The Case for Acknowledging Subjectivity in CS Education Research Data

Monica M. McGill (Moderator)
CSEdResearch.org
Galesburg, IL, USA
monica@csedresearch.org

Jean Ryoo
Director of Research, Computer
Science Equity Project
UCLA
Los Angeles, CA, USA
jeanryoo@ucla.edu

Allison Scott
CEO
Kapor Center
Oakland, CA, USA
allison@kaporcenter.org

Chris Stephenson
Head of CS Education Strategy
Google
Mountain View, CA, USA
stephensonc@google.com

Jayce R. Warner
Texas Advanced Computing Center
University of Texas Austin
Austin, TX, USA
jwarner@tacc.utexas.edu

ABSTRACT

Is quantitative data collected by CS education researchers objective? If we combine data from a set of studies that measure the same type of intervention, will that really show us the strength of that intervention? Are qualitative studies really less rigorous than quantitative because the number of participants may be as low as one?

In this panel, we will first present the different types of studies that are most common in CS education research and provide a working definition for what we mean by various types of research methodologies (e.g., quantitative, qualitative, mixed methods). Drawing upon our experiences in the field of studying computing education, we will then explore some of the myths surrounding data, highlighting where evidence presented through research data is rigorous and, when not, how we (as researchers) can mitigate the risks of collecting and sharing data that is unsound in publications.

We encourage you to attend. After all, this panel has been recommended by four out of five computer science education researchers.

KEYWORDS

Data, quantitative, qualitative, mixed methods, education, objectivity, subjectivity

1 SUMMARY

In his 1954 book How to Lie with Statistics, Huff starts his introduction with this story.

> "There's a mighty lot of crime around here," said my father-in-law a little while after he moved from Iowa to California. And so there was–in the newspaper he read. It is one that overlooks no crime in its own area and has been known to give more attention to an Iowa murder than was given by the principal daily [newspaper] in the region in which it took place."

Fast-forward to 2022, and we have witnessed a barrage of socio-political memes touting all sorts of statistical "facts" that infiltrate our social media feeds, with many wondering how sound those statistics are. Are we "smarter" than the average consumer of such statistics? After all, many researchers hold advanced degrees–surely that means something? What about as a researcher–is the data that we collect, analyze, and present in our research papers more sound than the average meme? And, do "hard numbers" that say an intervention is sound (or not) actually reflect the truth about the intervention? How do our interpretations of that data affect learners in the classroom?

Akkerman et al. consider objectivity and neutrality within education research more broadly (beyond data), stating "The problem at stake is that educational research often portrays objectification as a neutral act, as if researchers can study whatever situation or process as a given object and then study the object dispassionately." [1, p. 3] Our panel will explore some of these notions and timely questions around subjectivity within the context of CS education research. Our answers will be shaped by years of wandering through our own "methodology and data journeys" in CS education research as consumers, authors, and reviewers. The panel will weave together a story about the critical need for researchers to understand the nuances between subjectivity and objectivity in data collection and reporting and how a better understanding of these differences can contribute to a stronger body of research and more accurate use of evidence. Attendees can be expected to be motivated and encouraged to deeply reflect on their own methodologies and data collection efforts and how they can improve them in future research.

2 PANEL STRUCTURE

Each panelist will have five minutes to introduce their positions and perspectives. This will include how their own experiences inform their computing education research and practice. After each panelist has presented, the moderator will pose prepared questions to the panelists while also fielding questions from attendees in a shared document and/or via chat. Questions may include:

- How can data shape the future of computing education?
- How can misinterpretation of data be harmful for learners, particularly those who are underrepresented and underserved?
- Can we take quantitative data at face value?
- How can qualitative data be rigorous?
- What is the most important challenge to CS education research that as a community we should begin to address?
- What advice do you have for novice and experienced education researchers?

We prefer a hybrid presentation, but we can adapt easily to any mode SIGCSE TS will support. If we present as a combination of hybrid, in-person, or virtual, we will adapt the format accordingly. Though the goals will not change, we will rely on digital tools like chat and shared, online documents to ensure that *all* attendees have the opportunity to provide feedback and pose questions.

