.
Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record, 114*(10), 1–39.
Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. Cambridge, MA: National Bureau of Economic Research.
Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher, 39*(8), 591–598. https://doi.org/10.3102/0013189X10390804.
Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800.
Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001.
Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis, 36*(4), 476–500. https://doi.org/10.3102/0162373713519496.
Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). Teachers' everyday professional development mapping informal learning activities, antecedents, and learning outcomes. *Review of Educational Research, 86*(4), 1111–1150. https://doi.org/10.3102/0034654315627864.
Ladd, H. F. (2009). *Teachers' perceptions of their working conditions: How predictive of policy-relevant outcomes?* (No. Working Paper 33). Washington, DC: Center for the Analysis of Longitudinal Data in Education Research.
Lai, M. K., & McNaughton, S. (2016). The impact of data use professional development on student achievement. *Teaching and Teacher Education, 60*, 434–443. https://doi.org/10.1016/j.tate.2016.07.005.
Ma, L. (2010). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States* (2nd ed.). New York, NY: Routledge.
Magrogan, S. (2014). Past, present, and future of AP Chemistry: A brief history of course and exam alignment efforts. *Journal of Chemical Education, 91*(9), 1357–1361. https://doi.org/10.1021/ed500096f.
Marcotte, D. E., & Hansen, B. (2010). Time for school. *Education Next, 10*(1), 52–59.
Matsumura, L. C., Garnier, H. E., & Resnick, L. B. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis, 32*(2), 249–272. https://doi.org/10.3102/0162373710363743.
May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly, 47*(2), 332–352. https://doi.org/10.1177/0013161X10383411.
Meissel, K., Parr, J. M., & Timperley, H. S. (2016). Can professional development of teachers reduce disparity in student achievement? *Teaching and Teacher Education, 58*, 163–173. https://doi.org/10.1016/j.tate.2016.05.013.
Muthén, L. K., & Muthén, B. O. (2015). *Mplus User's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
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. (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: National Academies Press.
National Research Council. (2006). *America's lab report: Investigations in high school science*. Washington, DC: National Academies Press.
National Research Council. (2012a). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
National Research Council. (2012b). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.
NGSS 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.
Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257. https://doi.org/10.3102/01623737026003237.
OECD. (2013). *Country note. Results from PISA 2012*. United States. Paris, France: OECD Publishing.
Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407. https://doi.org/10.3102/0034654311413609.
Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics, 6*(2), 127–151. https://doi.org/10.1257/app.6.2.127.
Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics, 130*, 105–119. https://doi.org/10.1016/j.jpubeco.2015.02.008.
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.
Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal, 48*(4), 996–1025. https://doi.org/10.3102/0002831211410864.
Price, P. D., & Kugel, R. W. (2014). The new AP Chemistry exam: Its rationale, content, and scoring. *Journal of Chemical Education, 91*(9), 1340–1346. https://doi.org/10.1021/ed500034t.
Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167–190. https://doi.org/10.1007/bf02295939.
Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S. Y. Lee (Ed.), *Handbook on structural equation models* (pp. 209–227). Amsterdam, The Netherlands: Elsevier.
Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education, 67*, 227–236. https://doi.org/10.1016/j.tate.2017.06.016.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x.
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*(2), 117–148. https://doi.org/10.1002/tea.20408.
Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121*(1–2), 297–317. https://doi.org/10.1016/j.jeconom.2003.10.003.
Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics, 72*(2–3), 104–122. https://doi.org/10.1016/j.jue.2012.04.004.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.
Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338.
Secker, C. von (2002). Effects of inquiry-based teacher practices on science excellence and equity. *The Journal of Educational Research, 95*(3), 151–160. https://doi.org/10.1080/00220670209596585.
Shih, T.-H., & Fan, X. (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review, 4*(1), 26–40. https://doi.org/10.1016/j.edurev.2008.01.003.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.2307/1175860.
Steinberg, M. P., Allensworth, E., & Johnson, D. W. (2011). *Student and teacher safety in Chicago public schools: The roles of community context and school social organization*. Chicago, IL: Consortium on Chicago school research at the University of Chicago Urban Education Institute.
Supovitz, J. A., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*(1), 31–56. https://doi.org/10.1177/1094670509353043.
Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963–980.
Varghese, C., Garwood, J. D., Bratsch-Hines, M., & Vernon-Feagans, L. (2016). Exploring magnitude of change in teacher efficacy and implications for students' literacy growth. *Teaching and Teacher Education, 55*, 228–239. https://doi.org/10.1016/j.tate.2016.01.011.
Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. A working paper. Aurora, CO: Mid-Continent Regional Educational Lab.
Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics, 100*, 61–78. https://doi.org/10.1016/j.jpubeco.2013.01.006.
Woessmann, L. (2004). *How equal are educational opportunities? Family background and student achievement in Europe and the US* (CESIFO Working Paper No. 1162). Munich, Germany: Ifo Institute for Economic Research.
Zhang, X., Patel, P., & Ewing, M. (2014). *AP® Potential predicted by PSAT/NMSQT® scores using logistic regression* (No. Statistical Report 2014-1). New York, NY: The College Board.