Early Lessons from Evaluation of Computer Science Teacher Professional Development in Google's CS4HS Program

Article *in* ACM Transactions on Computing Education · August 2017

DOI: 10.1145/3077617

4 authors, including:

Jason Ravitz
Digital Promise

Juliane Blazevski
Hypothesi Applied Research and Evaluation

Early Lessons from Evaluation of Computer Science Teacher Professional Development in Google's CS4HS Program

JASON RAVITZ, CHRIS STEPHENSON, and KAREN PARKER, Google
JULIANE BLAZEVSKI, Hypothesi

This article compares self-reported learning gains and experiences of teachers in four professional development courses funded through Google's 2014 Computer Science for High School program. The courses were designed and taught independently at four universities and started late enough in the year to participate in our pre-post study. Two of the courses used a face-to-face approach, one was online only, and one used a hybrid format. Analyses from 314 pre-surveys and 129 post-surveys indicate CS teachers are far from homogenous, suggesting that some customization may benefit professional development. We also saw a stronger sense of community in the two face-to-face courses. Among the outcomes we measured, teacher concerns (Hall and Hord 1977) were more sensitive to change than our measures of self-efficacy, outcome expectations, readiness, or beliefs. Findings illustrate the variety of CS teacher professional development experiences and the need to study the best ways to scale effective CS teacher education.

Categories and Subject Descriptors: K.3.2 [**Computer Milieux**]: Computer and Information Science Education—*Computer science education*

General Terms: Online Learning, Teacher Education, Computer Science

Additional Key Words and Phrases: K12 education, attitudes, hybrid learning

ACM Reference format:
Jason Ravitz, Chris Stephenson, Karen Parker, and Juliane Blazevski. 2017. Early Lessons from Evaluation of Computer Science Teacher Professional Development in Google's CS4HS Program. *ACM Trans. Comput. Educ.* 17, 4, Article 21 (August 2017), 16 pages.
https://doi.org/10.1145/3077617

1 INTRODUCTION

In this study, we examine four attempts to address the growing demand for scalable professional development for computer science (CS) teachers. This work complements similar studies undertaken by Goode et al. (2014) and Ericson et al. (2014) in the U.S., and Sentance et al. (2014) in the U.K., as well as major efforts by the National Science Foundation (Cuny 2012) and Code.org (Wilson 2013). It also follows on the highly valuable landscape study conducted by researchers at the University of Chicago (Century et al. 2013), which surfaced the need for a more coherent system of professional development for computer science teachers, improved alignment of professional development content and teacher needs, and an increased focus on teaching and learning in the classroom.

Authors' addresses: J. Ravitz, C. Stephenson, and K. Parker, 1600 Amphitheatre Parkway. Mountain View, CA 94043; emails: {ravitz, stephensonc, kparker}@google.com; J. Blazevski, 2232 S. Main Street, Suite 363 Ann Arbor, MI 48103; email: jblazevs@umich.edu.

ACM 1946-6226/2017/08-ART21 $15.00
https://doi.org/10.1145/3077617

2 RATIONALE/MOTIVATION

Any study of CS professional development for teachers must be seen in the wider context of CS in education. In 2005, Stephenson et al. noted that the growing dependence on and interactions with computing technology require the acknowledgement of CS as a core element of all Science, Technology, Engineering, Mathematics (STEM) education. This recognition led to a bipartisan movement for a congressional act, introduced by Congressman Lamar Smith (2015), to officially include CS as a STEM field. Recent studies by the Computing Research Association (2017) indicate that growing interest in computer science is driving significant enrollment increases at the post-secondary level, and legislation is currently moving through several states to ensure greater access to computer science courses in schools. Despite these advances, schools continue to face profound challenges in preparing a sufficient number of well-trained teachers to meet these growing needs. A study commissioned by Google (2015) revealed that competition with other subject areas and lack of teachers trained in CS were some of the major barriers to offering CS in schools despite growing demand among parents and others.

In addition to a lack of access to CS education in schools, CS education also faces troubling and persistent inequities. The disproportionate participation of women and underrepresented minority students in formal education has been well documented. Camp (1997), Crutchfield et al. (2011), Kekelis et al. (2005), and Margolis et al. (2008) all point to the continued underrepresentation of low-income students, underrepresented minorities, and especially girls in the discipline. While a national study from Google found that academic exposure is a key predictor of interest in CS (Google 2014, p. 9), data analysis from the College Board's Advanced Placement Computer Science course conducted by Ericson (2016) showed that while there was an overall increase of 17.3% in student test takers in 2016, there were only marginal increases in representation with test takers being 23% female, 3.7% African American, and 11.5% Hispanic, below the U.S. population rates according to the U.S. Census (2015).

This unequal representation extends beyond the classroom and into the workplace. According to the National Science Foundation (2012), the U.S. is facing severe labor supply shortages that will make it impossible to meet the growing need for highly skilled CS workers. Moreover, the lack of diversity in the field likely limits creativity in problem solving and product development (Page 2008). Data from the Taulbee Survey conducted by the Computing Research Association (2015) highlight the extent to which lack of participation by minorities and females continues to be a significant part of the problem, with a low proportions of CS bachelor degree recipients being female (16%), Black or African-American (4.6%), or Hispanic (8%).

Efforts to improve access to and participation in computer science education in K–12 have been further hampered by a lack of well-prepared teachers. According to the U.S. Department of Education Office of Postsecondary Education (2015), at least 17 states have reported CS teacher shortages, some back as far as 1990. The Gallup study commissioned by Google (2015) found that this continues to be a key issue. Addressing current teacher shortages raises the question of how the educational community can provide rigorous, relevant professional development for current and potential CS teachers on a national scale. There is a rich body of research on models of professional development for teachers and, as Koellner and Jacobs (2016) point out, these models range from highly adapted to highly specialized. In addition, ubiquitous access to computing tools among current teachers has led to consideration of multiple methods for instructional delivery including, face-to-face, video conferencing, online learning, blended instruction (involving a mix of face-to-face and online learning) (McConnell et al. 2013), and the development of communities of practice (also known as professional learning communities) (Fincher 2007; McConnell et al. 2013; Ni et al. 2011; Wenger et al. 2002).

The use of online courses for teacher professional development in CS is a fairly recent development, and, as a result, there is, as of yet, not a large body of research on its efficacy. Vivian et al. (2014) noted that the varied backgrounds of participants created a challenge "to provide educators new to the discipline area with the fundamental knowledge but also opportunities for educators who are more comfortable with the learning area to extend themselves." There are also well-documented concerns regarding low completion rates (Perna et al. 2013) and other issues related to online teaching (EPFL 2014). These opportunities and concerns led us to launch a study of the efficacy of Computer Science for High School (CS4HS) courses in late 2014. Although the study had limited data, they highlight key issues for scaling CS teacher professional development and challenges in conducting studies to be addressed going forward.

