Research in Engineering Education Symposium &
Australasian Association for Engineering Education Conference
5 - 8 December, 2021 - Perth, WA

A Critique of Quantitative Methodologies to yield Critical Quantitative Methods in Engineering Education Research (EER)

Desen S. Ozkan [a]; David P. Reeping [b], Cynthia Hampton [c], and Cherie Edwards [d]
Tufts University[a], University of Cincinnati[b], University of Colorado Boulder[c], Virginia Commonwealth University[d]
desen.ozkan@tufts.edu

ABSTRACT

CONTEXT
A critical examination of quantitative research methods has been ongoing. Feminist and critical theorists have long problematized quantitative methods for their alleged 'view from nowhere' that offer neutral insights to the questions of inquiry (Haraway, 1988; Nagel, 1989). However, scholars have shown that these seemingly reproducible research methods have negative implications because they do not always consider the researcher assumptions or decisions that go into their design (Zuberi & Bonilla-Silva, 2008; Walter & Andersen, 2016).

PURPOSE
We position this paper as a bridge between the quantitative methods that are prized for their ability to offer scalability, order, and comparison with the critical methods that emphasize power relations and describe the historical and institutional context. We seek to examine the assumptions and decisions embedded within quantitative research methods by drawing on the analytics of power and knowledge.

APPROACH
We conducted a qualitative content analysis of a purposeful sample of recent engineering education research articles. Our review consists of recently published quantitative research articles from the Journal of Engineering Education (JEE), Australian Journal of Engineering Education (AJEE), conference papers from ASEE, as well as articles from Race Ethnicity and Education. This is not an exhaustive review of researcher decision-making in quantitative research but allows us to examine the primary modes of decision-making through a lens of power and knowledge relations.

ACTUAL OUTCOMES
This review results in a synthesis of the considerations that engineering education researchers make when conducting quantitative research. Our focus is not necessarily on reporting standards for specific methods like cluster analysis or regression. Instead, the anticipated outcome is a set of themes regarding how engineering education research can integrate criticality into the quantitative perspective. We, like others in engineering education, are cautious of research methods that lack transparency. These approaches reduce heterogeneous populations of engineering students and faculty to seemingly insular characteristics for the purpose of offering generalizable claims. Through this work we do not seek to promote qualitative research methods over quantitative methods but work to recontextualize researcher decisions through an examination of power in the production of knowledge.

CONCLUSIONS
Ultimately, research is a human endeavour and as such is entwined with complexities of power and knowledge relations. Through an analysis of decision-making in the quantitative research design process, we develop a practice of critical examination as researchers to make inferences about populations who are different from ourselves.

KEYWORDS

Research methods, Power/knowledge, Decision-making, Quantitative research

Introduction

Quantitative research has the power to find commonalities across different contexts. Reproducible quantitative methods hold the capacity to generate insights at scale and discern patterns across seemingly disparate localities, promoting generalization as their purpose. As such, many policy decisions are spurned from large-n quantitative studies (Gillborn, Warmington, & Demack, 2018). Although reproducibility and repeatability are prioritized as the pillars of most quantitative research, these ideals assume that the researcher and their implicit and explicit decisions are without consequence to the inferences drawn from the results. Past and present research teach us about the harmful consequences these seemingly repeatable and reproducible insights have had on minoritized populations (Zuberi & Bonilla-Silva, 2008; Walter & Andersen, 2016).

We do not aim to critique the quantitative methods from a technical standpoint in this article. Instead, we examine a subset of the countless researcher decisions that are often taken for granted in quantitative research. These decisions are not always explicit; some decisions are implicit assumptions, whereas others follow accepted disciplinary standards. However, although the decisions are accepted and normalized, we offer this critique on methodological practices in engineering education research to reassess and recontextualize researcher decisions.

Before we examine researcher decisions in present day quantitative research, we must acknowledge its historical formation. To historicize quantitative methods and quantification is to interrogate the cultural historical context of present research practices, which provide us with a deeper understanding of the power relations embedded in knowledge produced by quantitative methods (Foucault, 1977).

