Professional Development for the Redesigned AP Biology Exam: Teacher Participation Patterns and Student Outcomes

Barry Fishman
Christian Fischer
University of Michigan

Janna F. Kook
Abigail Levy
Yueming Jia
Education Development Center, Inc.

Arthur Eisenkraft
Ayana McCoy
University of Massachusetts at Boston

Frances Lawrenz
University of Minnesota

Chris Dede
Kim Frumin
Harvard University

For more information about this paper or project, contact:
Arthur Eisenkraft, arthur.eisenkraft@umb.edu

Acknowledgements

This work is supported by the National Science Foundation through the DRK-12 program, award number 1221861. The authors also thank the following people for their contributions to this work: Ted Gardella and Amy Wheelock of the College Board, Allison Scheff of the Massachusetts Board of Higher Education, and the thousands of AP teachers who helped shape and participated in this project. The views contained in this paper are those of the authors, and not their institutions, the College Board, or the National Science Foundation.

Professional Development for the Redesigned AP Biology Exam: Teacher Participation Patterns and Student Outcomes

In an era of high-stakes accountability and widespread calls for improved student performance in science, technology, engineering, and math (National Research Council, 2002), it is critical that we also focus on how to support and enhance teachers' learning. Teachers have long been understood to play a key role in the performance of students (e.g., Nye, Konstantopoulos, & Hedges, 2004). Educational policy makers have become increasingly focused on "value added" approaches to gauging teacher performance (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004), which attempt to directly link the contribution of individual teachers to their students' subsequent test performance, in both the near and far term. We take the position that, no matter what one thinks about the current testing and evaluation regime, it makes sense to conduct research to improve our understanding of how to support teachers' ongoing learning and efforts to improve their practice related to student outcomes. This paper reports on a study of teacher learning in a context that is especially apt in the current policy climate – how teachers learn to teach a curriculum associated with a recently-revised high stakes examination. In particular, we report early results from a study of high school teachers learning to teach the revised Advanced Placement Biology curriculum as they prepare students for a high-stakes examination. We examine the role of professional development in supporting teachers' learning to use the revised Advanced Placement Biology curriculum, and the relationship between teachers' professional development choices and subsequent student performance on the Advanced Placement Biology examination.

Background and Research Questions

The Advancement Placement (AP) program is offered by the College Board as a means of introducing rigorous, college-level material to high school students across a broad range of subject areas (The College Board, 2014). The College Board defines curriculum standards for AP courses, and offers corresponding examinations that are administered in centralized locations under controlled conditions and graded centrally for quality control and norming. The examinations are scored on a 1-5 scale. Students who earn a 3 or higher can, depending on the institution of higher education, use their scores towards college-level credit, both as a way to reduce the cost of college and as a way to place into advanced courses upon arrival at college. Increasingly, colleges view AP courses and AP exam performance as important information in the admissions process (Geiser & Santelices, 2006). There are no "official" College Board curriculum materials (though there are a range of well-regarded texts created by third parties in each subject), so teachers put together their own curriculum plan that must be reviewed and certified by the College Board before the course can officially be listed as "Advanced Placement" on high school transcripts.

The three primary AP science curricula and accompanying examinations in Biology, Chemistry, and Physics are in the midst of a major redesign in response to recommendations from the U.S. National Research Council (NRC) report *Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools* (National Research Council, 2002). The changes stress scientific inquiry and reasoning, reduce the emphasis on broad content coverage, and focus on depth of understanding, which aligns with both the new NRC Framework (2012) and the Next Generation Science Standards (Achieve, Inc., 2013). These are sweeping changes to a long-standing educational program with a well-defined examination process, and it is not hard to imagine that teachers facing such a change might be motivated to examine these changes and try to better understand the nature of the revised exam in order to better prepare their students. One logical response is professional development. To help teachers learn about the revisions, the College Board (developer of the AP) offers a range of professional development (PD) options, from week-long summer workshops to short face-to-face (FtF) courses, online self-paced courses, downloadable resources, and online peer-learning communities. There are also two high-quality PD offerings from outside providers included in this study.

This paper investigates data on how teachers responded to changes in AP Biology, which was the first AP exam and curriculum to be modified, in the 2012-13 academic year. The changes in AP Biology represented a range of challenges to teachers' knowledge and practice. The choices teachers made for PD offer a window into their thinking and, in turn, the ability to study how the PD relates to subsequent student outcomes on the AP Biology exam. This is a first report of data from a longitudinal study of teacher learning and student performance in relation to the revised AP courses and exams in Biology, Chemistry, and Physics, in an NSF-funded study called "Professional Development for the Revised AP." Eventually, we will report on differences among the three subject areas, and differences across years as teachers adjust to the new course and exam expectations and format. This paper addresses the following three research questions:

1. What are the patterns (type, number, and combinations) of PD choices that teachers made in response to the AP Biology revisions?
2. How are PD choices and patterns related to teacher characteristics, such as experience, gender, age, and specific concerns about teaching the revised AP Biology course?
3. What is the relationship between the PD patterns of teachers and their students' AP examination outcomes?

Theoretical Framework

Though there are many possible goals for teachers' participation in PD, our research begins from the premise that the ultimate goal of teacher PD is to increase student achievement, with interim objectives related to shifts in instructional practices (e.g., Mundry, Spector, Stiles, & Loucks-Horsley, 1999). A widely accepted theory of action for teacher learning from PD activities, summarized by Desimone (2009), posits that teachers participate in designed learning experiences, such as PD activities, from which they derive knowledge, skills, and attitudes. These skills and attitudes may lead to changes in classroom practice, which in turn may lead to changes in student outcomes. The entire process of teacher professional development exists in a broader policy context, which in the case of our study is the AP ecosystem that links high school science to higher education and to the broader discussion of science education reform, which includes the National Research Council and the Next Generation Science Standards.