3 THE PROGRAM

The goal of the CS4HS program is to meet the growing demand for professional development for CS teachers and to learn about effective professional development practices. The program was based on a model developed and offered by Carnegie Mellon University in 2006. In 2007 and 2008, it was piloted at multiple universities with support from Google. In 2009, Google officially launched CS4HS as an annual program that provides grants to universities and nonprofit organizations offering multiple-day workshops for CS teachers. Under CS4HS, interested organizations submit applications via a central application process that are reviewed based on criteria laid out in the application documents. The criteria for 2014 included the following: applicants must be affiliated with a college, university, technical college, community college, or an official non-profit organization; workshops must have a clear CS focus and either be a legacy program with face-to-face model or a new course to be delivered through an online format. While the criteria have evolved over the years, the program has consistently enabled the applying organizations to determine the learning content and format.

When CS4HS launched, very little professional development was available for computer science teachers and there was little agreement or research as to the efficacy of content and mechanisms. Since 2009, CS4HS has funded workshops in 294 locations for more than 20,000 teachers globally and has continued to evolve as a way to meet changing learning needs and explore critical questions with regard to scaling (Gray et al. 2015). With growing interest in using online technologies to support knowledge generation and sharing, in 2014 CS4HS added funding for programs using an online format to help scale CS professional development to more U.S. teachers. CS4HS was seen as a viable venue for exploring different approaches to professional development, building on the strong relationship between Google and universities and between the universities and classroom teachers. These relationships provided a sense of trust and collaboration and helped create an environment for experimentation and openness that allowed the study to be conducted on a voluntary basis. Google's online efforts have involved multiple courses that were widely available and free (Mindell et al. 2014). While these were not truly massive, the participating universities were encouraged to experiment with different approaches for delivering courses.

The pressing need to solve the scalability challenge for teacher professional development in CS education led us to begin trying different formats for CS teachers in Google's CS4HS program. Starting with this study in late 2014 we began identifying the more effective courses for different audiences and their characteristics, believing this could inform research and program development. Our study was not designed to thoroughly test the efficacy of any particular course; however, we were interested in initial outcomes and evidence of effectiveness for the different courses, including those experimenting with different approaches.

4 RESEARCH QUESTIONS

The guiding questions for our study were as follows:

(1) Who participates and persists in each course?
(2) How do knowledge and attitudes change?
(3) How do changes in knowledge and attitudes vary by course?
(4) Do any course characteristics appear to be associated with different outcomes?

We used analyses driven by these questions to draw tentative conclusions about the impact of the courses for different participants and to explore implications for effective design.

5 RESEARCH METHODS

5.1 Selection of Participants

To conduct this study, CS4HS identified four professional development providers that had not yet begun their courses so they could provide pre-post measures. This excluded CS4HS courses that started earlier in the year while the study was still being approved. From the four courses we were able to study, we obtained data from over 300 teachers from March 2014 to November 2014. All four courses recruited participants using similar methods, including the following:

- Emailing teacher listservs (i.e., CSTA, AP CS, Math and Science Teachers),
- Emailing school principals and curriculum coordinators,
- Emailing participants from CS4HS programs offered in previous years, and
- Emailing other educational teaching and outreach programs at the given institution.

Courses varied in length from 3 to 4 days to 2 to 4 weeks, and content from the four courses covered the following areas:

- Fundamental knowledge and skills of CS,
- Beauty and Joy of Computing,
- App development,
- Computational logic and problem solving, and
- Programming logic using Scratch.

We were fortunate to obtain data from courses that used different delivery approaches, but this was a convenience sample, and methodologies were not entirely consistent. In addition, funding for programs was provided as an unrestricted gift, so Google had limited ability to require consistent data collection and reporting, relying instead on each site to collect the data in a way that worked for them. As a result of this individualized data collection and analysis, two courses with online components sent the post-survey to all pre-survey participants, while the two face-to-face courses only included participants who continued to the end of the course. Because these factors could deflate post-survey scores for the Online and Hybrid courses, if non-completers filled out the survey, we are cautious to mention these and other differences in our case studies and we avoid generalizing beyond the participating courses.

5.2 Measures

We used our survey instrument to collect participant data for each course on demographics (age, gender), years teaching, and experience learning and teaching CS. We used pre-post measures based on the Stages of Concern approach developed by Hall and Hord (1987), who posited that concerns about an innovation often progress from earlier stages that tend to be informational and personal, such as whether they are interested and have sufficient knowledge, to managerial, such as what is needed to implement effectively (in our case preparing to teach CS and being able to assess students), and then to later stages involving impact and improvement, such as how the innovation improves learning and can itself be improved. A current summary of this framework is available from American Institutes for Research (2017). The study also addressed constructs such as self-efficacy in teaching CS, identified by Bandura (1997), and expectations, beliefs, and readiness using measures for self-efficacy, expectancy, and task value based on English (2013).

Before launching this study, we piloted the survey in two courses and used participants' feedback and "other" responses, as well as reliability and factor analyses to improve our measures. Although each set of items produced a reliable index (alpha >.75 or better), given the exploratory nature of this study, we focus as much on individual items that were sensitive to treatment as the scale scores.

We administered the survey to participants before and after each course. Using data from the 95% of the participants who were teachers, we analyzed 258 pre-surveys and 129 post-surveys. Because the pre- and post- responses were unmatched, we can only compare the overall responses on the surveys. To address the validity of these comparisons, we looked for systematic differences in backgrounds and demographics of respondents that might bias the results (e.g., if only those with extensive CS teaching experience completed the post-survey). In fact, we found very little systematic difference, as shown in Figure 1. Finally, survey completion was a condition for entering or completing the courses, so they represent a census of those who started and finished. We do not have response rates to address how many declined to participate or were unreachable. Instead, we focus on attrition rates and whether those who completed the post-surveys were similar in numbers and characteristics to those who took the pre-surveys.

5.3 Participant Demographics and Background

Overall, we found participants were experienced teachers with more than half between the ages of 40 and 59 years of age. This demographic is consistent with the 2013 CSTA Computer Science in High Schools Survey which indicated that 77% of CS teachers are 41 years of age or older (CSTA 2013). Around 80% of our participants had six years or more teaching experience, and over half had 11 or more years. About 80% taught in public schools, mostly at the high school level (60%, with an additional 10–15% teaching high school but also other grades). There were slightly more female (58% pre, 52% post) than male participants. A large majority (83%) was Caucasian.

