Evaluating Computer Science Professional Development Models and Educator Outcomes to Ensure Equity

Sloan Davis
Computer Science Education
Google Inc.
Mountain View, CA
sloandavis@google.com

Jason Ravitz
Computer Science Education
Google Inc.
Mountain View, CA
ravitz@google.com

Juliane Blazevski
Lead Researcher
Hypothesi
Ann Arbor, MI
jblazevs@umich.edu

***Abstract*—Google's educator professional development (PD) grants (formerly Computer Science for High School [CS4HS]) provide annual funding to education nonprofits that design and deliver Computer Science (CS) PD to educators in their local communities. As CS education is an emerging field, many education stakeholders can be ill-equipped to identify CS PD needs, evaluate options, and assess educator and student outcomes. As a result educators may participate in CS PD that fails to address their needs, which worsens equity gaps in CS education. Therefore, models of evaluating CS PD programs and outcomes are critical to equitable CS education. This paper provides an update on earlier research findings from 2014 with data from the 2015 and 2016 evaluation cycles, as well as updates to our evaluation measures and methodology.**

Keywords—Computer Science, Computer Science Education, Professional Development, Evaluation, Equity, Teachers, Educators

I. Introduction

Google's educator professional development (PD) grants (formerly Computer Science for High School [CS4HS]) provide annual funding to education nonprofits that design and deliver Computer Science (CS) PD to educators in their local communities. Google aims to scale equitable and sustainable access to CS education through localized PD that meets on-the-ground needs of educators and their school systems, with a focus on underrepresented and low-CS-momentum communities. Since 2009, Google has reached over 50,000 educators in more than 50 countries.

CS education is an emerging field (relative to other K–12 education disciplines), as is CS PD, which further complicates CS PD design and the evaluation of outcomes. As discussed by Darling-Hammond, Hyler, & Gardner (2017), "Few schools, districts, or state education agencies have created good systems of tracking PD, let alone systems for analyzing the quality and impact of PD" (p. 22). Given that CS is a field historically challenged with issues of representation, it is critical that K–12 education stakeholders are mindful of evaluating how CS is implemented in the classroom to ensure equitable access, participation, and outcomes.

Equity divides in CS education are worsened when educators are not prepared for the challenges they will likely face in the classroom. These challenges include students' lack of foundational skills, a lack of support and direction from administrators, and a lack of resources [1][6], as well as lack of a Community of Practice (COP) with shared purpose. Darling-Hammond et al. (2017) note that these communities reduce the "traditionally strong relationship between socioeconomic status and achievement gains in mathematics and science" (p. 17). Since 2014, Google has required that applicants incorporate a COP into their PD to provide ongoing support as educators implement CS in theclassroom.

Google is committed to understanding the impact of its funding on increasing the number of educators who are confident and competent in CS education, therefore broadening access to the discipline in K-12 schools. As a result, the grant program is run with an evaluation strategy that surveys PD providers and participating educators. The purpose of this annual evaluation is to examine the goals, objectives, and activities of the grant program and measure growth in attitudes and CS content knowledge over the course of the program. The data and findings are utilized to inform the growth and development of Google's CS education engagements. It is critical that educational leaders and institutions understand CS PD needs, evaluate PD options, and assess outcomes to ensure that education is equitably improving student participation, perceptions and proficiencies in CS. While Google currently funds a variety of PD programs representing different methodologies, we believe it is possible to design a common evaluation method that measures educator outcomes as they relate to participant demographics and CS PD design.

II. Method

Educators who attend Google-supported PD opportunities are requested to complete optional Pre- and Post-surveys at the first and last sessions of their PD/COP opportunity

978-1-5386-1363-4/18/$31.00 ©2018 IEEE

respectively. The research questions that guide these educator evaluations are:

- To what extent do Google-supported PD opportunities affect educator confidence and competence to teach CS in the classroom?

- To what extent do Google-supported PD opportunities equip educators with the skills, content and pedagogy needed to provide a quality learning experience for their students?

Since 2014, the evaluation methodology and measures have been refined to capture more precise data from educators while reducing the response burden. In 2014, we began measuring aggregate mean differences in pre-to-post responses for educators in the United States and Canada. In 2015, the evaluation process included a mechanism to link survey responses to a single individual with email address identifiers. The evaluation process has continued for the 2016 and 2017 grant years. In 2016, the evaluation process expanded to Google-supported PD programs in Africa, China, Europe, and the Middle East. In 2017, the evaluation process expanded to Google-supported PD programs in Australia and New Zealand.

