Towards a Common Framework for Evaluating Computing Outreach Activities

Adrienne Decker
Rochester Institute of Technology
2145 Golisano Hall
Rochester, NY 14623 USA
01-585-475-4653
adrienne.decker@rit.edu

Monica M. McGill
Bradley University
1501 W. Bradley Ave
Peoria, IL 61625 USA
01-309-677-4148
mmcgill@bradley.edu

Amber Settle
DePaul University
243 S. Wabash Ave
Chicago, IL 60604 USA
01-312-362-5324
asettle@cdm.depaul.edu

ABSTRACT
In the past six years, dozens of conference papers and journal articles have been presented in Association of Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE) educational forums concerning computing outreach activities. Nearly half of these (47.5%) appeared in SIGCSE venues. In this study, we used the free-form question "What type of data has been collected in formal, peer-reviewed research that has been conducted on computing outreach activities in recent years?" as a basis for a systematic literature review in these venues from 2009-2015. During the analysis of the articles, it was discovered that a majority of efforts focused on middle school and high school students, a majority of the reported events took place in the United States, and almost half had a goal of increasing gender diversity in computing. This paper summarizes the information about the studies, including their data collection techniques and the data that was collected. We also present a list of recommended practices for data collection, methodologies, and reporting for educational researchers engaged in these activities in an effort to provide comparative data and allow us as a community to more scientifically understand the impact that these activities are having on the participants.

Keywords
Outreach; Underrepresented groups in computing; Effectiveness; Literature Review

1. INTRODUCTION
At the 2015 Association of Computing Machinery (ACM) International Computing Education Research (ICER) conference, Jim Spohrer, the keynote speaker from IBM Global University Programs, highlighted the current shortage of technology workers and emphasized the exponential growth of this shortage that they anticipate in the next 10 to 40 years [18]. This has become a critical concern within the U.S., with the Gartner group and various Fortune 50 companies echoing the same sentiments [5].

While some say these concerns are driven by industry's desire to increase limits on H1-B visas that allow companies to hire foreign workers at significantly lower salaries, there are factors that still warrant consideration. The U.S. Department of Labor predicts a 17.7% increase in growth for computer occupations from 2012 to 2022. This is led by information security analysts (36.5%), computer system analysts (24.5%), and application software developers (22.8%), resulting in a need for 1,240,100 additional workers [3, 4].

At the same time, policy and decision makers have concluded that there is a lack of diversity within the field of computing and that this lack of diversity contributes to product performance challenges [12]. Much research has been conducted on this topic and industry profits of diverse teams have proven that diversity has an important role in creating better products [1, 13, 17].

In consideration of the above, industry and the government alike have poured millions of dollars of resources into outreach programs to address this issue [7, 11]. One primary goal has been to recruit and retain more students into the computing discipline through various outreach activities, with some efforts focusing primarily on the goal of broadening the participation of women and minorities [12]. Various outreach activities have been produced by academics and industry and have been primarily focused on students from Kindergarten through High School, with several of these programs (Code.org, Black Girls CODE, Girls Who Code, etc.) garnering national media attention [2, 16, 19].

But are these programs effective in recruiting and/or retaining students into studying computing? And are these programs effective in broadening the participation of women and minorities? Or simply, in industry terms, what is the return on investment and who is measuring the success of these activities?

Though industry may be measuring their effectiveness internally, the questions are valid and necessary to pose not only from a purely academic perspective, but also from the perspective of acknowledging and addressing the critical concerns voiced by government and by industry. Turning to reliable research available in computing education that reports on computing activities, the question becomes more basic: What data is reported and what types of successes do previous activities report?

To address these questions, we conducted a systematic literature review to determine what has been reported about the impact of computing outreach programs. After our initial experience with this data, we believed that more needed to be done to determine the impact of outreach activities on participant's choice of major and started towards that process [9].

However, the results of the literature review bear further discussion here. We conclude that a concerted effort must be made to collect data that can be compared and measured both in the long- and short-term and provide a framework for improving the quality of these studies. Therefore, the value of this study is two-fold and will be of

SIGCSE '16, March 02-05, 2016, Memphis, TN, USA

DOI: http://dx.doi.org/10.1145/2839509.2844567

interest to policy makers, those who invest time and resources in outreach activities, and researchers whose interests lie in broadening participation specifically or recruitment and retention in general.