There was a mix of CS experience. Just fewer than 50% had already taught a full CS course, 20–25% had a degree in CS, and 75% had at least some CS-focused professional development (with over 50% reporting a week or more). A majority (70%) of the participants reported that they had at least a little experience teaching CS or other IT-related subject area, and almost everyone else had taught other STEM subjects. Although a majority reported some experience *learning* CS, only a few (30%) reported "much" or "very much" experience *teaching* CS. When asked, "Why are you interested in taking this course," there was a good mix. About 50% of teachers had chosen to teach CS, approximately 20% had been assigned, and approximately 25% were thinking about teaching CS.

5.4 Comparing Demographics on Pre- and Post-surveys

Although the sample size was substantially larger for the pre-survey, participant demographics on the post-survey were very similar. This supports the validity of pre-post comparisons in the absence of matched data. Figure 1 shows that the average years teaching was very similar. In addition, Appendix A shows that gender, ethnicity, school type, and years of teaching were similar (within

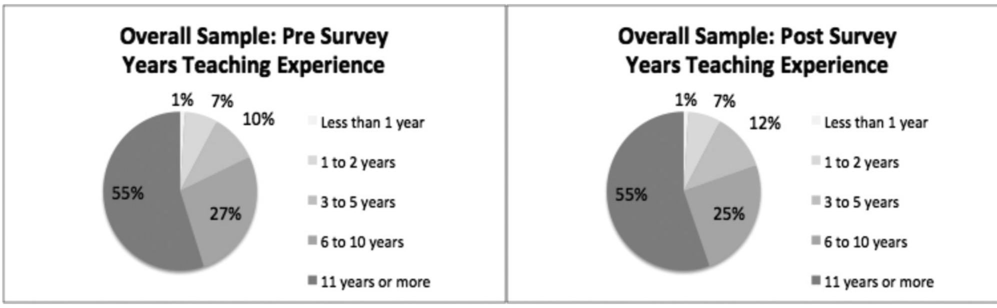

Fig. 1. Comparison of years teaching experience on pre-post surveys.

just a few percentage points) between pre- and post-survey participants. Additional comparisons suggesting the pre-post participants were similar are shown in Appendix B. For example, reasons for taking the course were rated in the same order and were reported with the same frequency (within 3%) between pre and post.

There were a few small differences. Appendix A shows the percent of teachers with degrees in CS rose from 20% on the pre-survey to 26% on the post-survey. In addition, the proportion of females dropped from 58% to 52%, and, consistent with findings from Haywood (2014), the percentage of 40-somethings dropped while the percent of 50-somethings went up.

5.5 Analysis

Mean differences between pre- and post-survey scores were analyzed using Analysis of Variance (ANOVA), with effect sizes using Cohen's D computed from the resulting means and standard deviations. Although there is some debate, scholars including Norman (2010) and De Winter et al. (2010) have supported using parametric statistics like these with Likert-style items. When we examined non-parametric differences through the Mann-Whitney analysis using the approach described by Fritz et al. (2012), all of the positive findings remained statistically significant. The one negative finding from the parametric analysis ("I can cover the material within the required timeframe") was no longer statistically significant. Effect sizes for the non-parametric analysis tended to be slightly smaller than those generated from the parametric analysis; however, the difference never exceeded an absolute value of .10 and on average only differed by .03. Because these differences were so minor, and because it is more intuitive to talk about means than ranks, we report ANOVA results with Cohen's D, and we provide non-parametric analyses in the appendices. For survey items where we examined percentages of categorical responses, chi-square analyses were utilized. The retrospective item on CS knowledge gain used a paired sample *t*-test. All analyses were conducted using SPSS software.

6 OUTCOMES

6.1 Self-reported CS Knowledge Gains

One of the post-survey questions asked participants to rate their overall knowledge of CS before and after the course on a five-point scale, retrospectively assessing their own growth in knowledge. At the end of the course, 66% said they had a high or very high level of CS knowledge, while only 22% felt this looking back at themselves before the course. There was a significant increase in mean responses (ES = 1.25) in self-rated knowledge from before the course (M = 2.8, SD = .99) to after the course (M = 3.9, SD = .75), $t(125) = 14.073$, $p < .001$. Although this item may be more valid as

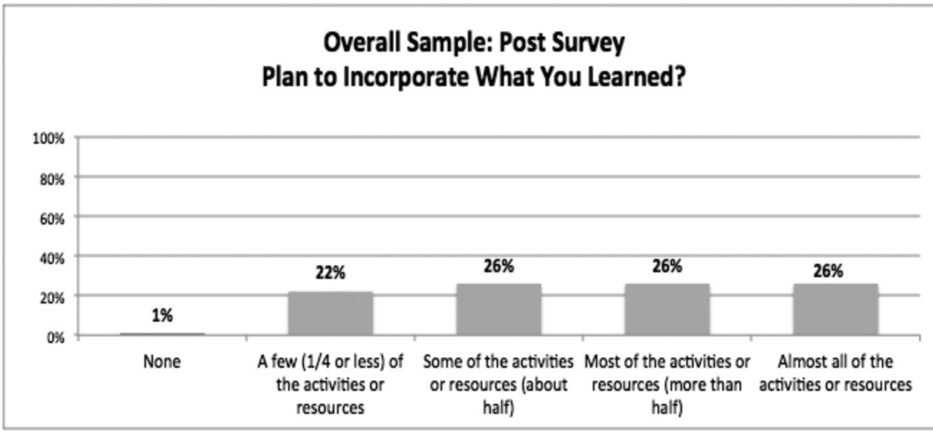

Fig. 2. Anticipated usefulness.

an attitudinal/motivational measure than a cognitive outcome (Sitzmann et al. 2010), it is still an indication of how the course experience varied for participants.

6.2 Intention to Use

Another outcome can be seen in participants' responses to the question regarding how many activities or resources from the course they planned to incorporate into their curriculum. Based on the post-survey responses, Figure 2 shows that over half of the participants (52%) planned to incorporate "most" or "almost all" of the activities or resources presented in the course, and 78% indicated at least some plans to use the resources that were provided in the course.

6.3 Pre-post Changes in Concerns and Other Attitudinal Measures

Both the pre- and post-surveys asked participants to assess their concerns about and interest in teaching CS, self-efficacy for teaching CS, expectations regarding the potential impact on students, beliefs about CS teaching (and its importance), and readiness for teaching CS. Appendix C provides the prompts for each of the items.

