2023 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT) | 979-8-3503-2832-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/RESPECT60069.2023.00012

Exposing Oversights in the Collection, Analysis, and Reporting of Demographic Data

Tom McKlin
The Findings Group
Atlanta, GA, USA
tom@thefindingsgroup.org

Clarissa Thompson
Department of Psychology
Kent State University
Kent, OH, USA
cthomp77@kent.edu

Susan Fisk
Department of Sociology and Criminology
Kent State University
Kent, OH, USA
sfisk@kent.edu

Audrey Rorrer
Computer Science
University of North Carolina, Charlotte
Charlotte, NC, USA
audrey.rorrer@uncc.edu

Tiffany Barnes
Computer Science
North Carolina State University
Raleigh, NC, USA
tmbarnes@ncsu.edu

***Abstract*—We propose engaging RESPECT attendees in a conversation on the use of demographic data for BPC research and evaluation. We present the decisions that teams face at four stages of a study: planning, collection, analysis, and reporting. Decisions at each stage may inadvertently hide, stifle, diminish, or erase important perspectives and findings; other decisions may expose confidential data. We seek perspectives that strengthen future research and evaluation work and prevent it from being dismissed a decade from now as insensitive or archaic.**

Keywords—demographics, broadening participation in computing, survey research

I. Introduction

This work encourages research/evaluation teams to engage in a process of using demographic data that allows teams to best represent the diversity of their participants. We expose the difficulties of demographic data, present guidelines, and encourage discussion to improve research/evaluation practice that honors the diversity of our participants while also providing rigorous findings. Importantly, this project discusses the pros and cons of different approaches in service of the specific questions we seek to answer rather than presenting a single, ideal approach. We discuss the use of demographic data at four stages of research/evaluation practices: planning, collection, analysis, and reporting.

II. Planning

At the outset of a project, teams select demographic elements that are broad enough to capture identities critical to the study while also reducing participant survey fatigue. They do so within an increasingly more diverse context. Gaither [3] notes a rapid increase in people identifying as two or more races: in Britain, almost 1 in 10 children are of mixed racial descent while the number of people identifying as multiracial in Canada increased by 25% between 2001 and 2006 and by over a third in the U.S between 2000 and 2010. Further, a Gallup study [3] found that 7.1% of U.S. adults identified as non-binary in 2021, double the percentage from 2012. One in five adults born after 1997 (Gen Z) identify as non-binary. In fact, the actual percentage of those in the Gen Z category that identify as non-binary or non-cisgender may be much larger because the Gallup study collected data only on Gen Z adults which comprises about 20% of the actual Gen Z population. Menier, Zarch, and Sexton [5] argue that transgender and nonbinary learners may be discouraged from participating in computing and encourage researchers and evaluators to include sexual identity and orientation in their data collection efforts.

Importantly, the planning phase is the stage in the process where teams make decisions about demographic data to omit. One potential approach is to engage in an exercise in which the team identifies three categories of demographic data: 1) demographic items that must be collected, 2) items that the team knows it will not collect (e.g. computer science education rarely collects data on religions/spiritual beliefs), and 3) items that fall in between the first two. Teams review items in the final category and formally decide which items to omit rather than omitting items by never considering them; that is, we encourage teams to justify their omissions.

III. Collection

Once a team decides on the demographic categories to collect, it must then operationalize each demographic item by selecting appropriate response options. Call et al. [1] highlight the need to balance respect for participants with generalizability and describe the spectra of collection approaches ranging from the most inclusive and least prescriptive (open-text responses for all demographic items) to the least inclusive and most prescriptive (forced, single-answer choice to a limited set of demographic response options). Importantly, they cite Strunk & Hoover [6] who argue that forcing participants to select an identity that does not apply to them is an act of oppression and reinforces the notion that research does not recognize, accept, or value their identity. Giving participants a set of items and response options that allows them to accurately report their identity is a step toward building trust in the research.

Call et al. [1] consolidate guidelines to consider when designing demographic data collection instruments. Some of those are:

979-8-3503-2832-5/23/$31.00 ©2023 IEEE
DOI 10.1109/RESPECT60069.2023.00012

- Document the rationale for including and excluding certain demographic variables.
- Consider demographic variables that might be relevant for future research even if it's not directly relevant to the current study.
- Do not use the "other" label when listing options for demographic items because it connotes abnormality; use "not listed" or "prefer to self-describe" instead.
- When asking about gender, avoid only binary options and consider options like non-binary, genderqueer, etc.
- Allow participants to select as many response options as they want.
- Use open-ended demographic items so participants are not limited to inaccurate options.
- Place demographic items at the end so participants can choose the demographic information to disclose based on the context of the information they have already shared.
- Allow participants to express concerns or questions about the methods.