Quantitative Methodologies

Quantitative methodologies abstract information across various local contexts to standardize insights broadly. These methodologies allow the “messiness of local context [to] be removed, ordered, scaled, compared, and rearranged as required by researchers” (Walter & Andersen, 2016). This universality is a common assumption of quantitative techniques. Notably, in the university setting, “mathematics [...] has long been almost synonymous with rigor and universality” (Porter, 1995, p. VIII). The implications that quantitative methodologies have more power in the production of knowledge is especially important as they have been used in policy and administrative interventions. Foucault and others have argued that quantification and statistics have been “an agency for acting on people, exercising power over them” (Walter & Andersen, 2016; Porter, 1995, p. 78).

While “numbers turn people into objects to be manipulated,” quantification provides communication lines across disciplinary and even national boundaries (Porter, 1995, p. 78). Similar to economic exchange, in which money can be converted across borders, “numbers are the medium through which dissimilar desires, needs, and expectations are somehow made commensurable” (Porter, 1995, p. 86). The seeming ability for these insights to efficiently travel and scale across knowledge and cultural boundaries was not unrelated to the increased effort through the 1960s and 1970s to incorporate quantitative measures in public decision-making (Porter, 1995; Gillborn et al., 2018).

Engineering Education Research

A critical discussion of research methodologies in engineering education is not new. Engineering education researchers have published some critical examinations of quantitative methods in engineering education (Douglas, Rynearson, Purzer, & Strobel, 2016; Godwin, 2020; Holly, 2020). Alison Godwin, in her recent editorial in *Studies in Engineering Education*

 https://doi.org/10.52202/066488-0085

provides an antiracist critique of quantitative research methods, specifically addressing notions of neutrality and objectivity in quantitative research. She notes that as researchers, “we leave our research fingerprints all over our work” (2020, p. 79). Moreover, Godwin et al. (2021) promote a distinction between person-centered analyses and variable-centered analyses.

Person-centered analyses can be identified by how individuals are treated during analytical procedures. A person-centered analysis treats the individual as a whole, one indivisible unit, to preserve variation. Thus, the response to an outlier in a person-centered analysis is not to remove it; instead, extreme values are put in conversation with other individuals’ responses as holistic comparisons. Moreover, person-centered analyses can incorporate data-driven methods. These methods take an inductive approach by focusing on relationships as given in the data as opposed to theoretically-driven frameworks. Person-centered approaches can use data-driven approaches to find hidden groups or structures to evaluate the patterns found in the data; however, not all data-driven approaches are person centered. In a broader categorization, variable-centered analyses - i.e., methods that prioritize predefined categories by mapping patterns among the chosen variables to them. Some data-driven approaches are designed to assign individuals to predetermined categories, but others can help researchers find subgroups that preserve variation within the individual.

We use this paper to build on and expand this work by Godwin (2020) and her coauthors (2021). We draw insights from neighboring disciplines who have implemented critical quantitative methods. In 2018, a special issue in *Race, Ethnicity, and Education* was published on QuantCrit, which integrates critical race theory (Garcia, López, & Vélez, 2018) and quantitative research methods. Moreover, *Indigenous Statistics* describes the incommensurability of indigenous ways of knowing with some statistical methods that have been imposed on indigenous communities (Walter & Andersen, 2016). By centering everyday experience in its sociopolitical context, these two examples of critical quantitative methods produce insights by addressing power/knowledge relations of traditional quantitative research.

Quantitative Methods and Social Justice

Statistical-based research, similar to power and knowledge, is not neutral or value-free (Gillborn, Warmington, & Demack, 2018). Statistics commonly used in quantitative methods are steeped in a history spanning decades of eugenics through the works of scientists such as Galson, Pearson, and Fisher (Clayton, 2020) and other inferences justifying the othering and subjugation of Black, Indigenous, disabled, and countless peoples seen by society as inferior on a scale of whiteness. The long-lasting impacts of the power imposed by researchers for knowledge production that supported the aims of white supremacy is present in all facets of inequity within the United States. The knowledge associated with statistical reasoning and applied through quantitative methods provides power to the beholder of said methods. As a foundation of quantitative methods, the use of statistical tests, treatments, and deduction automatically places a degree of power for those conducting said analyses. Other inherent characteristics within quantitative research is the notion of validity and reliability, however these signifiers often are limited to the validity and reliability of tests rather than measurement reliability related to the data alone (Fan, 2013). Within the context of social justice, the distillation of data and inferences made have been counter to the liberatory nature of justice work. Within engineering education, this countering has resulted in an overall lens within quantitative work that does not for instance take into consideration the “agency and asset based” (Holly, 2020; p.629) experiences of Black people. However, there exists paths forward that acknowledge, apply, and take deference to cultural sensitivity in steps of quantitative research from data collection and stakeholder involvement to recruitment of participants (Awad, Patall, Rackley, & Reilly 2016).