Overall, we see statistically significant pre to post differences. Table 1 lists each index score with its associated items, with mean scores and changes from pre- to post- using ANOVA (*t*-test), standard deviations, and effect sizes (based on the pre-survey standard deviation). Full analyses with standard deviations and F-values are shown in Appendix D, with similar non-parametric results shown in Appendix E.

Ratings of concerns about teaching CS on the overall index were significantly lower (ES = −.56) at the post-survey (M = 2.24, SD = .66) than they were at the pre-survey (M = 2.59, SD = .55), F(1, 432) = 29.98, indicating that these concerns were significantly lower after the courses. This was the case for all but one item ("deciding whether I want to teach CS"), which started and ended as the lowest-rated concern.

There was not a significant pre-to-post difference on the overall self-efficacy scale, but a couple of items did show gains: "I can teach the concepts required by the curriculum" and "I can effectively teach all students." There was also a significant decrease for "I can cover the material within the required timeframe," perhaps due to a more informed understanding of the required curriculum by the end of the course. For outcome expectation measures, the overall mean difference was statistically significant, based on the strength of two items "My students will be highly engaged"

Table 1. Changes in Pre-Post Measures

	Pre (n=314)	Post (n=124)	Diff	Effect Size	Sig.
Concerns Scale (I am concerned about...)	**2.6**	**2.2**	**−.4**	**−.56**	**.00**
deciding whether I want to teach CS	1.6	1.5	−.1	−.12	.3
understanding what CS is	2.1	1.7	−.4	−.44	0
understanding what teaching CS requires	2.6	2.1	−.5	−.5	0
assessing my ability to teach CS	2.4	2.2	−.2	−.26	.01
preparing to teach my course	2.8	2.5	−.3	−.31	0
finding out what students need to know	3	2.5	−.5	−.56	0
improving student learning outcomes	3.1	2.7	−.4	−.45	0
improving how I teach CS	3.1	2.7	−.4	−.42	0
working to improve how CS is taught	2.5	2.2	−.2	−.26	.02
Self-efficacy Scale (I can...)	**3.9**	**4.0**	**.1**	**.19**	**.09**
assess my students' learning and performance	3.9	4.0	.1	.13	.21
effectively teach all students	3.9	4.2	.3	.36	.00
teach the concepts required by the curriculum	3.7	4.1	.4	.50	.00
cover the material within the required timeframe	4.0	3.8	−.2	−.22	.05
motivate all or most of my students	4.1	4.0	−.1	−.15	.20
interest my students in CS	4.0	4.1	.2	.22	.05
Expectations Scale (My students...)	**3.9**	**4.1**	**.2**	**.33**	**.00**
will be highly engaged	3.8	4.1	.2	.34	.00
will be interested in CS degrees or careers	3.3	3.6	.3	.37	.00
will learn important skills	4.3	4.4	.1	.13	.24
will have better opportunities	4.3	4.4	.1	.10	.39
Beliefs Scale (Teaching CS...)	**4.4**	**4.5**	**.1**	**.15**	**.18**
is highly appealing to me	4.4	4.5	.1	.10	.36
is something I will enjoy	4.4	4.5	.1	.15	.18
will be rewarding and worth the effort	4.5	4.5	.0	.08	.42
will prepare my students for the future	4.7	4.8	.1	.16	.14
is my responsibility as an educator	4.2	4.3	.1	.10	.34
Readiness Scale	**3.5**	**3.5**	**.0**	**.02**	**.88**
My classroom time is sufficient	3.5	3.4	−.1	−.13	.18
My planning time is sufficient	3.2	3.0	−.2	−.16	.12
My curriculum will allow me to teach CS effectively	3.6	3.6	.1	.06	.61
I have support from a PLC beyond my school	3.4	3.7	.3	.27	.01
I have adequate access to the facilities, technology and materials	3.7	3.7	.1	.06	.56

Note: See Appendix D for full analyses and Appendix E for the non-parametric comparisons.

and "My students will be interested in CS degrees or careers." In contrast, the beliefs scale such as "teaching CS is something I will enjoy" or "I will prepare my students for future careers" did not exhibit changes, nor did any of the readiness items except "I have support from a PLC beyond my school," which was moderately correlated ($r = .26$, $p < .01$) to whether a sense of community was reported in the course.

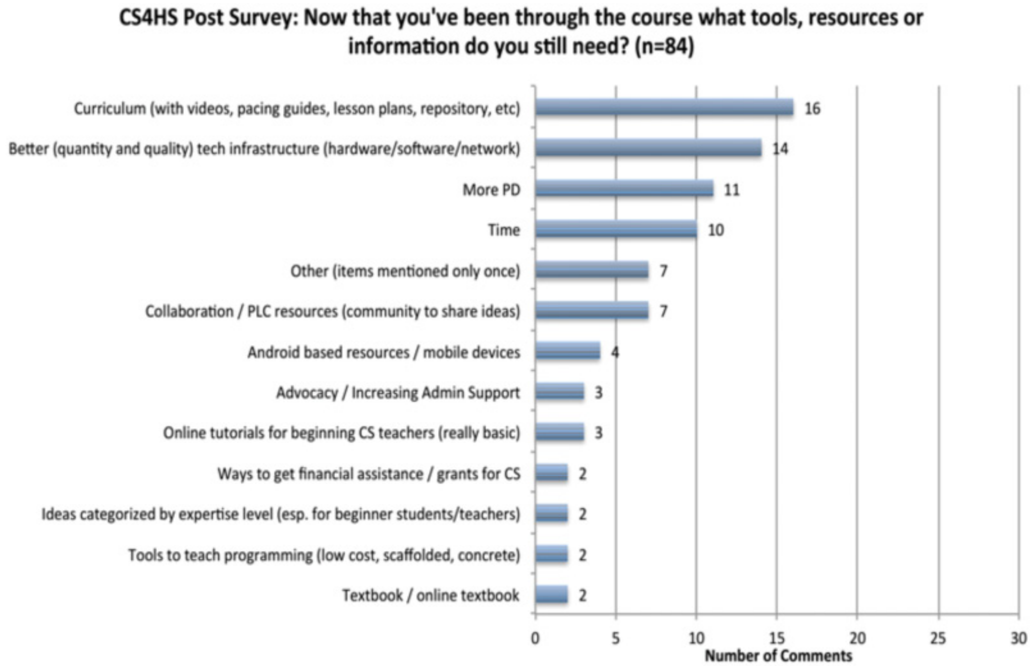

Fig. 3. Summary of qualitative responses about needed resources.

6.4 Ongoing Needs for Support

To identify needs for additional support, we asked an open-ended question about what tools, support, or resources were still needed after the course. Informal analysis of open-ended responses (coded by one of the authors and summarized in Figure 3) listed the types of resources that teachers wanted at end of the courses. Coding used an informal but iterative process to create the smallest number of categories.