2. METHODOLOGY

We undertook a systematic literature review to identify, evaluate, select, and synthesize results of high quality research involving computing outreach programs. To begin this initiative, we chose to primarily follow the framework developed by Khan, Kunz, Kleijnen, and Antes [8], with additional guidance from Petticrew and Roberts [15]. The framework has five foundational steps: frame the question, identify relevant work, assess the quality of the studies, summarize the evidence, and interpret the findings. This section describes the first three steps in detail, while the evidence summary and interpretation of the findings is presented in the following two sections.

2.1 Framing the research question

To begin the review process, a broad, free-form question was posed: "What type of data has been collected in formal, peer-reviewed research that has been conducted on computing outreach activities in recent years?" This question was carefully formulated to serve as a guide for identifying relevant work. Upon considering the research question, the following basic overarching characteristics were identified:

- Populations Studied—Students enrolled in computing outreach programs as defined by the researchers
- Interventions—Programs that exposed students to computing concepts
- Study designs—Quantitative, qualitative, or mixed methods studies
- Outcomes—Effects of the program on participants' behaviors, attitudes, skills, knowledge, or dispositions

2.2 Identifying relevant work

The ACM and Institute of Electrical and Electronics Engineers (IEEE) journal and conference publications have served as a reliable source for formal, blind, peer-reviewed computing education research for decades. Using this as a starting point for finding relevant literature, we further refined that to venues that emphasize education, identifying the following peer-reviewed journals and conference proceedings in electronic form as a reliable source for finding relevant research:

- SIGCSE Technical Symposium on Computer Science Education (SIGCSE)
- Frontiers in Education (FIE)
- Innovation and Technology in Computer Science Education (ITiCSE)
- International Computing Education Research Workshop (ICER)
- Taylor & Francis' Computer Science Education (CSE)
- Transactions on Computing Education (TOCE)

Publications from 2009 to 2015 inclusive[1] were considered. This resulted in 3,949 citations that were reviewed for relevance. An article was determined to be relevant if, upon human review, it had a title and abstract associated with outreach because it contained one or more of the actual or related identifying criteria: K-12, elementary school, high school, secondary school, after school clubs, summer camp.

After this initial search for relevance, 3,837 papers did not fit the criteria and were deemed irrelevant for the purposes of this review. This resulted in 112 articles to undergo a more thorough review. To evaluate these, the overarching characteristics were considered and collection points were created for organizing the data.

Table 1. Guidelines for identifying relevant work

Characteristic	Collection points
Populations Studied	Participant characteristics (age and/or grade in school, gender, ethnicity, location) Number of participants in study
Interventions	Goals and facets of the program
Study Designs	Research question Quantitative, Qualitative, Mixed Methods, or Other Longitudinal, cross-sectional, experimental, quasi-experimental, etc. Type of data collected (participants' behaviors, attitudes, skills, knowledge, or dispositions)
Outcomes	Results of the study

2.3 Assessing the quality of the studies

Data was collected for each point in Table 1 via a careful read of the 112 articles. Thirty-two of these did not use formal methods (quantitative or qualitative) to evaluate the impact of the associated outreach activities. Instead, these articles described the activity in general terms, provided advice on creating and implementing an activity, provided curriculum examples for these activities, or were work in progress papers that did not include any data or findings. This left 80 articles for in-depth review and analysis, both of which are discussed below.

3. RESULTS (SUMMARY OF EVIDENCE)

The next step in the systematic literature review was to summarize the evidence from our collection points of the remaining 80 articles. Table 2 provides an analysis of the venues for the papers by venue and year and note that the majority (47.5%) of the studies were published in SIGCSE forums over the last 6 years.

The papers were dominated by results from interventions in the U.S. (72%). Figure 1 shows the breakdown of where the interventions took place by country.

We converted the level of students in non-US activities to the U.S. system since that is where a majority of the interventions took place. A majority of the interventions (80%) were outreach efforts aimed at high school and/or middle school students. The others were aimed at various stages of students as outlined in Figure 2. Many outreach activities are aimed at broadening participation in computing by either increasing gender or ethnic diversity in computing. Of the interventions, 49% indicated that they were intended to increase gender diversity and 31% indicated that they were intended to increase ethnic diversity. There were also a number of interventions where the intentions were not at all clear

[1] FIE 2015 proceedings were not available at the time of writing and are not included.