https://doi.org/10.52202/066488-0085

Methods

We performed a qualitative content review of purposefully chosen, recent STEM education literature to identify and examine researcher decisions and justifications in their quantitative methodologies. Our review focused primarily on the archives of the *Journal of Engineering Education*, ASEE conference publications, *Race Ethnicity and Education*, and the *Australian Journal of Engineering Education*.

> RQ1 - What are the assumptions or decisions that researchers disclose in quantitative research publications?
>
> RQ2 - How do quantitative researchers justify their decision-making processes in their research processes?

This paper is not exhaustive in its review of quantitative researcher decisions in the STEM or specifically engineering education literature, and we hope to present a more systematic review in our following work. This preliminary review served to identify a preliminary list of different types of researcher decisions and justifications. The decisions around data collection include those around response rates and representativeness in sampling. The decisions regarding data analysis concern reliability, specifically measures of internal consistency as well as assumptions and enactments of normality in the datasets.

Notably, we do not focus explicitly on researcher decisions that are outside of the research design, analysis, and results sections, which can be a limitation to the work. We do acknowledge that researcher decisions are not limited to these sections. The examinations of these features of quantitative research are not exhaustive of all researcher decisions, but provide a starting point from which to conduct a more comprehensive review of quantitative engineering education research.

Examining Researcher Decisions

In this preliminary review of the research, we examined empirical engineering education research articles with a quantitative focus to understand what decisions and justifications to those decisions researchers pointed to in their work.

Decisions in Collecting Data

The quantitative studies examined in this article fall into two categories for data collection, either they cited a validated survey instrument as their primary tool for data collection or they used data from an existing data set like the High School Longitudinal Study (HSLS). The authors who used a validated instrument included citations from research that previously published scales. Here, we highlight the decisions of representation and categorizations of participants.

Representative samples and categorizations

For the studies in which the authors administered surveys, they provided the response rates and consequent representativeness of their samples. Johnson and Wislar (2012) explain that 60% is often used as a threshold for response rate but caution no scientifically credible rationale that substantiates this threshold - or any others presented in the literature. Nonetheless, the limits of these response rates often impact minoritized students the most.

In one paper, the authors cited a 15% response rate from the student body. Because of this lower response rate their study had a “small number of minoritized respondents,” to which they were “unable to disaggregate the data by race” (Jensen & Cross, 2021, p. 378). These authors do note that this small number limited their ability to perform meaningful statistical analysis. This type of awareness is not universal, as another paper discusses their demographic data collection as categorizing students as White and “underrepresented minority race other than White” (Jackson, Mentzer & Kramer-Bottiglio, 2021, p. 149) -

https://doi.org/10.52202/066488-0085

although, they did collect the finer demographic characteristics initially. This solution to binning data such that the groupings conform to conventional statistical practices has its own pitfalls. Shafer et al. (2021) show how regrouping students in such a fashion can mask disparities by subgroups within the aggregate category. The issue here is Simpson's Paradox; a trend can disappear or even reverse depending on how data are binned.

In a second journal article, the authors stated that "Unfortunately, the data collected do not contain demographic information for the students; thus, our analysis focuses on the population as a whole" (Chen, West, & Zilles, 2019, p. 578). These authors described their method for estimating "the demographic composition of the students in the data" which was to "[report] the demographic information of undergraduates who graduated with degrees in each discipline" (Chen et al., 2019, p. 578). While this method for estimation may have captured accurate demographic data, there is the concern that when they excluded various categories of data, they were excluding students from one demographic or gender. Even though the estimation was accepted, at least by the peer reviewers, the potential disparate impact of who was included and excluded cannot be identified. Thus, we caution assumptions around student representation in samples, especially when students are unable to describe who they are regarding data that represent them.

