Connecting Evaluation and Computing Education Research: Why is it so Important?

Adrienne Decker
Rochester Institute of Technology
Rochester, NY, USA
adrienne.decker@rit.edu

Monica M. McGill
Knox College
Galesburg, IL, USA
mmmcgill@knox.edu

Jason Ravitz
Google, Inc.
San Francisco, CA, USA
ravitz@google.com

Eric Snow
SRI International
Menlo Park, CA, USA
eric.snow@sri.com

Rebecca Zarch
SageFox Consulting Group
Amherst, MA, USA
rzarch@sagefoxgroup.com

SUMMARY

With the growth of computing education research in the last decade, we have found a call for a strengthening of empiricism within the computing education research community. Computer science education researchers are being asked to focus not only the innovation that the research creates or the question it answers, but also on validating the claims we made about the work. In this session, we will explore the relationship between evaluation and computing education research and why it is so vital to the success of the many computing education initiatives underway. It will also help computing faculty engaged in computer science education research understand why it is essential to integrate evaluation and validation from the very first conceptual stages of their intervention programs.

CCS CONCEPTS

• **General and reference~Evaluation** • **Social and professional topics~Computer science education** • *Social and professional topics~Student assessment* • *Social and professional topics~K-12 education*

KEYWORDS

Evaluation; computer science education research; empiricism in computer science education research; assessment; validation

ACM Reference format:
A. Decker, M.M. McGill, J. Ravitz, E. Snow, R. Zarch. 2018. Connecting Evaluation and Computing Education Research: Why is it so Important? *In SIGCSE '18: 49th ACM Technical Symposium on Computer Science Education, Feb. 21–24, 2018, Baltimore, MD, USA.* ACM, NY, NY, USA, 2 pages. https://doi.org/10.1145/3159450.3159642

SIGCSE'18, February 21-24, 2018, Baltimore, MD, USA.

ACM ISBN 978-1-4503-5103-4/18/02.
DOI: http://dx.doi.org/10.1145/3159450.3159642

1 OBJECTIVE

The objective of this special session is to engage the community at SIGCSE in the dialog about evaluation and its value to computing education. The SIGCSE Technical Symposium brings forward a number of pedagogical techniques and advances that are sometimes empirically tested. As we grow as a community and the number of initiatives grow (particularly those focused on pre-university courses, curriculum, and content), it is incumbent upon us to ensure that the information that is being conveyed about a particular pedagogical idea, tool, or curricular intervention can be verified and substantiated.

A key difference between research and evaluation is that evaluation is typically focused on a specific program or initiative and is designed with stakeholders of that initiative. Research is traditionally designed to provide knowledge that can be used more broadly. Evaluation can provide recommendations to improve CS education programs and raise larger research-oriented questions about what is still unknown.

This session brings together a number of evaluators currently working in the computer science education area who will share strategies and tools for ensuring valid, verifiable results.

2 OUTLINE

2.1 Introduction (Decker & McGill)

The introduction to the session will provide an overview and purpose of the session. It will provide the basis for the argument that evaluation is an important and necessary part of the research process and needs to be integrated from the first conceptual stages of an experiment. These presenters worked to bring this group of evaluators together and have previously presented at SIGCSE about the lack of reporting and evaluation on various activities at the pre-college level [3, 4].

2.2 The Evaluation Wrecking Crew (Ravitz)

Google is an active participant in the Computer Science education research community, creating products (CS First), running programs (CS4HS), and fund initiatives (see Google.org). Google has developed a core set of research and evaluation

protocols and tools. We have also created opportunities for others to develop their evaluation capacity [5, 9]. We rely on research [6, 11] to inform the field and our evaluations, and we want our findings to be believable (valid, reliable, etc.) and point to outcomes that matter. Our guideposts include using theory, designing studies and developing methods to demystify data and reduce error [10]. Common practices include developing and studying interventions, often in partnership with educators [7], conducting literature reviews, validating measures, analyzing results, developing narratives, and reporting.

When it became clear that there were many different groups working on this same problem, The Evaluation Wrecking Crew was formed to work together on these problems. We used a number of processes to help prioritize our initiatives including an empowerment evaluation exercise, needs assessment and collaborative review of existing resources. This led to a collaborative initiative to organize evaluation resources to be used by others [13]

2.3 Valid Measures Matter (Snow)

Despite the spread of computing into the pre-college domain, development of high-quality assessments has not kept pace with the current rate of growth in computing programs and curricula and teachers face a number of challenges finding valid and reliable assessments to use in their classrooms [14].

Evidence-Centered Design (ECD) is a framework for assessment design and development that views assessment as a process of gathering evidence to support inferences about what a student knows and can do [8]. It provides a structure for an approach that incorporates validity evidence into the assessment design process, and is particularly useful when the knowledge/skills to be measured involve complex, multistep performances, such as those required in computational thinking practices. SRI will share our experiences, lessons and best practices from using ECD and contemporary test validity theory [1] to guide the development and validation of assessments of computational thinking practices for the *Exploring Computer Science* curriculum [2, 12].