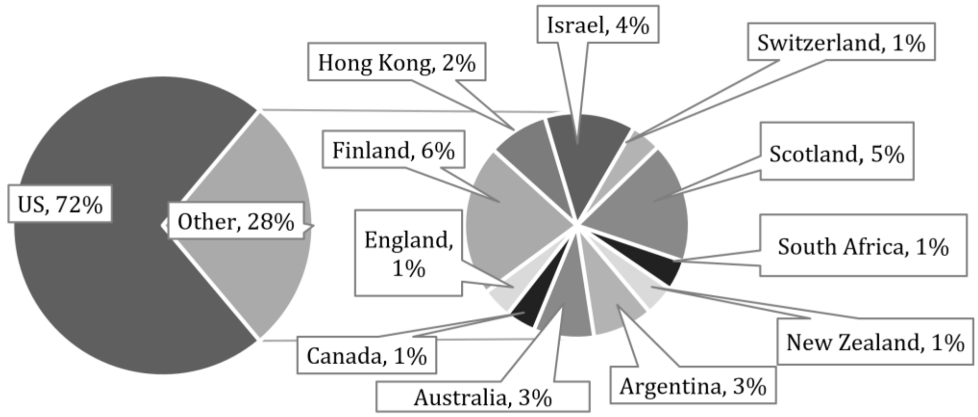

Figure 1. Countries where outreach activities took place

Table 2. Articles found by venue and year

	Number of articles meeting criteria							
	‘09	‘10	‘11	‘12	‘13	‘14	‘15	Total
SIGCSE	10	5	3	4	7	5	3	37
FIE	2	3	2	3			n/a	10
ITiCSE	3	2	2		2	1	2	12
ICER				1		2	3	6
CSE			1	1	2			5
TOCE			9	1		1		11
Totals	15	10	17	10	11	9	8	80

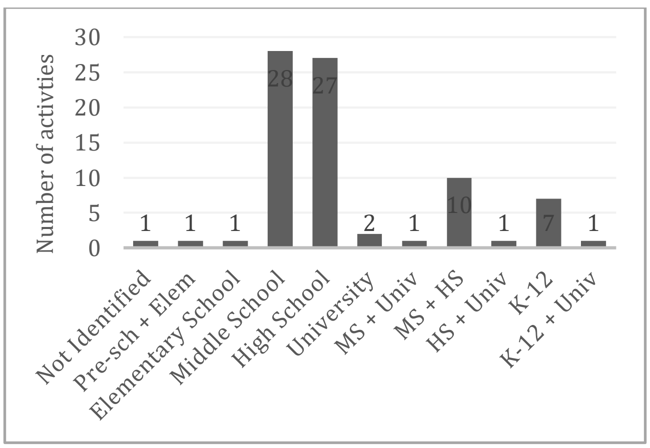

Figure 2. When outreach activities were offered

from the article. In fact, four of the articles (5%) were unable to be classified in either category. That is, the explanation of the intervention in the article did not allow for the ability for us to categorize them as either looking to improve diversity or not. Table 3 shows this breakdown.

The number of participants as reported by the interventions ranged from 2 to 9,999. Figure 3 shows in ranges the number of interventions that reported participants. The largest number of studies (11) had between 20-29 participants. The 16 studies (20%) that did not clearly report the total number of participants in the study are not included in Figure 3.

Table 3. Interventions designed to increase diversity

	Gender	Ethnic
Yes	39 (49%)	25 (31%)
No	37 (46%)	48 (60%)
Unknown	4 (5%)	7 (9%)

The gender of the participants was reported by 72.5% (58) of the studies. There were interventions that were single gender and interventions that were both genders. There were 22 studies (27%) that did not report the gender of the participants. Figure 4 shows the breakdown of gender in the interventions.

Of the activities that identified as mixed gender, 29 indicated the exact gender breakdown. Note that of those interventions previously identifying as attempting to address gender disparity (39), only 30 gave the exact gender breakdown of their interventions. Figure 5 shows percentages of female participation and how many studies were in each range.

The ethnicity of the participants was reported by 28 (35%) of the studies. There were only 3 studies (3%) that were strictly minority participation (two strictly Hispanic/Latino/Latina and one American Indian as reported by the articles). There were 25 studies (31%) that reported mixed ethnicity that included at least some minority participation. For those studies, the articles indicated the following races/ethnicities for the minority participants: American Indian or Alaskan Native; Asian; Asian/Pacific Islander; Black or African American; Filipino; Hispanic, Latino, Latina; Multi-racial; Other. There were 52 studies (65%) that did not indicate participant ethnicity.