The model put forth by Desimone (2009) is important because it explicitly links teacher learning to student learning, something which has been difficult to establish in studies of teacher learning for a variety of reasons (Loucks-Horsley & Matsumoto, 1999). One common reason is the conceptual distance between teacher PD activities and classroom teaching (Fishman, Marx, Best, & Tal, 2003). PD often includes generalized topics, such as collaborative learning or how to facilitate inquiry-based projects, but then the hard work of translating the information of the PD to the classroom is left up to the teacher. This creates two areas of difficulty for research. First, there will be a range of different interpretations of the PD for teachers at different grade levels and even subject areas, which is a fidelity challenge. What exactly were teachers expected to do? Second, there will be no common assessment of student outcomes, which is a measurement challenge. How can we compare student learning across classrooms? The AP exam offers a single common measure of student learning that allows us to overcome this limitation in PD studies. Although there will be various interpretations of PD by teachers in their local context, the importance of aligning instruction with the AP exam may provide a "dampening" effect on local variations, or at least a strong incentive to follow the curriculum framework as presented by the College Board.

Ruiz-Primo and colleagues (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002) discussed an analogous problem to the study of what students learn in relation to instruction in terms of the "instructional sensitivity" of assessments. The most direct information about what students are learning comes from direct or "over the shoulder" observation of student activity in classrooms. As tests become more distant from direct experience, they necessarily become more abstract and thus less likely to measure what was taught or learned by students. National-level tests, such as the SAT or ACT, are labeled "remote" because they may or may not inquire about what was taught to students and, if they do, the assessment may be far removed in time from the instruction. The AP exam operates as an end-of-year test, with explicitly stated content and structure that is closely aligned with the curriculum expectations for AP courses. As noted earlier, teachers still have leeway in designing their local course offering, but all courses are approved by the College Board. Thus, the AP exam is thus what Ruiz-Primo et al. (2002) would term a "distal" assessment, offered at the national level at some remove from the classroom, but guided by a curriculum framework. AP teachers, certainly, have a strong interest in aligning their teaching to the examination. It is for this reason that we find PD associated with the revised AP exam to be an interesting context for study.

Educational researchers have made good progress in understanding the design elements and structure of high-quality PD (Desimone, Garet, Birman, Porter, & Yoon, 2003; Garet, Porter, Desimone, Birman, & Yoon, 2001), but have not yet had much opportunity to examine either different PD offerings related to a particular curriculum (Fishman et al., 2013) or to study combinations of PD in relation to teacher and student learning. Moreover, as Hill, Beisiegel, and Jacob (2013) argue, there is still much we do not understand about how various PD features and designs combine to lead to student outcomes. Furthermore, the intentions of various PD designs will likely be received or interpreted differently by teachers starting at different levels of expertise or teaching in different contexts. In designing the survey that serves as our primary means of examining teacher engagement with PD related to the revised AP Biology curriculum, we sought to better understand how different components of PD interact with individual teacher motivations, interests, and context. We relied primarily on two different discussions of these interactions. The first, by Borko, Jacobs, and Koellner (2010) considers the content, process, and structure of high quality PD, and concludes that the best PD offerings are both situated in teachers' practice and focused on students' learning. In terms of process, Borko, et al. (2010) find that effective PD is ongoing and sustainable, is integrated with other aspects of change within the school, and focuses on modeling preferred instructional strategies building professional learning communities. Opfer and Pedder (2011) take these ideas further in a conceptual discussion that considers the dynamic interactions of individual teachers, school contexts, and PD activity designs (see Figure 1). These conceptualizations of effective PD and the interactions between different components of PD are consistent with the earlier observations of Garet, et al. (2001) and Desimone, et al. (2003), but provide a more nuanced and interactive approach that better suits the nature of our inquiry into teachers' choices from among a range of possible PD offerings to support the revised AP Biology course. These interactions of teacher and PD exist within the larger chain of logic presented by Desimone (2009), and are linked to student outcomes through the AP exam.

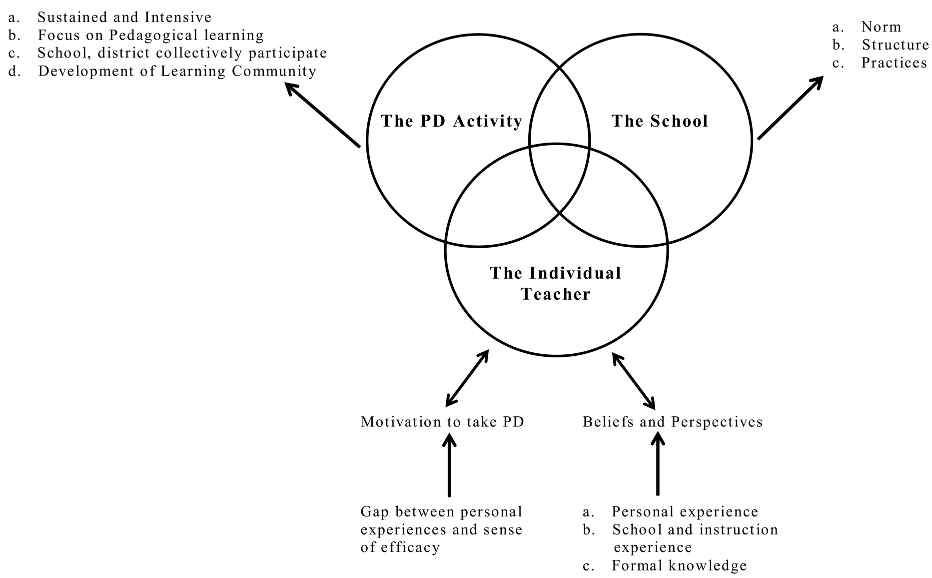


Figure 1. A dynamic model of teacher learning and change, after Opfer and Pedder (2011).

Our study is designed to uncover possible reasons why teachers select particular PD options, and the resulting relationship to their students' performance on a high-stakes, well-validated examination. Explanations for teacher selection of PD opportunities include a range of possibilities, such as teacher experience in science or with the AP exam, preference for online or FtF PD, student demographics, or specific areas of concern about content changes in the revised AP curriculum. Furthermore, we recognize that AP Biology teachers have a range of *choices* for their PD engagements, and thus we examine the effects of different combinations of PD.