Even for studies that do collect data regarding minoritized status of students, the question of representativeness is critical. A recent ASEE publication notes that they used "chi-squared analyses [...] to determine the significance of the discrepancies in representation rates of marginalized students" (Bowen, Johnson, & Powell, 2021, p. 6). The authors calculated whether the minoritized students were participating in the study at different levels of representation than the students who are traditionally served by engineering departments. Additionally, this paper noted that their sample "must actually exhibit *better* representation rates or quantitative outcomes than non-marginalized populations to a statistically significant degree" (Bowen et al., 2021, p. 6). These authors note this decision as striving for equity rather than equality.

Finally, in an article from the journal, *Race Ethnicity and Education,* the author discusses their decision in sample restriction to Black and White women "as a racial comparison." The author goes on to say "this comparison is made not to normalize Whiteness, but as a way to indirectly understand how power drives policy decisions" (Campbell, 2020, p. 6). The differences in how authors discuss race and gender variables (and in other studies: disability status or sexuality) reveal some assumptions or accepted normalizations of what they represent. In quantifying each variable, we reduce the social relationships that make up the categorization. Additionally, this quantification stabilizes concepts that are not necessarily stable. For race:

> *'It' is not a thing, a reified object that can be measured as if it were a simple biological entity. Race is a construction, a set of fully social relationships.'" (Apple 2001, p. 204 emphasis kept).*

Further, the use of race in this fashion is what Zuberi (2008) calls a "form of racial reasoning" (p. 131). While seemingly stable categorizations are useful in helping insights travel, methodological transparency and awareness to the limits of such categorical reification is necessary to ensure that insights do not perpetuate harm onto vulnerable populations (Gillborn et al., 2018).

Decisions in Analyzing Data

In the sections pertaining to data analysis, including but not limited to sections titled data analysis, we observe several commonalities in the authors' decisions and justifications for their analyses. One of these decisions is to emphasize that various statistics match those found in previous literature. Notably, data analysis occurs throughout multiple sections of a research article. From scholars who discuss data cleaning as analysis (D'Ignazio & Klein,

https://doi.org/10.52202/066488-0085

2020) to writing results and discussion as a form of analysis, the process of data analysis is not bound to the section titled analysis. Here, we discuss the decisions of determining internal consistency, normality, and the treatment of outliers.

Internal Consistency

Internal consistency has to do with how well a set of items measuring a certain construct produces similar scores when administered. This concept has been applied in several different ways, such as average inter-item correlation, closeness to unidimensionality, and internal consistency reliability (Tang et al., 2014). In engineering education, internal consistency often manifests through Cronbach's alpha. Several studies point to the similarity of internal consistency coefficients between their study and previous studies. In one study, the authors "measured Cronbach's alpha scores for the [X] subscales comparable to previous studies using the short form (Henry & Crawford, 2005; Osman et al., 2012)" (Jensen & Cross, 2021, p. 377). Similarly, authors note the Cronbach alpha score for a different subscale to be "consistent with previous work (0.905 compared with 0.84 and 0.89) (Jones et al., 2010)" (Jensen & Cross, 2021, p. 377). In a different example of internal consistency, authors used "ωh, which relaxes assumptions about the structure of measurement scales and is more appropriate than coefficient α in most cases (Zinbarg et al., 2005)" (Jackson et al., 2021, p. 152). These practices around reliability coefficients are often required in these research practices. Jensen and Cross provide transparency and robustness as they report the measures of internal consistency for each subscale dataset. Additionally, they compare these internal consistency values to several previous studies with similar values to justify the consistency of their data. In Jackson et al., authors note that they use the omega coefficient, ωh, to evaluate consistency instead of Cronbach's alpha, which they explain is an improved practice. They cite Zinbarg et al. to justify this decision.