Each comment was first broken down into distinct codeable segments (e.g., a single comment could list more than one tool or resource). Each segment was assigned a code. Identical or nearly identical content received the same code. After the initial round of coding, related codes were combined to create a smaller, yet still distinct, set of categories. Codes with a frequency of “1” were combined into the “other” category. It appeared teachers continued to want more classroom-relevant content. In addition, they identified access to hardware as a critical issue, a finding that is consistent with Phillips and Stephenson (2013), who found lack of adequate hardware access for CS courses can be a barrier to CS, even as computers may be available and used for other purposes in the school (e.g., testing).

7 EXPLORATORY CASE STUDIES

7.1 Comparing Primary Outcomes by Course

To learn more from our study, we disaggregated results and analyzed how the implementation of CS4HS differed for each course, including how the populations in these courses differed, had different rates of attrition and retention, and changed attitudes from pre-post. In addition to changes in self-reported knowledge and attitudes, key points of comparison included the

Table 2. Comparison of Post-survey Outcome Measures by Course

	A FTF1 (n=22)	B Hybrid (n=40)	C FTF2 (n=29)	D Online (n=30)
Response rate on post-survey (post/pre)	22/28 (79%)	42/70 (60%)	30/45 (67%)	31/137 (23%)
Sense of Community among Participants				
Tend to agree or Strongly agree	**77%**	38%	**73%**	47%
Recommend to Others				
Yes	100%	95%	96%	**63%**
Incorporate into your curriculum				
A few (1/4 or less) of the activities or resources	5%	17%	21%	**40%**
Some of the activities or resources (about half)	**50%**	17%	17%	27%
Most of the activities or resources (more than half)	32%	13%	**48%**	20%
Almost all of the activities or resources	14%	**50%**	14%	13%
Pre-Post Attitude Changes (Effect Sizes)				
Concerns	−.24	**−.34+**	**−.83***	**−.67***
Self-Efficacy	.04	**.34+**	.05	.12
Expectations	.19	**.46***	.19	.21
Beliefs	.03	.20	**.53***	−.30
Readiness	−.02	.19	.16	−.02
Retrospective CS Knowledge Gain (Effect Size)				
	1.19**	1.25**	**2.12** **	1.07**

+ $p < .10$, * $p < .01$; ** $p < .001$ See Table 3 for ANOVA statistics for these results.

overall sense of community reported, whether the course would be recommended to others, and how much of the content participants planned to use in their teaching.

These findings are viewed as exploratory. We cannot presume the results will generalize to other courses or that our survey data fully represent what happened in the courses. The results of these exploratory analyses, summarized in Table 2, illustrate how much variation is evident even in just four courses.

There were statistically significant differences in the percent of respondents agreeing or strongly agreeing that the course had a sense of community, X2 (3, 120) = 13.13, $p < .005$, and affirming that they would recommend the course to others, X2 (3, 119) = 24.31, $p < .001$ The sense of community reported by teachers appeared to be stronger in the courses that had a face-to-face format (70%+), while Course D (Online) had lower recommendation ratings (63%) than the other courses.

There were also statistically significant differences (Chi-Sq = 39.74, N = 121, df = 12, $p < .001$) regarding intention to incorporate activities or resources into the curriculum. Within each category, the difference between the most- and least-frequent response was statistically significant ($p < .05$, ANOVA with posthoc Bonferroni adjustments). For example, Course D (Online) reported the least intention to use (40% "1/4 or less") and Course B (Hybrid) reported the most (50% "almost all").

For the pre-post attitude changes, Course C (FTF2) showed the largest decrease in concerns (ES = −.83) and an increase in purported beliefs (ES = .53). Course D (Online) also showed a decrease in reported concerns (ES = −.67), while Course B (Hybrid) participants reported a

Table 3. Pre-post Changes in Attitudes by Course, ANOVA Analyses

		Pre			Post					Effect Size
Course	**Attitudes**	Mean	N	SD	Mean	N	SD	df	F-value	(Cohen's D)
A	Concerns	2.48	26	.60	2.33	21	.65	45	0.69	−.24
(FTF1)	Self-Efficacy	4.10	26	.56	4.12	22	.49	46	0.02	.04
	Expectations	4.11	25	.48	4.19	22	.41	45	0.40	.19
	Beliefs	4.66	26	.39	4.67	22	.37	46	0.01	.03
	Readiness	3.91	26	.73	3.89	22	.70	46	0.01	−.02
B	Concerns	2.52	65	.51	2.34	40	.61	103	2.91	−.34+
(Hybrid)	Self-Efficacy	3.98	64	.51	4.16	40	.52	102	2.91	.34+
	Expectations	3.98	64	.46	4.18	39	.42	101	4.99	.46*
	Beliefs	4.55	65	.69	4.67	40	.50	103	0.86	.20
	Readiness	3.45	65	.79	3.60	40	.75	103	0.84	.19
C	Concerns	2.48	44	.53	2.01	29	.59	71	12.36	−.83*
(FTF2)	Self-Efficacy	3.83	43	.82	3.86	29	.53	70	0.03	.05
	Expectations	3.82	43	.61	3.94	29	.68	70	0.63	.19
	Beliefs	4.20	43	.60	4.47	29	.39	70	4.30	.53*
	Readiness	2.99	44	.78	3.11	29	.76	71	0.44	.16
D	Concerns	2.71	123	.56	2.28	30	.71	151	12.4	−.67*
(Online)	Self-Efficacy	3.89	121	.59	3.97	30	.60	149	0.36	.12
	Expectations	3.94	119	.55	4.05	30	.52	147	1.04	.21
	Beliefs	4.37	122	.52	4.22	30	.47	150	2.04	−.03
	Readiness	3.47	122	.71	3.41	30	.72	150	0.20	−.02

** $p < .001$, * $p < .01$, + $p < .10$.

Table 4. Retrospective Knowledge Gains by Course, Paired Sample *t*-tests

Course	Pre N	Pre SD	Post N	Post SD	t-value	Effect Size (Cohen's D)	p <
A (FTF1)	21	.79	21	.59	5.46	1.19	.001
B (Hybrid)	39	1.06	39	.86	7.81	1.25	.001
C (FTF2)	29	.82	29	.78	11.40	**2.12**	.001
D (Online)	30	.79	30	.62	5.89	1.07	.001

significant increase in outcome expectations (ES = .46) and marginally significant changes in self-efficacy (ES = .34) and concerns (ES = −.34). Table 3 provides ANOVA statistics for these findings.