The methodology of the studies is shown in Figure 6. Most of the studies were quantitative in nature (61%), but there was a significant amount of qualitative and mixed methods studies represented as well. For one of the studies, we were unable to classify its exact methodology from the description.

Table 4 breaks down the different types of data collected by the studies and the number of studies that collected this type of data. Computer science, information technology, and related disciplines are coded as “computing” for purposes of distinguishing the field.

Many of the studies collected data on more than one variable, so the studies were coded for each variable that was collected and reported. The most frequent data collected was participant attitudes towards computing (31% of studies). The second most frequent data collected was about interest in future study and/or careers in

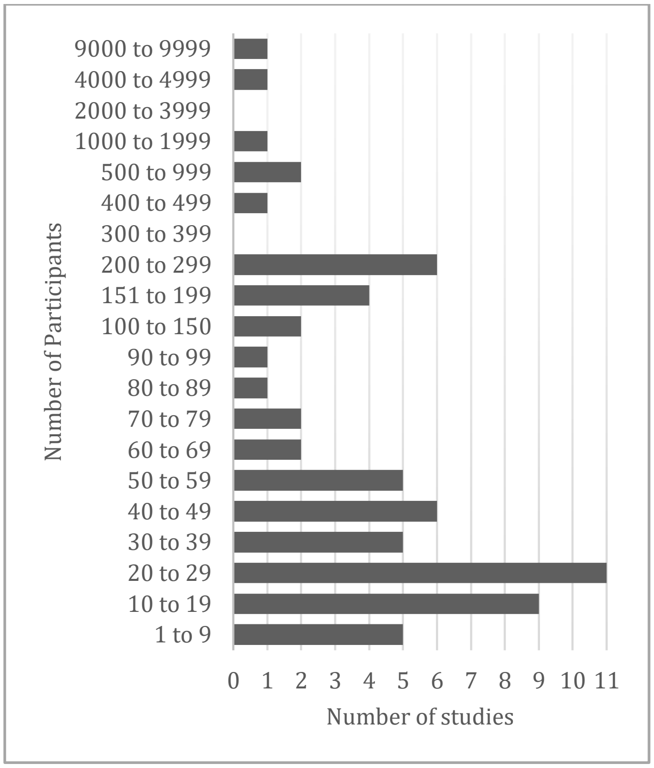

Figure 3. Number of participants in each study

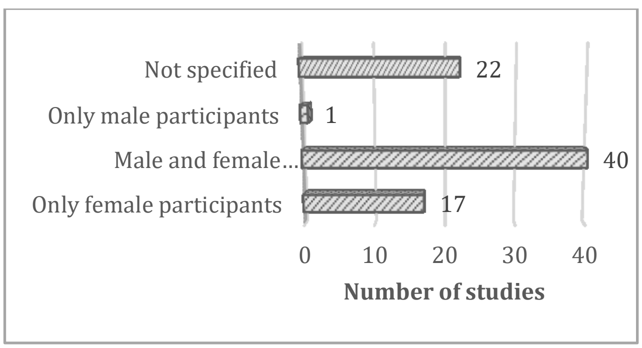

Figure 4. Breakdown of participants' gender in all studies

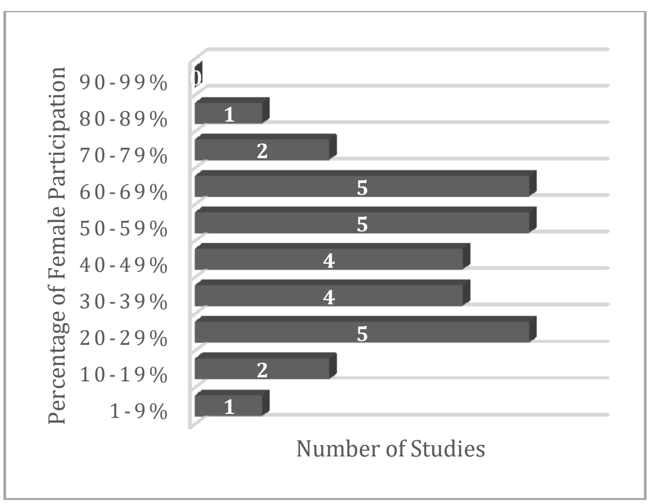

Figure 5. Female participation in mixed gender studies

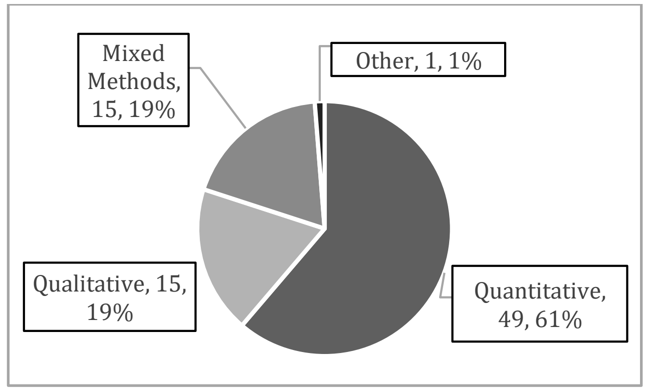

Figure 6. Study methodologies

computing (23%) and assessment of computing skills (23%) through some type of test.

Several studies were difficult to categorize. One study simply discussed the experiences of the students during the intervention, but not using a rigorous qualitative method. One study wanted to understand student differences by categorizing their personal study orientations. One study reported on the results of giving a university course as-is to high school students as an intervention. Finally, one study indicated that data was collected, but did not give information on what type or results.

Longitudinal data was only reported by 7 (8%) of the studies in the literature review. The durations of the longitudinal study were 9 months to 10 years, with an average duration of 4 years. The studies (short-term or longitudinal) all reported positive or neutral findings. There were no reports of negative findings for the outreach initiatives.

4. DISCUSSION

As noted previously, the question that this literature review was attempting to answer was, "What type of data has been collected in formal, peer-reviewed research that has been conducted on computing outreach activities in recent years?"

Upon analysis, we found that there was considerable data presented on outreach activities, but there is also missing data. First and foremost, there is little longitudinal evidence on the impact of these activities on the participants. Only 8% of the articles reflected a long-term study. In order to adequately address the larger question of the long-term effectiveness of these programs, more longitudinal studies are needed.