In this paper, we present the evaluations' scale reliability, and pre-post outcomes (using paired t-tests through voluntary, user-submitted email address identifiers) of educators who participated in the 2014, 2015, and 2016 cohorts of Google-funded CS PD opportunities in the United States and Canada (US/CA). The 2016 data are analyzed by demographic subgroups of prior CS teaching experience, middle or high school teaching, COP expectations, and content implementation.

III. RESULTS

Results of the 2014 evaluation process were published as an "early findings" article in TOCE [5] This article compared self-reported learning gains and experiences of educators in four Google-funded PD opportunities. The findings were based on unmatched pre-post data, however we were able to determine participants in both the pre-survey and post-survey were demographically and experientially very similar which suggested our analyses still had merit. Analyses from 314 pre-surveys and 129 post-surveys illustrated that the CS educator participants were quite heterogeneous, suggesting that some ability to customize PD based on educator background and needs would benefit educator outcomes. We reported a preliminary finding that the two face-to-face PD experiences appeared to engender a stronger sense of community than the online or blended experiences. Finally, among the outcomes we measured, educator concerns [5] were more sensitive to change than our measures of self-efficacy, outcome expectations, readiness, or beliefs. We highlighted the variety of CS educator PD experiences and the need to study the effective ways to scale CS teacher education to meet the needs of a wide range of educators and contexts, and we highlighted methodological and measurement challenges to assessing online PD outcomes.

Since the previous results were shared, we have had the opportunity to improve our data collection and replicate and improve on many of the findings. The updated findings we present for RESPECT 2018 include the 2015 and 2016 evaluation cycles:

TABLE I. GRANT PROGRAM AND SURVEY PARTICIPATION

	2014[a]	2015	2016
Program sites evaluated	n = 4	n = 14	n = 36
Pre-survey response rate	n = 314	18% (n = 348)	68% (n = 672)
Post-survey response rate	n = 129	8% (n = 148)	38% (n = 373)
Linked response rate	n = 0	4% (n = 75)	7% (n = 68)

a. 2014 cycle did not identify original participant counts; response rate not calculated.

Not only did the validity of the data increase from 2014 to 2016 as a result of adding the voluntary email identifiers for paired t-test analysis, but also the response rates have consistently increased, indicating that the conclusions are increasingly representative. Further, we have revised the evaluation process to include more effective measures that produce actionable learnings across sites and outcomes, analysis by demographic subgroups, analysis of COP influence on educator outcomes in CS education, and analysis of educator outcomes as they relate to classroom implementation of content learned in the PD opportunity.

A. *More effective measures*

The 2014 study indicated the Concerns items were the measures for which we saw the most significant changes. Reliability for the measures continued to be strong ($\alpha \geq 0.82$) for all US/CA scale scores for both the 2015 and 2016 evaluation cycles (Tables II and III). While the Concerns measures still show statistically significant change, new "Readiness" items added in 2015 are also sensitive to change. We still are not seeing a statistically significant change in self-efficacy for the scale and most items. Given that the self-efficacy scale scores are already positive at the pre-survey, we suspect that participants' speculative analysis of the self-efficacy items (e.g. "I can effectively teach the concepts required by the curriculum") is more difficult for the respondents to answer because they have not completed a full cycle of classroom implementation with the PD-learned CS content. In contrast, educators have the opportunity to reflect on their concerns and readiness attitudes throughout the PD, potentially demonstrating to themselves that they have improved in those areas.

The measures used are refined each grant year for multiple reasons. First, we identify gaps in what we are measuring relative to the priorities of the grant program or CS education landscape (e.g. whether educators implement PD-learned content and if it is successful was added for the 2016 grant year). Further, the 2014 study included exploratory measures intended to generate data about how different PD delivery models relate to educator outcomes. Through the 2015 and 2016 evaluation cycles, items that had no relationship to the outcomes educators gained nor to the site-by-site contexts of the PD opportunities funded were removed. These items were

978-1-5386-1363-4/18/$31.00 ©2018 IEEE

not answering the most important questions about whether outcomes change as a result of participating in PD, or which elements of PD relate to those outcomes. We replaced those dropped measures with attitude items on “Readiness” that were more effective in showing variation across sites and relationships to outcomes. This resulted in a reduced response burden and more actionable learnings. The items we added include the following:

- I am confident in my ability to teach CS effectively;
- I have the knowledge and skills I need to teach CS effectively;
- I have the curriculum tools and resources I need to teach CS effectively;
- I have a social network that enables me to teach CS effectively.