Additional details on attitudes changes by course are provided in Appendix F. Finally, all of the courses saw statistically significant gains in self-reported knowledge, especially Course C (ES = 2.12), based on paired sample *t*-test (Table 4).

These exploratory findings must be taken with caution, in part because Courses A and C (with their face-to-face formats) only conducted post-surveys with participants who completed the course. In contrast, Courses B and D included online participation and sent post-surveys to all pre-survey participants, allowing course non-completers to respond in a way that might deflate post-survey averages. This was particularly a concern in Course D, where the number of participants in the post-survey was less than a quarter (22%) of those who participated in the pre-survey (Table 2).

Additional exploratory comparisons concern differences in teacher background (Appendix G) and baseline differences in attitudes (Appendix H). Not surprisingly, Course A's (FTF1) participants, who had considerable CS teaching experience, were less concerned about whether they wanted to teach CS and were more concerned about how to teach CS better and improve learning outcomes. Course C (FTF2) had the fewest experienced CS teachers (a large proportion of middle-school STEM teachers) and had more teachers at earlier "stages of concern" (Hall and Hord 1987), still learning what CS is. Course D (Online) had sizable minorities (40%+) of both experienced and less experienced CS teachers and, with these widely varied backgrounds, also had the highest mean score for concerns overall.

Finally, concerning retention patterns (shown in Appendix I), all of the courses more frequently retained teachers with more PD or CS experience and lost teachers with less experience, especially Course D (Online). Course C (FTF2) seemed best suited for those who had not taught CS yet (the percentage increased by +11%) but still had −22% attrition if teachers had no prior PD at all.

7.2 Summary Findings from Case Studies

In summary, Course A (FTF1) had a much higher proportion of teachers who had already taught a full CS course (73%) and who had a degree in CS (46% on the post survey). Despite a very strong sense of community and significant gains in self-reported knowledge, other attitudes did not shift for participants in this course

In contrast, Course B (Hybrid) also had numerous teachers with CS experience (Appendix G). Despite not reporting as strong a community of practice as other courses, this course exhibited pre-post differences in concerns and was the only course with statistically significant changes in expectations and self-efficacy measures. This course also had the highest proportion (50%) intending to use "almost all or all of the activities or resources."

Course C (FTF2) used a face-to-face approach and included many middle school STEM teachers (about 50%), and only a few with any experience teaching CS-related subjects. This course demonstrated a strong sense of community with 73% agreeing that "I felt a sense of community among participants of this workshop." It also showed the largest pre-post gains in knowledge (moving from 2.2 to 4.0 on a five-point scale). These results indicate that the greatest knowledge benefit was perceived where teachers had least experience and perhaps the most to learn.

In Course D (Online), fewer than half of the participants agreed there was a sense of community, and only one-third planned to incorporate most or all of the activities or resources. Fewer teachers (63%) would recommend this course compared to over 95% in the other courses. Despite this less-enthusiastic endorsement, the course reduced participants' concerns about teaching CS more than any other, and there was still a significant increase in self-reported CS knowledge gains. Moreover, as indicated previously, participants who did not complete the course may have completed post-surveys, and this could be deflate post-responses more than in the other courses.

8 DISCUSSION

The above summaries draw from surveys administered voluntarily using protocols determined by each course. Our program could not require identical methods or additional qualitative data to be provided because of the nature of the funding we provided. However, based on the information that is available, we see evidence of the complexity of implementing and studying CS teacher professional development across varied contexts.

It is difficult to interpret the outcomes of individual courses without knowing more about the participants, qualitative data about how they interacted with the content, what they did after the course, the pedagogical approaches of the instructor, the quality of the materials, the work that teachers and students produced, and so on. However, our study highlights challenges in online CS teacher education and challenges that arise when studying outcomes. For example, the relative novice CS teachers in Course C (FTF2) reported particularly large knowledge gains, and changes in concerns and beliefs, while Course A (FTF1) with teachers who started with more CS experience self-reported smaller growth in knowledge and none of the other attitude changes. This suggests that while providing professional development opportunities for beginning CS teachers is important, there could be a use for advanced courses that are customizable to the needs of experienced CS educators.

One of our hopes was to see in which courses participants reported a strong sense of community and how this might relate to outcomes. It appeared that the face-to-face courses had the strongest sense of community, but we cannot conclude there is a relationship to course format or outcomes. Instead, we note that it is difficult to tease apart community of practice outcomes from the overall course design and delivery, and other factors that would predict outcomes (e.g., why the teachers enrolled in the course). It would be helpful to see more evidence that a sense of community is important for outcomes for both novice and experienced CS teachers, including in what ways CS teachers who have a sense of community are more effective than their peers, for example, Ni et al. (2011).

Although we were limited to the surveys we collected, a follow-up effort to collect more qualitative data might help us understand better, for example, why Course A (FTF1) with its strong sense of community and large self-reported knowledge gains did not results in larger pre-post attitude changes (Table 1) or how the online-only course (Course D-Online) served teachers with widely varying experience levels and still reported a strong sense of community compared to Course B (Hybrid) with its face-to-face component and relatively strong outcomes for efficacy, expectations, and intention to use. These few cases make clear how much we have to learn about how community does or does not contribute to teacher learning outcomes across professional development courses. In future studies, we hope to use matched pre-post data and qualitative data to identify and help us tell the stories of teachers who completed and did not complete the courses, including female teachers and others who appear to complete at slightly lower rates. We agree with Vivian et al. (2014) who noted that "Understanding the motivations and experience of those who did not complete the last module may provide insight into how we can improve." In future research, we hope to collect more qualitative data and include a question about the extent of course completion on the post-survey to rule out the chance of non-completers negatively biasing the post-survey findings.

This study tried using a variety of conceptual frameworks to evaluate outcomes. The most useful set of measures for detecting differences, being sensitive to change over time and between courses (Table 2), were our measures of concerns about teaching CS that were based on stages developed by Hall and Hord (1987). A few other items detected differences, but overall the self-efficacy and expectations measures were less sensitive, while beliefs and readiness items were rarely if ever worthwhile for drawing inferences about the efficacy of the different courses. It seems less worthwhile, therefore, to focus on abstract self-perceptions, or "high inference" items (Babad 1996; Chavez 1984) and more worthwhile to focus on immediate issues that teachers face and actions they can take. The efficacy item "My curriculum will allow me to teach CS effectively" seems to have countless dependencies a teacher must judge about his or her own teaching, the curriculum, and students. Meanwhile, "I am concerned about preparing to teach my course" seems like a fairly direct indicator of teachers' sense of preparedness, even if the reasons are not specified.