2.1 Monica McGill (Moderator)

In addition to being formally trained in qualitative, quantitative and mixed methods education research, I have had the unique experience of reviewing thousands of abstracts and overseeing curation of data from hundreds of K-12 CS education research articles across a variety of topics, interventions, and design methodologies. I will share my experiences of seeing gaps in data reporting as well as the harms these gaps can cause. I will also draw upon my past studies and exploration of evidence in data–my curiosity in how these gaps are being mitigated in other fields has led me to consider how these practices could enable a higher quality body of evidence across CS education research [3]. I will also touch on what shapes our attention in research studies and how that process may be flawed, why it may be flawed, and how we may be able to contribute to mitigating the risks of the subjective interpretation of data.

2.2 Jean Ryoo

Dr. Ryoo is the Director of Research of the UCLA CS Equity Project. Her work focuses on building research-practice partnerships with administrators, educators, and students, attending to issues of equity in teaching and learning through a sociocultural theoretical lens. Ryoo will add to the conversation on how quantitative and qualitative research methods might better work in dialogue, building on her experiences conducting critical ethnographic studies, qualitative case studies, and participant observations of K-12 CS education spaces. Ryoo was trained in qualitative research methods by Drs. Frederick Erickson and Kris Gutiérrez, whose work emphasizes critical examinations of power and agency in qualitative methods.

2.3 Allison Scott

Dr. Scott's research in CS education focuses on examining racial, gender, and socioeconomic gaps in interest, access, participation, and success in K-12 CS education. To clearly understand the challenges to equity that exist, and how best to address inequities, it is critical to (a) utilize rigorous frameworks for understanding underlying causes of inequality in education, (b) ask the rigorous and robust questions of the data, and (c) utilize multiple sources of data (including standardized tests, surveys, interviews). We also must consider the perspectives and orientations brought to bear on the research process and the demographic backgrounds of researchers, as well as the sources of information that are viewed as "valuable" or accurate in the examination of data. Understanding the subjective nature of education research, and acknowledging the limitations of any approach, are critical as we seek to create greater knowledge and understanding to move the needle on equity in CS education.

2.4 Chris Stephenson

Although I have conducted a significant amount of research (quantitative, qualitative, and mixed-methods) over the long course of my career, I make no claims to being a professional or exemplary researcher. But I am a voracious consumer of CS education research, precisely because my job has always depended upon it. Making data-driven decisions has always been a critical aspect of my responsibilities precisely because making decisions about possible interventions in CS education in the absence of reliable research is risky at best and a mug's game at worst. The last four years, I have been privileged to lead the CS Education Research grants program at Google. In this role, I read a great many research proposals each year and ultimately must choose the ones that are most rigorous and best meet our funding criteria. This is a challenging task because there is (finally) a great deal of high-quality research being proposed and conducted. But when compared to other more mature fields of education research, there is a concerning gap in our ability, or perhaps willingness, to grapple with the challenging complexities of methodology, objectivity, and truth itself. My hope is to bring that impassioned consumer perspective to this conversation and to draw from my own research experiences evidence of why we need to be having these conversations.

2.5 Jayce Warner

Dr. Warner is Research Associate at the Texas Advanced Computing Center at The University of Texas at Austin. He is an applied statistician whose work focuses on quantifying inequities in computing education and using multilevel modeling to assess student and teacher outcomes. He will contribute to the conversation by providing insight into how quantitative data are sometimes misrepresented and misinterpreted. This includes results of experimental studies that use advanced statistical methods as well as basic descriptive statistics that are used to identify and track inequities (e.g., disparities in enrollment rates) in CS education over time.

REFERENCES

[1] Sanne F Akkerman, Arthur Bakker, and William R Penuel. 2021. Relevance of Educational Research: An Ontological Conceptualization. *Educational Researcher* (2021), 0013189X211028239.
[2] Darrell Huff. 1954. *How to Lie with Statistics.* W.W. Norton & Company.
[3] Monica M McGill, Tom McKlin, and Errol Kaylor. 2019. Defining what empirically works best: Dynamic generation of meta-analysis for computer science education. In *Proceedings of the 2019 ACM Conference on ICER.* 199–207.