TABLE II. US/CA PRE-POST CHANGES IN ATTITUDE SCALES (2015)[A]

	a[b]	N		Mean (SD)		Change	Effect Size[c]
		Pre	Post	Pre	Post		
Concerns							
Linked	0.83	75	75	2.48 (0.56)	2.08 (0.64)	-0.40	-0.71**
Unlinked	0.83	345	128	2.54 (0.56)	2.09 (0.67)	-0.45	-0.80**
Self-Efficacy							
Linked	0.84	75	75	3.87 (0.70)	4.02 (0.57)	0.15	0.21
Unlinked	0.84	345	126	3.89 (0.68)	4.08 (0.55)	0.19	0.28*
Readiness							
Linked	0.82	75	75	3.31 (0.86)	4.00 (0.65)	0.69	0.80**
Unlinked	0.82	343	128	3.44 (0.88)	3.97 (0.71)	0.53	0.60**

a. Using paired t-tests and ANOVA (unpaired). ** p < 0.001. * p < 0.01.

b. Reliability. Cronbach’s alpha.

c. Effect sizes are based on pre-SD.

TABLE III. US/CA PRE-POST CHANGES IN ATTITUDE SCALES (2016)[A]

	a[b]	N		Mean (SD)		Change	Effect Size[c]
		Pre	Post	Pre	Post		
Concerns							
Linked	0.89	64	64	2.50 (0.56)	2.15 (0.60)	-0.35	-0.62**
Unlinked	0.89	662	369	2.51 (0.53)	2.08 (0.66)	-0.43	-0.81**
Self-Efficacy							
Linked	0.89	67	67	3.90 (0.71)	4.00 (0.70)	0.10	0.14
Unlinked	0.89	670	373	3.86 (0.73)	4.04 (0.65)	0.18	0.25**
Readiness							
Linked	0.87	67	67	3.41 (0.94)	3.84 (0.81)	0.43	0.46**
Unlinked	0.87	668	374	3.27 (0.96)	3.91 (0.81)	0.64	0.67**

a. Using paired t-tests and ANOVA (unpaired). ** p < 0.001. * p < 0.01.

b. Reliability. Cronbach’s alpha.

c. Effect sizes are based on pre-SD.

B. *Sub-group outcomes*

As of 2015, we were able to analyze pre-post gains by subgroup. We see significant changes in attitudes (in intended directions) for both those who have taught CS before and those who have not, as well as for both middle school and high school educators. Although we only had 19 matched cases in 2016 for teachers who had never taught CS before, their gains were significantly higher than gains reported by teachers with prior experience teaching CS. This finding underscores the importance of a focus on CS fundamentals and pedagogical content knowledge, so that educators are resilient in the face of technology changes and prepared to support diverse student needs. From 2015 to 2016, we saw a 52% increase in educator respondents who indicated no prior CS teaching experience. It is critical that PD providers continue to address the varied needs of this educator population to ensure an equitable rollout of CS education to all students.

C. *Communities of Practice*

We introduced communities of practice (COP) as a focus in 2014 and have continued studying how a sense of community relates to educator outcomes. Here we focus on changes in readiness attitudes according to whether or not participants’ expectations for the COP were met. Educators whose COP expectations were met improved their attitudes more than those whose were not met. For example, the pre-post change on the readiness scale for those whose expectations were met ($n = 31$) was a statistically significant 0.62 ($ES = 0.75$, $p < 0.001$), while for those whose expectations were not met ($n = 35$) the change of 0.30 did not reach statistical significance ($ES = 0.36$, NS, $p < 0.10$). Moreover, the pre-post change was significant on all four readiness items ($p < 0.05$) for those whose expectations were met, but for those with unmet expectations the change was significant only for the knowledge and skills-related item. The biggest difference related to COP expectations had to do with having a social network. For teachers whose expectations were met 81% (out of 16) of those with a negative or neutral pre-score shifted to a positive response, while only 29% (out of 21) with unmet expectations had a similarly positive shift (Chi-Sq, $p < 0.002$).

D. *Implementation of content*

Unlike in the previous study, we linked data on self-reported implementation of PD content to attitude shifts in the 2016 evaluation cycle. Overall, 64% of educator respondents estimated they implemented around 50% or more of the content they learned in their Google-supported PD. The pre-scores on readiness attitudes were similar for those who did and did not implement 50% or more of the content. However, those who implemented 50% or more of the content improved their readiness attitudes more than others. The pre-post change for those who implemented 50% or more ($n = 33$) was 0.59 ($ES = 0.69$, $p < 0.001$) while for those who did not implement 50% ($n = 30$) it was 0.29 ($ES = 0.34$, NS, $p < 0.10$). For those who implemented 50% of the content, the pre-post change was significant on all four readiness items, while for those who did not implement 50% of the content the change was significant only for the knowledge and skills-related item. The biggest difference related to implementation of content involved curriculum tools. Specifically, 67% (out of 18) of the high implementing teachers with a negative or neutral pre-score shifted to a positive response, while only 27% (out of 15) who implemented 50% or less had a similarly positive shift (Chi-Sq, $p < 0.02$).