Additionally, the concerns framework appears to have broad utility, because participants in Course D (Online) exhibited large pre- and post-survey differences (the largest overall change in concerns) despite entering the course with widely different levels of CS teaching experiences. Given the need for more research on teacher professional development in CS education, we recommend looking for additional frameworks that can help assess change, building on measures from other CS-focused studies (e.g., McKlin et al. 2013; Wiebe et al. 2003) to help inform and improve teachers' experiences.

We would want to study larger numbers of courses and multiple cohorts before drawing conclusions about the impact of the course format, but this study helps us move in that direction. Partially as a result of these findings, the CS4HS grant criteria for 2015 were expanded to keep exploring different formats and to create a new emphasis on supporting stronger communities of practice. It is expected that future iterations of the program will bring additional revisions based on data from 2015–2016 grantees and future sessions, as well as an evaluation of the long-term impact of the CS4HS program. We hope the lessons learned will help future programs address the needs of CS educators and conduct more useful studies.

To conclude, this exploratory study examined four professional development programs carried out independently and using different approaches, under the auspices of Google's CS4HS program. It examined differences in teacher attitudes in relation to the program contexts and the characteristics of the teacher participants. Although the goal was to inform the CS4HS program, the study raises broader questions about how professional development outcomes vary for teachers with different backgrounds and working in different contexts. The variations in outcomes suggest possible strengths and weaknesses of professional development courses using different online and face-to-face approaches. Although these results are exploratory, they highlight worthwhile questions and challenges for future research on how teachers experience different kinds of professional development.

The study contributes conceptual frameworks relating to how effective professional development might be structured and how programs might be designed or modified to better support the learning needs of different populations of teachers. The study also provides useful information about different attitude measures that can reveal changes over time, both across and within programs. The evidence that teachers' concerns were among the most sensitive measures for exploring differences across sites and at different points in time may be of use to future researchers. Finally, the study suggests practical implications for curriculum and professional developers, including the need for awareness of teachers' needs in online courses and the challenge of providing learning opportunities for diverse participants.

The results of our preliminary study suggest different models and formats of professional development could impact learning outcomes in ways that vary depending on the participants and context. These results could inform future learning designers and researchers, who can benefit from recognizing that delivery methods, including face to face, blended, and online, are not experienced in just one way and could have benefits for some and disadvantages for others. It appears that attitudes and learning outcomes may interact with delivery mechanisms, contexts in which participants work, and the characteristics of the participants. Further studies could provide greater understanding of the interplay between these factors. By raising these issues and suggesting useful outcome measures, this article opens areas for investigation and contributes to the growing and much-needed body of research into CS education professional development.

ELECTRONIC APPENDIX

The electronic appendix for this article can be accessed in the ACM Digital Library.

ACKNOWLEDGMENTS

The authors acknowledge the helpful suggestions from Julia Wilkowski and Cameron Fadjo, who provided valuable input as early reviewers of our work, as well as critical feedback from the editors and anonymous reviewers.

REFERENCES

American Institutes for Research. 2017. *SEDL Archive: Stages of Concern*. Retrieved February 25, 2017 from https://www.sedl.org/cbam/stages_of_concern.html.

Elisha Babad. 1996. How high is high inference? Within classroom differences in students' perceptions of classroom interaction. *J. Classroom Interact.* 31, 1 (1996), 1–9.

Albert Bandura. 1997. *Self-Efficacy: The Exercise of Control*. Freeman, New York, NY.

Tracy Camp. 1997. The incredible shrinking pipeline. *Commun. ACM* 40, 10 (Oct. 1997), 103–110. DOI: http://dx.doi.org/10.1145/262793.262813

Jeanne Century, Michael Lach, Heather King, Sarah Rand, Courtney Heppner, Baker Franke, and Jean Westrick. 2013. Building an operating system for computer science. Retrieved March 23, 2015 from http://outlier.uchicago.edu/computerscience/OS4CS/.

Rudolfo Chávez. 1984. The use of high-inference measures to study classroom climates: A review. *Rev. Educ. Res.* 54, 2, 237–261.

Computing Research Association. 2017. *Generation CS: Computer Science Undergraduate Enrollments Surge Since 2006*. Retrieved from http://cra.org/data/Generation-CS/.

Computing Research Association. 2015. Taulbee Survey. *Comput. Res. News* 28, 5, 19.

Computer Science Teachers Association. 2013. CSTA national secondary school computer science survey: Comparison of results from 2005, 2007, 2009, 2011, and 2013 surveys. Retrieved April 25, 2015 from https://csta.acm.org/Research/sub/HighSchoolSurveys.html.

Jan Cuny. 2012. Transforming high school computing: A call to action. *ACM Inroads Mag*. 3, 2, 32–26. DOI: 0.1145/2189835.2189848

Eden Dahlstrom, J. D. Walker, and Charles Dziuban. 2013. *ECAR Study of Undergraduate Students and Information Technology*. EDUCAUSE Center for Analysis and Research, Louisville, CO.

Joost C. F. De Winter and Dimitra Dodou. 2010. Five-point likert items: T test versus mann-whitney-wilcoxon. *Practical Assessment, Research & Evaluation* 15, 11, 1–12.

École Polytechnique Federale de Lausanne. 2014. *MOOCS Annual Report 2012-2014*. Retrieved July 6, 2016 from https://documents.epfl.ch/groups/m/mo/moocs/www/MOOCS_Report_2012-14_1006_HD.pdf.

Mary English. 2013. *The Role of Newly Prepared PBL Teachers' Motivational Beliefs and Perceptions of School Conditions in Their Project Based Learning Implementation*. Unpublished doctoral dissertation. George Mason University, Fairfax, VA.

Barbara Ericson. 2016. Detailed data on pass rates, race, and gender for 2016. *Retrieved March 20*, 2015 from http://home.cc.gatech.edu/ice-gt/556.

Barbara Ericson, Mark Guzdial, and Tom McKlin. 2014. Preparing secondary computer science teachers through an iterative development process. In *WiPSCE'14 Proceedings of the 9th Workshop in Primary and Secondary Computing Education*. ACM, New York, NY, 116–119. DOI: http://dx.doi.org/10.1145/2670757.2670781

Sally Fincher and Josh Tenenberg. 2007. *Opening the Door of the Computer Science Classroom: The Disciplinary Commons*. In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education*.

Catherine O. Fritz, Peter E. Morris, and Jennifer J. Richler. 2012. Effect size estimates: current use, calculations, and interpretation. *J. Exp. Psychol.* 141, 1, 2.