Methods

Sample Population

This is a quantitative, quasi-experimental research study employing survey methods. The primary data sources are a web-based survey emailed to 6,796 AP Biology teachers in late May 2013, and student scores on the AP Biology examination taken by students in May 2013. 2,058 teachers responded to our survey, for a total response rate of 30.28% (see below for non-response analysis). The College Board provided data on all students who took the AP Biology exam in May 2013, and combining teacher and student data resulted in a sample size of 50,044 students, out of a total of 187,013 students who took the AP Biology Exam in 2013. A limitation of College Board data is that linking students to individual teachers within a school is not possible, which means that our combined data set excludes teachers and students in cases where there is *more than one* teacher within a school. The combined sample also excludes teachers (and thus students of teachers) who opted-out of communication from the College Board.

Survey Design and Student-level Data

The survey inquired about teacher experience with AP, general teacher background and experience, AP-related PD experience in the prior year, and general attitudes towards PD. In addition, teachers were asked about their concerns related to teaching the revised AP Biology course, and about the nature of their AP Biology course, including school climate issues related to support for AP courses. The survey took between 30 and 45 minutes to complete, and all questions were piloted with experienced AP Biology teachers using cognitive interviewing techniques (Desimone & Carlson Le Floch, 2004).

Student AP exam data, provided by the College Board, is matched to individual teachers, allowing for comparisons of the relationship of teacher PD engagement to student performance on the AP exam, controlling for teacher characteristics such as teaching experience, school context, and teachers' classroom practices with respect to the revised AP Biology course and exam. The AP exam is identical for all students, and administered under controlled conditions, increasing its validity as a comparative measure. AP scores are reported on a scale of 1-5, with 5 as the highest score. In addition to data on student AP Biology performance in 2013, the College Board also provided data related to school socioeconomic status including the percentage of students enrolled in free or reduced-price lunch programs, and students' PSAT examination scores, which can be used as a control for prior academic performance.

Non-Response Analysis

Though a response rate of greater than 30% is considered good for a web-based survey (Shih & Fan, 2009), it is important to understand both what the response population represents and what it does not represent. For evaluating the patterns of non-respondents (students/schools whose teacher respond to the survey), we conducted a threefold analysis using a case-wise deletion approach for missing data. On the student level, we compared students' PSAT and AP Biology scores, using PSAT as a control for prior knowledge. On the teacher/school level, we compared socio-economic status, evaluated through the percentage of students in a school enrolled in free or reduced lunch programs. See Table 1 for a summary of these analyses.

Table 1. Summary of non-response analyses.

		Observations		Mean (SD)
		[n]	[%]	
PSAT Scores	Respondents	41,182	(22.02 %)	169.01 (26.65)
	Non-Respondents	110,829	(59.26 %)	164.95 (27.27)
	Missing	35,002	(18.72 %)	
AP Biology Scores	Respondents	50,044	(26.76 %)	3.04 (.99)
	Non-Respondents	136,969	(73.24 %)	2.84 (1.00)
	Missing	0	(0.00 %)	
Free & Reduced-Price Lunch	Respondents	1,831	(22.43 %)	24.53 % (23.61 %)
	Non-Respondents	5,891	(72.17 %)	30.08 % (25.11 %)
	Missing	441	(5.40 %)	

We conducted non-parametric Mann-Whitney tests to compare the means between responders and non-responders, and calculated the effect size using Cohen's *d*. These tests indicated that the PSAT scores of students whose teachers did not respond to the survey were significantly lower ($z = -25.34$, $p < .001$, $d = -.15$) than students whose teacher did respond to the survey. The AP Biology scores of students whose teachers did not respond to the survey were also significantly lower ($z = -37.42$, $p < .001$, $d = -.20$) than the scores of students whose teachers responded to the survey. And finally, schools where teachers did not respond to the survey have significantly higher percentages of students enrolled in free and reduced lunch programs ($z = 8.28$, $p < .001$, $d = .22$) than schools in which teachers responded to the survey.

In sum, the population of teachers described in our dataset is skewed toward those who teach higher SES students (as measured by free/reduced-price lunch data), and higher academic performance than the AP Biology population as a whole. This is important to consider when interpreting our findings, especially given concerns about equity issues in the AP process (Geiser & Santelices, 2006). In future work, we will pursue strategies to increase response rates from lower SES and lower-performing schools and also investigate reasons for this bias.

Findings

Descriptive Data

The 2,058 AP Biology teachers who responded to our survey ranged in age from 22 to 75 years old, with a mean age of 47.02 years (SD = 10.34 years). The sample population was self-identified as 62.78% female, and 25.56% male (with 11.66% of teachers not reporting gender).

Teachers had a range of experience in teaching at the K-12 level, teaching secondary science, and teaching AP Biology. Additionally, teachers had taken substantial numbers of Biology courses in college and/or graduate school. These data are presented in Table 2.

Table 2. Range of teaching experience, range is from 0 years (new teacher) to "greater than 15 years" and 0 courses to "more than 15 courses" of Biology in college and graduate school.

	25^{th} Percentile	50^{th} Percentile	75^{th} Percentile	Missing [%]
K-12	9 years	15 years	>15 years	0.10
HS Science	8 years	14 years	>15 years	0.24
AP Biology	2 years	5 years	11 years	0.39
# Biology Courses	10 courses	>15 courses	>15 courses	0.29

What did Teachers Find Challenging in the Revised AP Biology Curriculum?

We asked teachers what they found most challenging about the revised AP Biology curriculum. Teachers were presented with a range of specifically-defined features of the new curriculum framework and asked to rate each one on a 5-point scale where 1= "no challenge at all" and 5 = "a large challenge." Teachers indicated concerns with multiple areas of the revised AP Biology curriculum (see Table 3), with the largest numbers of teachers concerned in particular with the revised exam and the need to design new student assessment for use in their courses that would be aligned with the exam.

Table 3. Areas of potential challenge for teachers in the revised AP Biology curriculum. All numbers are percentages and data is presented in descending order of items with the greatest challenges (column "5").