2.4 SageFox Consulting Group (Zarch)

The NSF CS10K program "aims to have rigorous, academic computing courses taught in 10,000 high schools by 10,000 well-prepared teachers." We are challenged by the lack of a coordinated set of definitions, measures, and system for data collection. The Evaluator Working Group (EWG) coordinates data collection among project evaluators to count how many teachers are prepared through CS10K. The EWG has been successful at collecting aggregated information from all projects and has a meaningful set of participation indicators. Yet, the quality of the data varies between projects, and the high level of aggregation by project data precludes the EWG from doing any outcome-based analysis or pursing research interests such as understanding the quality of the professional development, the strength of implementation, and student learning. Ideally these items are captured at the project level, but there is no system for shared reporting.

2.5 Discussion and Q&A

There will be 30 minutes for discussion and questions.

3 EXPECTATIONS

The intended audience are those engaged in the scholarship of computing education. If not formally trained in educational research methods and evaluation, we are hopeful that those in the audience will begin their journey into understanding the important role evaluation plays in the research process. For those already engaged at some level in the research-evaluation pipeline, we hope to provide some perspective on the multiple types of evaluation that is going on within the community.

ACKNOWLEDGMENTS

The authors would like to acknowledge that this work is supported by the US National Science Foundation under Award Nos. DUE-1625335, DUE-1625005, CNS-1132232, CNS-1240625, DRL-1418149, CNS-1433065, and CNS-1640237.

REFERENCES

[1] American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 2014. Standards for educational and psychological testing. Retrieved August 31, 2017 from http://www.apa.org/science/programs/testing/standards.aspx

[2] M. Bienkowski, E. Snow, D.W. Rutstein, & S. Grover. 2015. *Assessment design patterns for computational thinking practices in secondary computer science: A first look* (SRI technical report). Retrieved August 31, 2017 from http://pact.sri.com/resources.html

[3] A. Decker & M.M. McGill. 2017. Pre-College Computing Outreach Research: Towards Improving the Practice. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17). ACM, New York, NY, USA, 153–158. https://doi.org/10.1145/3017680.3017744

[4] A. Decker, M.M. McGill, & A. Settle. 2016. Towards a Common Framework for Evaluating Computing Outreach Activities. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16). ACM, New York, NY, USA, 627–632. https://doi.org/10.1145/2839509.2844567

[5] D. Fetterman & J. Ravitz. 2017 *in press.* "A Google-Enhanced Empowerment Evaluation Approach." In Fetterman, D., Rodríguez-Campos, L., & Zukoski, A. (Eds.). *Collaborative, Participatory, and Empowerment Evaluation: Stakeholder Involvement Approaches to Evaluation.* New York, NY: Guilford Press. 105-117.

[6] Google Inc. & Gallup Inc. 2016. Diversity Gaps in Computer Science: Exploring the Underrepresentation of Girls, Blacks and Hispanics. Retrieved from August 31, 2017 http://goo.gl/PG34aH

[7] Erin C. Henrick, Paul Cobb, Kara Jackson, William R. Penuel, and Tiffany Clark. (2017). Assessing Research-Practice Partnerships: Five Dimensions of Effectiveness. New York, NY: William T. Grant Foundation. Retrieved November 20, 2017 from http://wtgrantfoundation.org/library/uploads/2017/10/Assessing-Research-Practice-Partnerships.pdf

[8] R.J. Mislevy and G.D. Haertel. 2006. Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice* 25, 4: 6–20.

[9] National Girls Collaborative Project. 2017. CS Outreach Program Evaluation Network (CS OPEN). Retrieved August 25, 2017 from https://ngcproject.org/cs-outreach-program-evaluation-network-cs-open

[10] J. Ravitz. 2002. Demystifying data about technology impacts in schools. Paper presented in National Educational Computing Conference. San Antonio, TX. Retrieved August 25, 2017 from http://academia.edu/1854297

[11] S.L. Smith et al. 2017, *in press.* Cracking the Code: The prevalence and nature of computer science depictions in media. Los Angeles, CA: USC Annenberg Media, Diversity & Social Change Initiative.

[12] E. Snow, C. Tate, D. Rutstein, M. Bienkowski. 2017. *Assessment design patterns for computational thinking practices in Exploring Computer Science* (SRI technical report). Menlo Park, CA: SRI International. Retrieved August 31, 2017 from http://pact.sri.com/resources.html

[13] STEM Evaluation Repository. Online resource. Retrieved 11/17/17. http://comm.eval.org/stemeducationandtraining/stem-tig-repository/viewrepository

[14] A. Yadav, D. Burkhart, D. Moix, E. Snow, P. Bandaru, and L. Clayborn. 2015. *Sowing the Seeds: A Landscape Study on Assessment in Secondary Computer Science Education.* Comp. Sci. Teachers Assn., NY, NY. Retrieved January 18, 2016 from http://csta.acm.org/Research/sub/Projects/ResearchFiles/AssessmentStudy2015.pdf