However, focusing on what has been reported, we see that the majority of the reported interventions are from the United States. The reason for this may be that the venues for publication are heavily US-centric. It should also be noted that this is the only category for which there are no ambiguities about the data. This is because the affiliation of the authors was used as a guide for determining location of the interventions. In many cases, the exact location for the interventions was not reported in the paper.

A majority of the interventions were aimed at middle school and high school students, which is not surprising given that several of the interventions were designed to increase student awareness of computing as a discipline, major, and career. There were a fair number targeting younger children--many target multiple age ranges or allow students to progress through them as they age.

Table 4. Summary of types of data collected by studies

Type of Data Collected	Number (Percent) of activities collecting
Attitudes towards computing	25 (31%)
Interest in future study of computing and/or interest in pursuing a career in computing	19 (23%)
Assessment of computing (programming) skills	19 (23%)
Perception of the field of computing	15 (18%)
Enjoyment of intervention	13 (16%)
General interest in computing	9 (11%)
Self-reported abilities with computing concepts	8 (10%)
Self-efficacy	4 (5%)
Number of majors in computing at university	4 (5%)
How material presented in intervention related to participants or the real world	4 (5%)
Relevance of computing	3 (3%)
Ability to express creativity with computing	3 (3%)
Engagement (general)	3 (3%)
Engagement with tools beyond the scope of the assigned task	3 (3%)
Assessment of spatial reasoning ability	2 (2%)
Participant GPA	2 (2%)
Motivation/persistence	2 (2%)
Completion of assigned task	2 (2%)
Assessment of other STEM skills	1 (1%)
Participant drop rate from computing program	1 (1%)
Identity within computing	1 (1%)
Belonging within computing	1 (1%)
Future enrollment in computing course	1 (1%)

Nearly half (49%) of the studies indicated that they were designed to increase either gender and/or ethnic diversity of the field of computing. However, only 31% indicated that increasing ethnic diversity was a goal. We know that both the lack of gender and ethnic diversity are problems in the field, but these numbers indicate that the problem of gender balance is being tackled more often than ethnic balance.

The number of participants in the studies varied greatly, but over half had less than 100 participants, and 36 (45%) had less than 50 participants. This is not surprising given the nature of these activities as summer camps or after school programs. The number of students corresponds to a typical size of a class/cohort or perhaps up to two classes. However, such small numbers and the lack of repetition of the intervention lead to problems for generalization of the activity's impact and effectiveness.