Challenge	1 (No Challenge at All)	2	3	4	5 (A Large Challenge)	Missing
Design of new student assessments	4.96	10.25	26.04	22.55	25.90	10.30
Format of questions/problems/exam	6.22	12.34	26.43	22.21	22.16	10.64
Inquiry Labs (feature of the Redesign)	7.39	9.33	27.70	23.23	21.23	11.13
Pacing of the course	7.19	11.08	29.98	24.78	16.42	10.54
Development of new syllabus	9.14	13.9	32.07	20.60	13.99	10.30
Moving students to a concept. understanding of Bio.	8.79	14.33	29.69	24.88	11.86	10.45
Quantitative Biology	9.04	16.13	29.88	22.74	11.86	10.35
Labs (used in prior AP)	9.18	13.95	34.65	20.17	11.32	10.74
Appropriate use of text	11.52	16.18	32.22	19.19	10.25	10.64
Organization of Biology content	16.38	16.96	32.85	13.27	9.33	11.22
Application of science practices to the content	14.29	20.80	31.68	17.40	5.25	10.59
Use of illustrative examples	18.22	25.07	30.17	11.71	4.13	10.69
Work with a new or different textbook	51.65	11.22	14.29	6.12	3.89	12.83
Biology Content	34.45	20.26	23.91	6.12	3.40	11.86

What Professional Development Options Did Teachers Choose?

We asked teachers to indicate which of twelve professional development activities they participated in. These twelve activities included ten PD activities offered by the College Board: Four face-to-face (FtF) workshops of varying lengths, four self-paced online courses, one resource-based web site (AP Central), and an online teacher community (AP Teacher Community). In addition, we asked teachers to indicate whether or not they participated in either of two extended FtF PD workshops offered by major outside providers. Teachers could indicate whether they were "unaware," "aware but did not attend," "attended," or "led" the PD activity (if a teacher led one of the activities, we did not include them in analyses as a participant). Teachers could select multiple activities. The most frequently attended FtF activities were a 4-5 day Summer Institute, and a shorter 1-day workshop (offered in multiple locations around the country). We also know from College Board data that roughly 50% of the overall AP Biology teacher population attends the 4-5 day Summer Institute, indicating that our survey response pattern is, at least on the surface, consistent with one available indicator of PD participation from objective records. Relatively few teachers used the self-paced online courses, but almost all teachers (88%) reported using the AP Central web site as a support resource, and nearly 66% of teachers indicated some level of participation in the AP Teacher Community. (However, given low levels of actual activity in the AP Teacher Community, teachers may have indicated "membership" as a proxy for professional development, rather than frequent participation.) In the case of low-participation PD activities, many teachers indicated that they were not aware of the option. See Tables 4 and 5 for a summary.

Table 4. Professional Development awareness and participation, reported as percentages, excluding missing data.

		Unaware	Aware / did not attend	Attend	Led
1	AP Biology (1 day, FtF, fall workshop)	21.23	54.91	20.65	1.31
2	AP Biology Summer Institute (4-5 days, FtF)	3.40	45.34	47.52	2.48
3	AP Biology: Transitioning to Inquiry-Based Labs (1 day, FtF)	37.46	48.35	11.95	0.49
4	Day With an AP Biology Reader (1 day, FtF)	75.66	20.12	1.55	0.15
5	Self-Paced Online Course: Transitioning to Inquiry-Based Labs (5-6 hours, Online)	51.17	42.76	4.47	0.00
6	Self-Paced Online Course: Introduction to AP Biology (5-6 hours, Online)	52.53	42.66	3.01	0.10
7	Self-Paced Online Course: AP Biology: Quantitative Skills (2 hours, Online)	66.96	29.83	1.17	0.00
8	Self-Paced Online Course: AP Biology: The Labs (2 hours, Online)	64.24	32.26	1.55	0.00
9	College Board Web Site: AP Central	3.06	7.29	88.05	0.73
10	College Board Web Site: The AP Teacher Community	4.13	28.18	65.94	0.49
11	Laying the Foundation (LTF) Summer Institute (4-5 days, FtF)	76.77	18.61	2.48	0.44
12	BSCS & NABT AP Biology Leadership Academy (4-5 days, FtF)	72.79	24.73	1.17	0.00

Table 5. Professional Development participation, reported as numbers of teachers.

	Professional Development Activity	n
1	AP Biology (1 day, FtF, fall workshop)	425
2	AP Biology Summer Institute (4-5 days, FtF)	970
3	AP Biology: Transitioning to Inquiry-Based Labs (1 day, FtF)	246
4	Day With an AP Biology Reader (1 day, FtF)	32
5	Self-Paced Online Course: Transitioning to Inquiry-Based Labs (5-6 hours, Online)	92
6	Self-Paced Online Course: Introduction to AP Biology (5-6 hours, Online)	62
7	Self-Paced Online Course: AP Biology: Quantitative Skills (2 hours, Online)	24
8	Self-Paced Online Course: AP Biology: The Labs (2 hours, Online)	32
9	College Board Web Site: AP Central	1,805
10	College Board Web Site: The AP Teacher Community	1,354
11	Laying the Foundation (LTF) Summer Institute (4-5 days, FtF)	51
12	BSCS & NABT AP Biology Leadership Academy (4-5 days, FtF)	24

Research Question 1. What are the patterns (type, number, and combinations) of PD choices that teachers made in response to the AP Biology revisions?

Given the large number of PD options available to teachers, the type, number, and combination of PD choices varied widely. In characterizing the PD choices of teachers, the research team decided to exclude the College Board Website, AP Central, as a PD option given the overwhelmingly high proportion of teachers who reported accessing the website, and the ambiguous nature of its use as a PD resource. Across the remaining eleven PD options, teachers participated in an average of 1.62 PD options (SD = 0.99) with a range of zero to seven, and 46 unique patterns of PD participation. In order to distill larger, more generalizable patterns of PD participation, we ran a series of Latent Class Analyses (LCA). 212 participants reported either that they did not participate in any of the eleven PD options (201), or reported not participating in a subset and left the remainder blank (11). This "No PD" group was excluded from the LCA, leaving a total of 1,842 participants' patterns to analyze.

Beginning with a 1-class model and ending with a 7-class model, each model was evaluated for significant improvement in model fit relative to the previous model using a chi-square test (see Table 6). Models were compared based on log-likelihood, BIC, entropy, and the p-value of Lo-Mendell-Rubin likelihood ratio (LMT; Nylund, Asparouhov & Muthén, 2007), and also based on face validity of the groupings and group size. The 7-class model did not have a significantly better fit than the 6-class model. The 6-class model did have a significantly better fit than the 5-class model, but it resulted in a breakdown of groups, which was unevenly distributed across the sample. Consequently, the 5-class model was chosen as the final model.