IV. Analysis

While data collection seeks to account for all identities in the survey sample, analysis often simplifies and reduces those options and may erase diversity beyond a single descriptor. In computer science education research and evaluation, we often want to know how our interventions work for those who have historically been excluded from computer science education resources such as Latino/a and Black or African American students. Therefore, we collect data on race and ethnicity and run inferential statistics to determine whether there is a significant difference among participants by race/ethnicity categories on a particular outcome variable (e.g. belongingness, self efficacy, intention to persist).

Consider a survey participant who selects both Asian and Black racial identities. Oftentimes, we simply lump those who identify as more than one race into a "more than one" category to create groups large enough for statistical comparison, a practice that enables statistical comparison but assigns an identity that the participant did not select.

In some cases, a small sample size forces researchers to analyze only two groups: those who have been historically included in CS education and those who have not. In CS, white and Asian students often comprise the "included" category while all others comprise the comparison group. Here, almost all diversity from the original instrument is erased. Further, participants like the example Asian and Black participant may have identities associated with both the "included" and "excluded" categories which forces the researcher to decide how the participant should be grouped for analysis. Some survey designers anticipate analytic reduction and ask, "if you had to select the one race/ethnicity category that best describes you, which would you choose?" This signals that reduction will be necessary and asks the participant to make the decision. However, this practice may cause psychological conflict for individuals who are multiracial if they are being asked to choose a singular category that does not align with their identity [2].

Call et al. [1] provide additional guidelines to apply at the analysis stage:

- Avoid collapsing demographic groups with a small sample size into an "other" or "minority" variable. If necessary, justify the decision and describe its limitations.
- For race/ethnicity analyses, avoid using white as the reference group; it assumes that white is the standard to which other racial/ethnic groups should be normed.

V. Reporting

While analysis focuses on using demographic data as a set of grouping or contextual variables used in inferential statistics, reporting covers the display of inferential findings and also the overall presentation of methods and descriptive statistics. Call et al., [1] provide further guidelines to consider during the reporting stage:

- Report the full sample demographics even for demographic groups not included in the analysis. The team must navigate the tension of maintaining participant anonymity with data transparency.
- Report intersectional identities relevant to the study.
- Include a positionality statement [2].
- Report limitations.

VI. Conclusion

Demographic data collection involves more than simply adding a standard set of demographic items to a survey; instead, it is a multi-step process applied during each phase of the study, and it requires additional resources and expertise. We encourage research/evaluation teams to apply these or similar guidelines as a formal practice governing the collection and use of demographic data at all four stages of a study.

References

[1] Call, C. C., Eckstrand, K. L., Kasparek, S. W., Boness, C. L., Blatt, L., Jamal-Orozco, N., Novacek, D. M., Foti, D., & Scholars for Elevating Equity and Diversity (SEED). (2022). An Ethics and Social-Justice Approach to Collecting and Using Demographic Data for Psychological Researchers. *Perspectives on Psychological Science*, 174569162211373. https://doi.org/10.1177/17456916221137350

[2] Darwin Holmes, A. G. (2020). Researcher Positionality—A Consideration of Its Influence and Place in Qualitative Research—A New Researcher Guide. *Shanlax International Journal of Education*, *8*(4), 1–10. https://doi.org/10.34293/education.v8i4.3232

[3] Gaither, S. E. (2015). "Mixed" Results: Multiracial Research and Identity Explorations. *Current Directions in Psychological Science*, *24*(2), 114–119. https://doi.org/10.1177/0963721414558115

[4] Jones, J.M., "LGBT Identification in U.S. Ticks up to 7.1%," Gallup, February 17 2022. https://news.gallup.com/poll/389792/lgbt-identification-ticks-up.aspx

[5] Menier, A., Zarch, R., & Sexton, S. (2021). Broadening Gender in Computing for Transgender and Nonbinary Learners. *RESPECT*, 1–5. https://doi.org/10.1109/RESPECT51740.2021.9620612

[6] Strunk, K. K., & Locke, L. A. (Eds.). (2019). *Research Methods for Social Justice and Equity in Education*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05900-2