The measures of effectiveness varied across studies. Not surprisingly, participant attitudes about computing, potential further study of computing and interest in computing careers dominated the studies. Another highly measured outcome was participant knowledge about computing concepts, particularly programming constructs, which were considered by 23% of the studies. The way this data was collected varied from study to study, but for the most part, this data was collected using surveys that were created by the leaders of the interventions/studies.

The rigor in which results were reported varies considerably. While some report on the use of statistical means of analyzing the data, others report raw scores or means and indicate a raising or lowering of the means as success. Overwhelmingly, the studies reported positive or neutral findings. Because of the low numbers of participants for some of the interventions studied, statistical methods may not be viable for assessing the success of the intervention. However, case study or other rigorous qualitative methods could have been used in these cases. Except in a very few cases, qualitative studies simply gave anecdotes and observations about the intervention rather than following grounded theory or other acceptable qualitative study techniques.

Several studies did not clearly provide raw number or percentage for one or more pieces of data. There were studies where we were unable to categorize their intentions on increasing diversity of gender or ethnicity. For those that did indicate goals of increasing diversity, we discovered that a significant portion did not report on the exact gender or ethnic breakdown of participants. Furthermore, twenty percent of the studies did not clearly indicate the total number of participants in the interventions in any way.

As with any study, there are limitations. This literature review is focused on venues for academic researchers within the computing community. Research that does not appear in the aforementioned academic venues, like research from the National Center for Women in Technology (NCWIT) and the Girl Scouts of America, are not included in this study [10]. We evaluated articles in the selected venues, and these articles served as representation for the whole—though we recognize that there may be relevant articles in other venues. Further, despite our careful analysis, there may have been articles that should have been included, but were not. If a title did not seem to indicate an association with outreach activities, its abstract was not examined. Despite these limitations, the findings present an image of the quantity and quality of research being conducted and presented within the SIGCSE and larger ACM/IEEE community.

5. RECOMMENDATIONS

Based on this review as well as on standard research methods and practices defined by expert researchers [6, 8, 14, 15], we find that the results of these studies are difficult to compare and, in general, at best lack adequate data and at worst follow research methodologies that are weak and/or unsupported by formally accepted practices.

For reviewing, analyzing, and synthesizing qualitative data, particularly in grounded theory research, prescripting a framework or guidelines for such data contradicts accepted sound methodology. However, for quantitative data, studies with similar data collection techniques can provide a better pathway for cross-comparison and analysis of data. We therefore propose a basic framework to consider when conducting quantitative studies centered on computing activities designed to broaden the pipeline:

Preliminary steps

- Define overarching research question(s) to be studied
- Ensure that the data collection and reporting of data has been approved by your local institutional review board
- Consider variables outside of the study that may influence the outcomes and include these as part of your report
- Define data to be collected to provide answers to the research question

Data to be collected

- Collect basic demographic data on the participants, including gender, ethnicity, age, grade in school
- Collect any other unique characteristics about the participants that may influence the study (participated in

previous activities, were all gifted students, etc.)

- Use reliable, validated survey instruments, when possible, to gauge participant attitudes, self-efficacy, and skills if one or more of these are used to answer your research question(s)
- Consider the number of students in the group; statistical analysis such as a t-test typically requires 26 or more to be considered valid

Reporting

- Provide the research question and/or the purpose of the intervention (computing activity)
- Describe type of activity and where activity was held (including country)
- Provide amount of time participants were engaged in the activity (hours/days/weeks)
- Provide information on who ran the activity
- Provide data that was collected, reporting at a minimum the gender, ethnicity, age, and grade in school of participants, both in count and in percentages.
- For each piece of data collected, report count and percentages

Though there are many other ways to report data, and data collection is often very dependent on the nature of the study, these basic items may provide a measure of consistency among types of data that can be compared.

Even though these suggestions apply most acutely to quantitative studies, there are still many more things to be learned through rigorous approaches to qualitative studies surrounding the impact of outreach activities. While not highlighted explicitly in this paper, there were several very good examples of qualitative research exploring outreach initiatives. Given that our study shows that the number of participants in computing activities is generally low, sound qualitative methods when appropriate will enhance understanding of the personal impact on participants.

6. CONCLUSION

There is a significant number of outreach activities that are reported in the literature that impact a variety of different constituents across ages, countries, ethnicities, and gender. The activities are reporting impact on the participants through various metrics that are overwhelmingly positive. However, we did not find sufficient meta-evidence to conclude that, as a whole, computing outreach activities are effective, especially in the long-term.