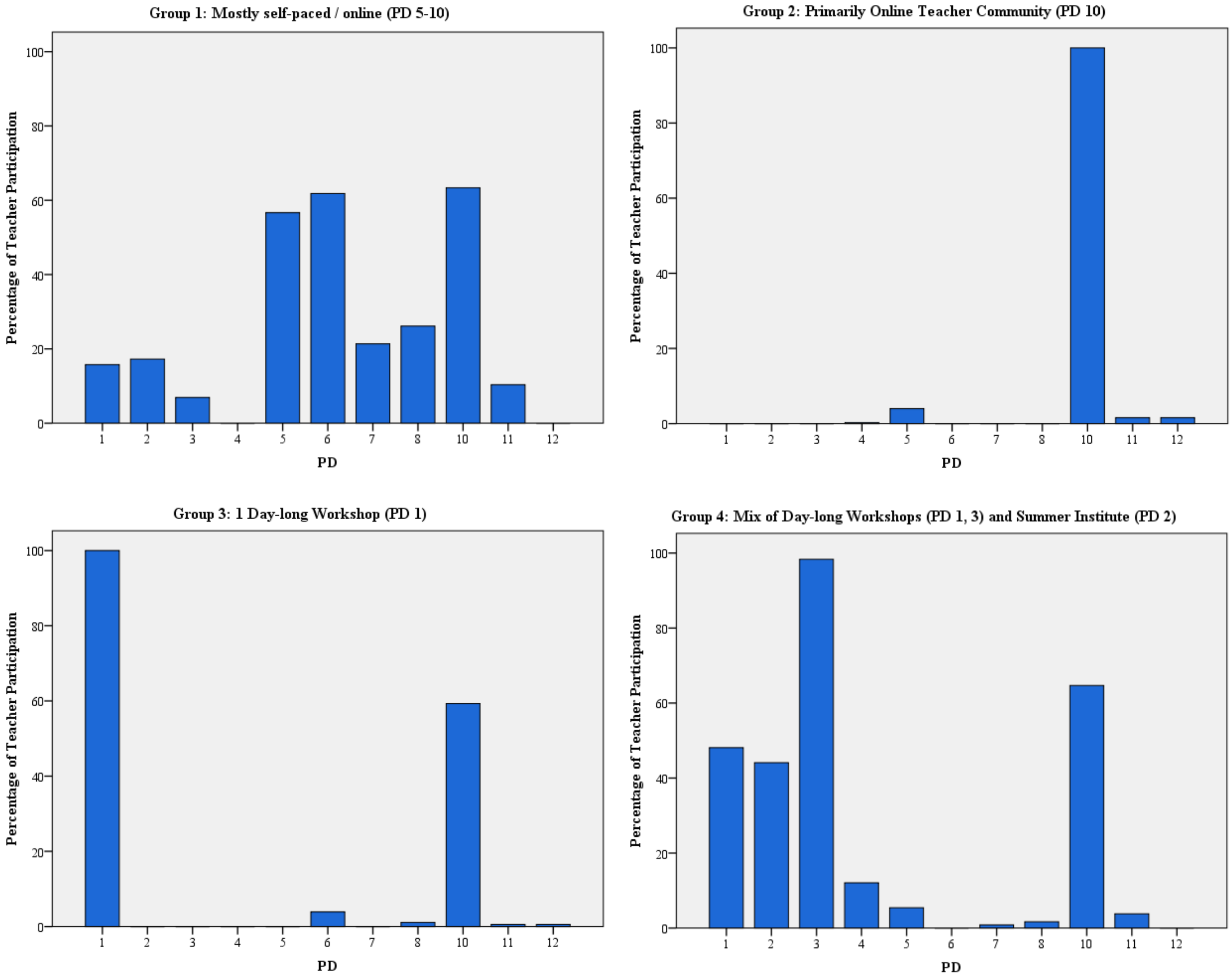
Table 6. Model fit indices for 1- through 7-class solutions

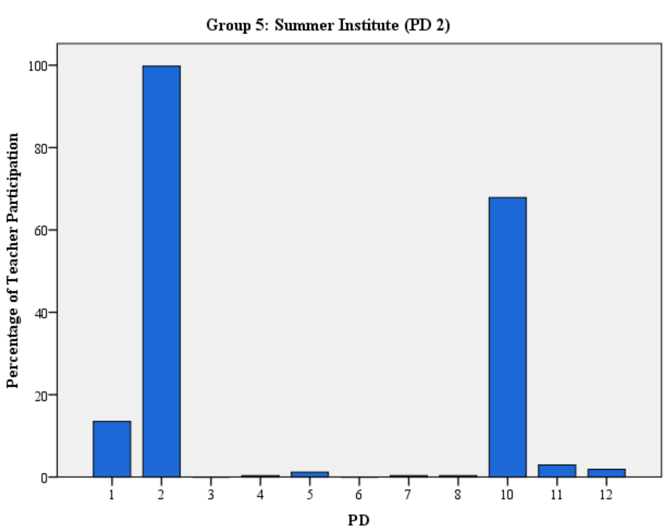
	1-class	**2-class**	**3-class**	**4-class**	**5-class**	**6-class**	**7-class**
Loglikelihood	-5442.573	-5309.730	-5213.675	-5109.391	**-5065.840**	-5040.568	-5028.516
BIC	10967.850	10792.388	10690.502	10572.157	**10575.277**	10614.957	10681.076
Entropy	-	.477	.639	.768	**.889**	.902	.894
LMR *p*-value (for k vs. k-1 classes)	-	0	0	0	**.001**	.0086	.3516

Combining the five classes of teachers derived from the LCA and the group of teachers who reported participating in no PD yielded a total of six PD participation groups. Within the "No PD" group (n = 212), teachers who reported that they were leaders of PD (n = 25) were excluded from subsequent analyses, leaving a total of 187 teachers. Specific participation patterns of each group are depicted graphically in Figure 2. Within Groups 1, 3, 4 and 5, roughly 60% of teachers participated in the online Teacher Community (PD10); all teachers in Group 2 participated. Although all groups included small numbers of participants across several PD options (e.g. in Group 2, less than 5% participation in PD4, 5, 11, 12), overall descriptions of each group were derived based on the PD participation of the majority of group members (e.g. in Group 2, 100% participation in PD10). Overall descriptions of PD participation for each group are provided below in Table 7.