Joanna Goode, Jane Margolis, and Gail Chapman. 2014. Curriculum is not enough: The educational theory and research foundation of the exploring computer science professional development model. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*. ACM, New York, NY, 493–498. DOI: http://dx.doi.org/10.1145/2538862.2538948

Google. 2015. Searching for computer science: Access and barriers in U.S. K-12 education. Retrieved October 19, 2015 from http://services.google.com/fh/files/misc/searching-for-computer-science_report.pdf.

Google. 2014. Women who choose computer science—what really matters? Retrieved March 23, 2015 from http://tinyurl.com/girlsincs.

Jeff Gray, Kathy Haynie, Sheryl Packman, Mary Boehm, Carol Crawford, and Deepa Muralidhar. 2015. A mid-project report on a statewide professional development model for CS principles. In *SIGCSE'15 Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, New York, NY, 380–385. DOI: http://dx.doi.org/10.1145/2676723.2677306

Gene Hall and Shirley Hord. 1987. *Change in Schools: Facilitating the Process*. SUNY Press, Albany, NY.

Jeff Haywood. 2014. Lessons from learning at scale: learners, technologies and directions. In *Learning at Scale Conference*.

Linda Kekelis, Rebecca Ancheta, and Etta Heber. 2005. Hurdles in the pipeline: Girls and technology careers. *Front.: J. Women's Stud.* 26, 1, 99–109. DOI: http://dx.doi.org/10.1353/fro.2005.0013

Karen Koellner and Jennifer Jacobs. 2016. Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *J. Teacher Educ.* 66, 1, 51–67.

Jane Margolis, Rachel Estrella, Joanna Goode, Jennifer Jellison Holme, and Kim Nao. 2008. *Stuck in the Shallow End: Education, Race, and Computing.* MIT Press, Cambridge, MA.

Tom McConnell, Joyce Parker, Jan Eberhardt, Matthew Koehler, and Mary Lundeberg. 2013. *Virtual Professional Learning Communities: Teachers' Perceptions of Virtual Versus Face-to-Face Professional Development. Journal of Science Education and Technology.*

Tom McKlin, Jason Freeman, Brian Magerko, and Mike Reilly, M. 2013. *Earsketch: Retrospective Student Engagement Survey Lanier High School (March 2013).* Retrieved January 27, 2016 from http://earsketch.gatech.edu/wp-content/uploads/2012/11/Lanier-High-School-Retrospective-Student-Engagement-Spring-2013.pdf.

Erin Mindell, Karen Brennan, Gwendolyn Britton, Jennifer S. Kay, and Jennifer Rosato. 2014. CS professional development MOOCs. In *Proceedings of the SIGCSE'14 Technical Symposium. ACM*, New York, NY, 369–370. DOI: http://dx.doi.org/10.1145/2538862.2538872

National Science Foundation. 2012. *Science and Engineering Indicators 2012.* Washington, DC.

Lijun Ni, Mark Guzdial, Allison Elliott Tew, Briana Morrison, and Ria Galanos. 2011. Building a community to support HS CS teachers: The disciplinary commons for computing educators. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education.* ACM, New York, NY, 553–558. DOI: http://dx.doi.org/10.1145/1953163.1953319

Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ.* 15, 5, 625–632.

Scott E. Page. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton University Press, Princeton, NJ.

S. L. Orpheus, Christopher D. Crutchfield, Guy Haas Harrison, Daniel D. Garcia, Sheila M. Humphreys, Colleen M. Lewis, and Peter Khooshabeh. 2011. Berkeley foundation for opportunities in information technology: A decade of broadening participation. *Trans. Comput. Educ.* 11, 3, Article 15 (October 2011). DOI: http://dx.doi.org/10.1145/2037276.2037279

Laura Perna, Alan Ruby, Robert Boruch, Nicole Wang, Janie Scull, Chad Evans, and Seher Ahmad. 2013. The Life Cycle of a Million MOOC Users. In *Proceedings of the MOOC Research Initiative Conference.* Arlington, TX.

Pat Phillips and Chris Stephenson. 2013. *Bugs in the System: Computer Science Teacher Certification in the U.S.* ACM, New York, NY.

Sue Sentance, Simon Humphreys, and Mark Dorling. 2014. The network of teaching excellence in computer science and master teachers. In *WiPSCE'14 Proceedings of the 9th Workshop in Primary and Secondary Computing Education.* ACM, New York, NY, 80–88. DOI: http://dx.doi.org/10.1145/2670757.2670789

Traci Sitzmann, Katherine Ely, Kenneth Brown, and Kristina Bauer. 2010. Self-assessment of knowledge: A cognitive learning or affective measure? *Acad.f Manag. Learn. Educ.* 9, 2, 169–191.

Lamar Smith. 2015. *House Passes Bipartisan STEM Education Act.* (February 2015). Retrieved October 19, 2015 from http://lamarsmith.house.gov/media-center/press-releases/house-passes-bipartisan-stem-education-act.

Chris Stephenson, Judith Gal-Ezer, Bruria Haberman, and Anita Verno. 2005. *The New Educational Imperative: Improving High School Computer Science Education.* Computer Science Teachers Association (CSTA), New York, NY.

U. S. Census Bureau. 2017. Quickfacts. Retrieved February 25, 2015 from https://www.census.gov/quickfacts/.

U. S. Department of Education, Office of Postsecondary Education. 2015. *Teacher Shortage Area Nationwide Listing 1991-1991 through 2014-15.* Retrieved January 26, 2016 from https://www2.ed.gov/about/offices/list/ope/pol/tsa.pdf.

Moshe Y. Vardi. 2012. Will MOOCs destroy academia? *Commun. ACM* 55, 11, 5.

Arto Vihavainen, Matti Luukkainen, and Jaakko Kurhila. 2012. Multi-faceted support for MOOC in programming. In *Proceedings of the 13th Annual Conference on Information Technology Education (SIGITE'12).* ACM, New York, NY. DOI: http://dx.doi.org/10.1145/2380552.2380603

Rebecca Vivian, Katrina Falkner, and Nickolas Falkner. 2014. Addressing the challenges of a new digital technologies curriculum: MOOCs as a scalable solution for teacher professional development. *Res Learn. Technol.* 22, 1 (2014), 24691. DOI: http://dx.doi.org/10.3402/rlt.v22.24691

Etienne Wenger, Richard Arnold McDermott, and William Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge.* Harvard Business Press.

Cameron Wilson. 2013. What's up next for Code.org? *IEEE Comput.* 46, 8, 95–97. DOI: http://dx.doi.org/10.1109/MC.2013.292

Eric Wiebe, Laurie Williams, Kai Yang, and Carol Miller. 2003. Computer science attitude survey. *Comput. Sci.* 14, 25 (January 2003), 0-86.

Received January 2016; revised March 2017; accepted March 2017