This is in part due to the fact that we also uncovered many reports that are lacking in several details that are important from a research perspective for those interested in studying outreach efforts on a broader scale. We encourage conducting such studies to collect data that can be compared across programs. We also encourage researchers to track participants beyond the end of the activity. Though one-time activities can often provide a boost to a participants' self-efficacy or beliefs about computing, these can radically change over two years, two months, or even two days' time. By tracking longer periods, the results become more meaningful and will aid researchers in identifying practices that have higher success rates over the long-term.

7. REFERENCES

[1] Armstrong, C., Flood, P. C., Guthrie, J. P., Liu, W., MacCurtain, S., & Mkamwa, T. 2010. The impact of diversity and equality management on firm performance: Beyond high performance work systems. Human Resource Management, 49(6), 977-998.

[2] Bracy, C. 2013. Closing the Tech Industry's Gender Gap Requires Better Data. National Public Radio (NPR). Retrieved August 20, 2015 from http://n.pr/1WRblQt

[3] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, Employment by detailed occupation. Retrieved August 18, 2015 from http://1.usa.gov/1NSVkTT

[4] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, Fastest growing occupations. Retrieved August 18, 2015 from http://www.bls.gov/emp/ep_table_203.htm

[5] Charette, R.N. 2013. Is There a U.S. IT Worker Shortage? IEEE Spectrum, 3 Sep 2013. Retrieved August 20, 2015 from http://bit.ly/1V8d8ic

[6] Creswell, J.W. 2008. Educational Research. Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Pearson Education, Upper Saddle River, New Jersey.

[7] Google Rise Awards. 2015. Retrieved August 19, 2015 from http://bit.ly/1zvi86f

[8] Khan, K.S., Kunz, R., Kleijnen, J., Antes, G. 2003. Five steps to conducting a systematic review. *J R Soc Med.* 96, 3, 118–121. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC539417/

[9] McGill, M.M., Decker, A., and Settle, A. 2015. Does Outreach Impact Choices of Major for Underrepresented Undergraduate Students?. In *Proceedings of the eleventh annual International Conference on International Computing Education Research* (ICER '15). ACM, New York, NY, USA, 71-80.

[10] Liston, C., Petereson, K., Ragan, V. 2008. Evaluating Promising Practices in Informal Information Technology (IT) Education for Girls: Phase III: Women in IT – Survey Results. Retrieved July 2015 from https://www.ncwit.org/sites/default/files/resources/ncwit-gsusaphaseiiireport_final.pdf.

[11] National Science Foundation CE21 Award List. (2013). Retrieved May 16, 2013 from http://1.usa.gov/1JksPh4

[12] National Science Foundation, Increasing Accessibility to computer science education across the U.S., 2013. Retreived July 22, 2015 from http://1.usa.gov/1NSUZkc

[13] Page, Scott E. The difference: How the power of diversity creates better groups, firms, schools, and societies. Princeton University Press, 2008.

[14] Patton, M. (2002). Qualitative Research & Evaluation Methods. USA: Sage Publications.

[15] Petticrew, M., and Roberts, H. 2006. Systematic Reviews in the Social Sciences: A practical guide. Blackwell Publishing. Retrieved August 15 from http://www.cebma.org/wp-content/uploads/Pettigrew-Roberts-SR-in-the-Soc-Sc.pdf

[16] Richtel, M. 2014. Reading, Writing, Arithmetic, and Lately, Coding. New York Times. Retrieved August 20, 2015 from http://www.nytimes.com/2014/05/11/us/reading-writing-arithmetic-and-lately-coding.html?_r=0

[17] Slater, S.F., Weigand, R.A., and Zwirlein, T.J. 2008. The business case for commitment to diversity. Business Horizons, Vol(51)3, May-June 2008, pp. 201-209.

[18] Spohrer, J. 2015. Empowering Makers in the Cognitive Era. *In Proceedings of the eleventh annual International Conference on International Computing Education Research* (ICER '15). ACM, New York, NY, USA, 1-1. DOI=http://dx.doi.org/10.1145/2787622.2787623

[19] Westervelt, E. 2015. Disrupting Tech's Diversity Problem with a Camp for Girls of Color. National Public Radio (NPR). Retrieved August 20, 2015 from http://www.npr.org/sections/ed/2015/08/17/432278262/hacking-tech-s-diversity-problem-black-girls-code