Table 7. The five classes of teacher PD participation patterns.

Group	**n**	**Description**
0	187	No PD
1	91	Mostly self-paced & online (PD5-8)
2	457	Primarily online, AP Teacher Community (PD10)
3	182	1-day FtF Workshop (PD1)
4	244	Mix of 1-day FtF Workshop (PD1, PD3) and 4-5 day Summer Institute (PD2)
5	868	4-5 day Summer Institute (PD2)





Key to PD Activities

1	AP Biology (1 day, FtF, fall workshop)
2	AP Biology Summer Institute (4-5 days, FtF)
3	AP Biology: Transitioning to Inquiry-Based Labs (1 day, FtF)
4	Day With an AP Biology Reader (1 day, FtF)
5	Self-Paced Online Course: Transitioning to Inquiry-Based Labs (5-6 hours, Online)
6	Self-Paced Online Course: Introduction to AP Biology (5-6 hours, Online)
7	Self-Paced Online Course: AP Biology: Quantitative Skills (2 hours, Online)
8	Self-Paced Online Course: AP Biology: The Labs (2 hours, Online)
10	College Board Web Site: The AP Teacher Community
11	Laying the Foundation (LTF) Summer Institute (4-5 days, FtF)
12	BSCS & NABT AP Biology Leadership Academy (4-5 days, FtF)

Figure 2. Distribution of PD patterns in each of the 5 groups (PD9 is excluded from analyses).

Research Question 2. How are PD choices and patterns related to teacher characteristics, such as experience, gender, age, and specific concerns about teaching revised AP Biology?

We ran a series of one-way ANOVAs to test differences in teacher characteristics across the six groups (see Table 8), with Bonferroni post-hoc comparisons. First, to characterize these groups based on overall level of PD participation, a variable was computed to reflect an estimated total number of hours spent in PD. Because hours spent on the AP Central website or in the Online Teacher Community could not be estimated, these PD options were excluded from the calculation. Hours spent in FtF PD, online self-paced PD, and online web events were calculated by first estimating the hours for each PD (e.g. 8 hours for a 1-Day Workshop); for each teacher, these values were then summed for PDs in which they had participated. The groups varied significantly in the number of hours spent in PD. Each group was significantly different from each other group; from least hours to greatest was Group 0, Group 2, Group 3, Group 1, Group 4, Group 5.

Table 8. ANOVAs comparing teacher characteristics by PD group.

	Group 0 M (SD)	Group 1 M (SD)	Group 2 M (SD)	Group 3 M (SD)	Group 4 M (SD)	Group 5 M (SD)	F Test
Age	48.1 (11.2)	48.0 (11.2)	46.1 (10.2)	47.9 (9.86)	48.7 (10.4)	46.4 (10.1)	3.16**
Years Teaching AP Biology	7.05 (5.77)	7.03 (5.50)	7.07 (5.45)	7.05 (5.86)	6.98 (5.40)	5.80 (5.16)	5.21***
Proportion Female	.50 (.50)	.81 (.36)	.67 (.47)	.70 (.46)	.71 (.46)	.77 (.42)	10.7***
Importance of PD	3.89 (0.83)	4.35 (0.67)	4.24 (0.64)	4.20 (0.68)	4.40 (0.61)	4.48 (0.54)	30.3***
Challenge with AP Redesign Features	2.76 (0.83)	2.95 (0.69)	2.96 (0.70)	2.91 (0.72)	2.99 (0.71)	3.04 (0.69)	4.68***
Frequency of Implementing Redesign Features	3.23 (0.65)	3.45 (0.65)	3.33 (0.68)	3.23 (0.66)	3.46 (0.67)	3.30 (0.65)	4.07***
Total PD Hours	0.00 (0.00)	18.3 (18.8)	1.29 (6.13)	8.67 (4.01)	29.2 (20.8)	38.8 (8.19)	1056***

** $p < .01$, *** $p < .001$

The groups varied significantly in age; Group 2 was significantly younger than Group 4. The groups also varied significantly in years of experience teaching AP Biology; Group 4 had less experience than Group 2; Group 5 had less experience than Groups 2 and 4. In terms of gender, the groups also varied significantly; Group 0 had significantly more men than all other groups; Group 2 had more men than Group 5.

These groups of teachers also differed on their attitudes toward the value of PD, their reported level of challenge with teaching AP Biology, and their AP redesign practices. First, the groups differed significantly in their rating of how important they believed PD to be for their performance as teachers and their students' performance; Group 0 reported that PD was less important compared to all other groups; Groups 2 and 3 also reported that it was less important than Groups 4 and 5. The groups also differed significantly in their reported level of challenge in implementing AP redesign features (e.g. refer to the "Big Ideas" of Biology, having students report lab findings to other students); Group 0 reported less challenge than all other groups. Finally, the groups differed significantly in the reported frequency with which they employed AP redesign features. Group 4 used redesign features more frequently than Group 0, Group 3, and Group 5.

Research Question 3. What is the relationship between the PD patterns of teachers and their students' AP examination outcomes?
Multilevel structural equation modeling was conducted using MPlus software in order to evaluate direct and indirect associations between student AP scores and teacher characteristics, PD participation, and reported classroom practices. Multilevel modeling was employed in order to account for the nested structure of the data, with students (Level 1) nested within teachers (Level 2).

A series of multilevel path models were tested to evaluate direct effects on student AP scores. In all models, student PSAT scores were included as a Level-1 covariate and teacher years of experience teaching AP Biology were included as a Level-2 covariate. First, all individual PD choices were included as predictors of student scores. Only participation in the online Teacher Community, $B = 0.08$ (0.02), $p < .001$, and the AP Biology Leadership Academy, $B = 0.18$ (0.06), $p = .003$, had significant positive associations with student scores. Laying the Foundation Summer Institute had a significantly negative association with student scores, $B = -0.12$ (0.06), $p = .046$. Next, all the five PD groups, derived from the latent class analysis and converted into dummy variables, were used to predict student scores relative to the "No PD" group. Two groups were associated with higher student scores: Group 4 (Mix of 1-day FtF Workshop and 4-5 day Summer Institute), $B = 0.08$(0.04), $p = .032$ and Group 5 (Summer Institute), $B = 0.06$ (0.03), $p = .045$.