IV. Discussion

Since 2014, the evaluation process has refined the measures used to evaluate educator outcomes, increased the number of educators who respond, and replicated trends we found in pre-post educator outcomes. Additionally, the measures have produced reliable results, despite the highly varied PD opportunities in which educators have participated (e.g. face-to-face versus online, AP Computer Science Principles versus Exploring Computer Science). Current survey measures and further analyses can be obtained by emailing the authors.

Considering the greater and more significant outcomes of certain educator subgroups (no prior CS teaching, COP expectations met, majority of PD-learned content implmeneted), the findings of the 2015 and 2016 evaluation cycles support the 2014 conclusion that PD customized to educator background and needs benefits outcomes. This further underscores the need for robust evaluation of PD needs, options, and outcomes to scale CS education equitably.

There are two primary limitations to this evaluation process. First, the Google team has no direct engagement with educators who participate in PD opportunities, nor their students. While we provide documentation about the evaluation process, PD providers ultimately decide if they will distribute surveys and how they are communicated to educators. This lack of direct access to educators and students prevents Google from conducting longitudinal/multi-year analysis of educator outcomes, or evaluation of student changes in belief, attitude, or proficiency in CS. A second limitation is the varying nature of PD opportunities supported by Google; a hallmark of the program is that applicants identify the CS PD needs of local educators and tailor their PD opportunities accordingly. While tailoring CS PD to educator needs produces greater outcomes for both educators and students [2], it prevents Google from designing and executing a controlled, comparative assessment of PD models to identify what formats, curricula, and so on are most effective in CS teacher training.

V. Conclusions

Google maintains that educator professional development can affect widespread participation in CS through institutional change. However, it is critical that educator and student outcomes of CS PD opportunities are appropriately measured to ensure that perception and proficiency gaps are not widening inequitably.

Despite limitations, the Google evaluation process has effectively measured pre-post changes for educators with significantly different backgrounds in demographics, education, CS, and CS education. In addition to identifying measures and methodologies for tracking the outcomes of educators who participate in a variety of CS PD oppotunities, we have also learned that the programs funded by Google can be linked to objective increases in CS educator confidence and competence. Finally, extant research has shown that a localized PD model produces greater outcomes for educators and students that can help overcome traditional equity barriers in education and CS [1][2][3][6] Our evaluation process indicates that the model of Google's CS educator PD grants program (i.e., localized PD coupled with an academic year COP that support educators during CS implementation) is consistent with these findings from the literature.

It is possible to evaluate attitudinal outcomes of a wide variety of CS PD opportunities through shared evaluation measures while also identifying what PD elements are most impactful. However, more controlled studies of CS PD programs are needed to better understand which PD models provide the best outcomes for students and educators in varying contexts. As the CS education field looks to broaden participation in CS to ensure sufficient representation, developing models to identify CS PD needs, evaluate PD options, and assess outcomes are paramount to maintaining consistent engagement of all students with CS.

Acknowledgment

The authors would like to acknowledge the PD providers and educators who participated in the evaluations process and contributed their experiences to better our understanding of the CS PD landscape. We would also like to thank Chris Stephenson, Karen Parker, Olga Garcia and many others at Google who have supported this evaluation process since 2014. S.D. would like to thank Kat and Steve Davis, Elliot Bartz, Joshua Garoon and all of the family, friends, and educators who supported the journey to this research.

References

[1] Armoni, M. (2011). Looking at secondary teacher preparation through the lens of computer science. ACM Transactions on Computing Education (TOCE), 11(4), 23.

[2] Century, J., Lach, M., King, H., Rand, S., Heppner, C., Franke, B., & Westrick, J. (2013). Building an Operating System for Computer Science. Chicago, IL: CEMSE, University of Chicago with UEI, University of Chicago. Retrieved August 25, 2017, from http://outlier.uchicago.edu/computerscience/OS4CS/

[3] Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Palo Alto, CA: Learning Policy Institute.

[4] Hall, G. E., & Hord, S. M. (1987). Change in Schools: Facilitating the Process. Albany, NY: State University of New York Press.

[5] Ravitz, J., Stephenson, C., Parker, K., & Blazevski, J. (2017). Early Lessons from Evaluation of Computer Science Teacher Professional Development in Google's CS4HS Program. *ACM Transactions on Computing Education*, 17(4). https://doi.org/10.1145/3077617

[6] Sentance, S., & Csizmadia, A. (2016). Computing in the curriculum: Challenges and strategies from a teacher's perspective. Education and Information Technologies, 22(2), 469-495. doi:10.1007/s10639-016-9482-0