Finally, all redesign practices were tested as predictors of student scores. These included teachers' ratings of the challenges represented by revised curriculum practices listed in Table 3, as well as the number of labs they completed, the number of labs completed from the guide, and the number of student-generated labs completed. The only practices that were positively associated with student scores were the number of labs completed, $B = 0.07$ (0.02), $p < .001$, and the number of student-generated labs completed, $B = 0.02$ (0.01), $p = .019$. The frequency with which teachers provided guidance on test questions that integrate Biology content and scientific processes was negatively associated with student scores, $B = -0.04$ (0.01), $p = .003$.

The results of these tests of direct effects were used to formulate hypotheses about potential mediating effects on student scores. Based on the finding that PD Group 5 (most of whom participated in the Summer Institute) was associated with higher student scores, but the Summer Institute itself was not associated with scores, it was hypothesized that some other characteristics associated with this group may have been related to student scores. This group had relatively higher ratings of the importance of PD and also had the highest number of total hours spent in PD. Participation in the Online Teacher Community was also uniquely associated with student scores, although the PD group that primarily used this resource, to the exclusion of other PD options, did not have significantly higher scores relative to the No PD group. (This may be due to the attribution by survey participants of "membership" as equaling professional development whether or not they actively participated on a frequent basis.) For these reasons, beliefs about the importance of PD, hours in PD, participation in the Online Teacher Community, and the number of labs were selected as variables to include in a larger model. We hypothesized that rated importance of PD would be positively associated with greater number of hours spent in PD as well as in participation (or at least membership) in the Online Teacher Community, that these PD behaviors would be associated with completing more labs, and that the number of labs would be positively associated with student scores.

A path model (see Figure 3) was tested in which all hypothesized mediating paths were specified as well as direct effects on student scores. This model had good fit (RMSEA = .017, CFI = .998; Hu & Bentler, 1999). Rated importance of PD was a significant predictor of both hours spent in PD, B = 6.99 (0.74), $p < .001$, and participation in the online Teacher Community, B = 0.07 (0.02), $p < .001$. Number of hours in PD was a significant predictor of the number of labs completed, B = 0.003 (0.001), $p = .002$, as was participation in the Online Teacher Community, B = 0.17 (0.05), $p < .001$. Finally, the number of labs was a significant predictor of student scores, B = 0.07 (0.01), $p < .001$. Participation in the online Teacher Community had a significant direct effect on student scores, B = 0.09 (0.02), $p < .001$, but hours in PD did not, B = 0.00 (0.00), $p = .491$. Rated importance of PD had a small, but significantly negative direct effect on student scores, B = -0.05 (0.01), $p = .001$, suggesting that students whose teachers believe that PD is important but do not participate in PD have lower scores. This model is presented below with nonsignificant paths removed.

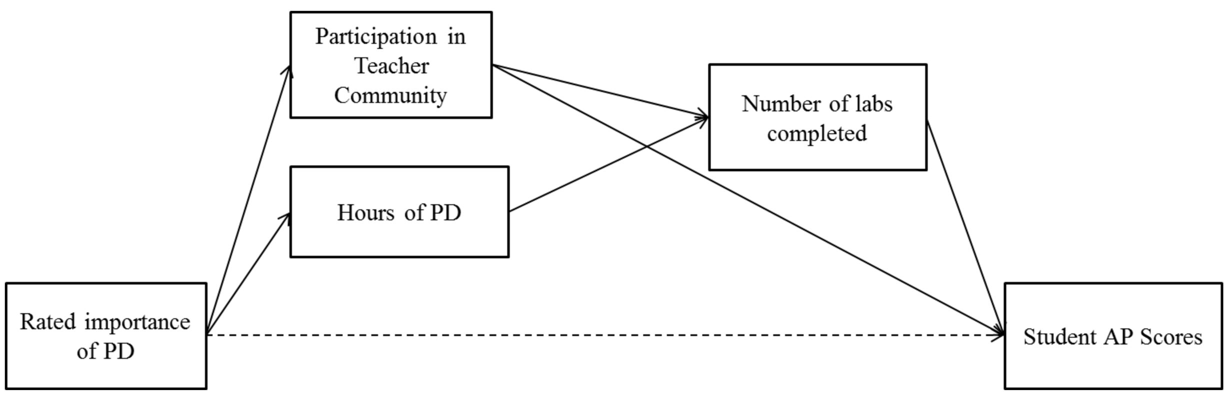


Figure 3. Path model showing relationship between teacher beliefs about PD, participation in PD, classroom practice, and student AP score outcomes. Solid lines represent significant positive effect. The dotted line represents a small, but significant, negative effect.

Discussion

In this paper, we examined data about teacher PD choices in the context of the revised AP Biology examination and curriculum. Given a major change in a high-stakes examination and the course designed to prepare students for that examination, what PD choices did teachers make in order to prepare? Are there recognizable patterns, in terms of the type, number, and combinations of PD activities, chosen by teachers (RQ1)? We found that yes, there are patterns. Using Latent Class Analysis to reduce the total combinations of PD to the most common combinations, we found that the "best fit" solution comprised five different combinations of PD, in addition to a group that participated in no PD. Are these patterns related to identifiable characteristics of teachers (RQ2)? Again, we found that yes, these patterns are related to teachers' individual characteristics. Differences that are salient include teachers' age, their years of experience teaching AP Biology, their gender, how valuable they generally feel PD is, what they found challenging about the AP redesign, and how frequently they reported implementing features of the AP redesign in their classroom. Are teachers' PD choices related to subsequent student performance on the AP Biology examination (RQ3)? This may be the most important question for educational policy, as students' performance on the AP Biology exam is the most direct measure of the "success" of teachers' efforts in preparing for and teaching the AP Biology curriculum. We found that, yes, some of the patterns of PD selected by teachers were predictive of student outcomes on the AP Biology exam, although effect sizes were small. In particular, it appears that teacher belief in the importance of PD, as mediated by participation in PD and use of laboratory activities, significantly predicts student AP exam scores, though the impact is limited. Although the number of hours of participation in PD was not directly related to student outcomes, it was indirectly related through teacher's use of laboratory activities. In contrast, participation in the online AP Teacher Community was directly related to teacher practice and student outcomes, though our understanding of "participation" lacks nuance using the data available to us. Overall, it appears that student achievement is more likely to be related to a spectrum of variables associated with teachers and their PD choices, rather than any individual factor or choice, which provides support for the dynamic model of teacher learning and change suggested by Opfer and Pedder (2011).

Limitations

There are several limitations in this work. Most notably, our sample was biased towards higher SES schools and higher performing students. This compressed variation in student outcome scores may have affected our results, and certainly means that our findings should not be used when considering PD options for teachers working with lower-SES or lower-performing students. Future research may help us to understand why our sample population has these biases. In terms of how we characterize PD in this study, our analyses were limited to the PD activity as a whole, as opposed to using the individual characteristics that make up the various components of each PD activity (e.g., Garet, et al., 2001) as predictors. It is therefore difficult to understand precisely what might mediate the relationship between teacher PD, classroom practice, and student outcomes. The model we selected has a good fit, but there may be other models that also fit our data that were not explored.

Future Work

An important component of this work is a series of teacher case studies, designed to illuminate the data from the surveys. Case study data gathering related to the data presented in this paper is underway, and we will be able to use data from these cases to make more nuanced interpretations of the survey data in future publications.

The data and findings reported in this paper are part of a longitudinal study of PD for the revised AP courses in Biology, Chemistry, and Physics. The AP Chemistry revisions will be introduced in the 2013-14 academic year, followed by AP Physics in the 2014-15 academic year. Our initial funding allows us to survey teachers about their PD participation related to AP Biology for three years, AP Chemistry for two years, and AP Physics for one year. Will teacher participation in PD remain constant across multiple years, or shift as teachers' experience grows and their felt needs with respect to the challenges of the AP curriculum shifts? As part of our ongoing work, we plan to increase our focus on how particular challenges of the AP revisions, as identified by teachers, are related to PD choices. And we will seek to develop more nuanced understandings of the characteristics of individual PD activities that are related to improved student outcomes.

Teacher participation in the online AP Teacher Community was a reliable predictor of student outcomes on the AP Biology exam. We therefore plan to delve more deeply into the nature of teacher participation in this online resource in future surveys and case studies. Is it important that teachers make frequent contributions to the online community? How valuable is “lurking,” where one regularly follows online conversations without participating actively? How often must one visit the online community in order to realize the benefits?

Finally, we intend to address the response bias in our survey population through several means. One is to seek to increase the participation of teachers from lower-SES schools, perhaps through different or more direct contact or through differential survey incentives. Another approach is to examine the data we *do* have from lower-SES schools, in order to see whether within those schools some students perform better on the AP exam than demographically comparable students. We might then be able to examine differences in professional development participation patterns among teachers in this subset of the larger study population.

Conclusion

The significance of this research is in its contribution to knowledge about the professional development choices teachers make in the context of large-scale mandated curriculum change. The finding that one can form distinct groups of teachers according to professional development choices, and the finding that professional development choices are related to the degree to which teachers are challenged by the AP redesign, together suggest that teachers might be aligning their participation in professional development with their perceived needs. This is a potentially important observation, suggesting that there is value in providing a range of professional development options to serve teachers with different needs. It is well-known that high-stakes examinations create leverage for change in teaching practice, sometimes with de-generative effects when tests are poorly designed (Pellegrino, Chudowsky, & Glaser, 2001). The revised AP science examinations represent a thoughtfully designed test, and the hope of the NRC and the College Board is that this test will drive positive changes in science teaching among a highly influential group of instructors – AP science teachers. We believe that a better understanding not just of what types of professional development are related to particular outcomes, but in what combination and for what types of teachers, will yield valuable information both for policy and practice towards the improvement of professional development offerings for all science teachers.

References

Achieve, Inc. (2013). *Next Generation Science Standards*. Achieve, Inc. Retrieved from http://www.nextgenscience.org/next-generation-science-standards

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

Desimone, L. (2009). Improving impact studies of teachers’ professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Desimone, L., & Carlson Le Floch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve survey results in education research. *Educational Evaluation and Policy Analysis*, *26*(1), 1–22.

Desimone, L., Garet, M. S., Birman, B. F., Porter, A., & Yoon, K. S. (2003). Improving teachers' in-service professional development in mathematics and science: The role of postsecondary institutions. *Educational Policy*, *17*(5), 613–649.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education*, *64*(5), 426–438. doi:10.1177/0022487113494413

Fishman, B., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, *19*(6), 643–658. doi:10.1016/S0742-051X(03)00059-3

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945.

Geiser, S., & Santelices, V. (2006). The role of Advanced Placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. L. Horn (Eds.), *Expanding opportunity in higher education: Leveraging promise* (pp. 75–114). Albany, NY: SUNY Press.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional Development Research Consensus, Crossroads, and Challenges. *Educational Researcher*, *42*(9), 476–487. doi:10.3102/0013189X13512674

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics*, *99*(5), 258–271.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67–101. doi:10.3102/10769986029001067

Mundry, S., Spector, B., Stiles, K., & Loucks-Horsley, S. (1999). *Working toward a continuum of professional learning experiences for teachers of science and mathematics*. Madison, WI: National Institute for Science Education, University of Wisconsin-Madison.

National Research Council. (2002). *Learning and understanding: Improving advanced study of mathematics and science in U.S. high schools: Report of the content panel for biology*. Washington, DC: Committee on Programs for Advanced Study of Mathematics and Science in American High Schools, National Research Council. Retrieved from http://www.nap.edu/catalog.php?record_id=10365

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: National Academies Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Nylund, K.L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535-569.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, *81*(3), 376–407. doi:10.3102/0034654311413609

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, *39*(5), 369–393.

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. doi:10.1016/j.edurev.2008.01.003

The College Board. (2014). *AP Central – Advanced Placement Scores, Courses & Exam Center*. Retrieved March 8, 2014, from http://apcentral.collegeboard.com/home