Internal consistency is not the aim of Cronbach's alpha, however. As Sijtsma (2009) notes, Cronbach (1951) originally wrote that alpha provides a lower bound to the "true reliability" (p. 299) and does not say much of anything about the internal consistency of the items. Moreover, alpha is not a property of the itemset themselves; instead, it is the property of the itemset within the context of a specific population (Miller, 1995; Thompson & Vacha-Haase, 2000). Thus, comparing alphas to previous literature without considering the differing contexts can be misleading. These decisions and their justifications are commonly accepted practices in quantitative research nonetheless (Dunn et al., 2014).

The practice of testing for reliability has been one that goes further back than Cronbach in 1952. However, as we think about non-homogenous populations, we ask if there is a disparate impact with who the error variance explains and if this practice measuring internal consistency should remain homogenizing. A .70 coefficient for internal consistency at minimum is generally accepted (Tavakol & Dennick, 2011). This 0.70 implies that 70% of the variance in the scores is reliable variance and 30% is error variance. We question whether there are overrepresented demographics of students in the error variance. Through internal consistency calculations, do researchers unknowingly elevate the dominant student population's scores? Additionally, as the common practice of evaluating a consistency coefficient is to cite previous scholarship that may or may not have adequate demographic and gender representation, what does that say around whose survey answers produce dominant knowledge in the education research?

Normality and Outliers

The next feature of quantitative research is the way that authors evaluate normality in their data. Notably, not all quantitative research articles work under the assumption of a normal distribution in their data, but many do with statistical tests to support an assumption of normality. However, to fit into a normal distribution, authors have discussed various methods for excluding data that fall outside of the necessary parameters. Godwin et al. (2021) note that the treatment of outliers is a decision that can affect minoritized individuals most severely. Because a subset of the data does not conform to the appropriate distribution,

https://doi.org/10.52202/066488-0085

students outside of the bell curve's main body are not included in the analysis - erasing their contribution to the study.

Two concepts are commonly used to assess the normality of a variable, skewness and kurtosis. Skewness is a measure of asymmetry, examining the extent to which side of the distribution has a longer tail than the other. Kurtosis measures the tailedness of the variable's distribution, which corresponds to how flat or peaked the distribution is. The cutoffs for these variable distribution properties vary. In one example, the authors "excluded asynchronous exams whose score distribution's kurtosis was more than 10" (Chen et al., 2019, p. 578). For variables with at least approximately normal properties, outliers are still threats to the performance of classical statistical tests. These authors note that their filter "eliminates asynchronous exams that have large deviations from the mean, which could have unstable effects on the regression coefficients" (Chen et al., 2019, p. 578). These types of justifications are worrisome as we question who is excluded when the data are forcefit to statistical normality.

In addition, this article notes another group of records they excluded as those that "Were outside the corresponding exam periods." They provide a footnote that states "in exceptional circumstances such as long-term illness, students take an asynchronous exam outside the normal exam period" (Chen et al., 2019, p. 588). Long-term illness or chronic illness are recognized by the American with Disabilities Act (ADA) as a disability. For the authors to note that these students were excluded because of their disability is jarring and reinforces the ableism all too common in the university setting (Brown & Leigh, 2020) as well as the ableism in research that upholds notions of normality but from a statistical perspective and from a societal perspective (Wong, 2020; Hendren, 2019). As Godwin et al. (2021) would contend, this approach embraces a variable-centered approach to analysis.

Person-centered approaches embrace heterogeneity in their data and seek to preserve the variation in individual responses within the measures. In a different paper, the authors collect demographic and gender information that they use to disaggregate their data. Notably, these authors note that the research focus is to understand "differential experiences of students based on their characteristics and contexts instead of trying to normalise engagement 'for "average" students' (Polmear, Chau, & Simmons, 2020, p. 66). In their analysis, they conduct the "Levene's test [...] to examine homogeneity of error variances assumption" (Polmear et al., 2020, p.68). Additionally, they note that a "Histogram, Q-Q normal probability plot, skewness, and kurtosis were constructed and computed to examine if data met the normality assumption (Hair et al. 2019)" (Polmear et al., 2020, p.68). Specifically, because these authors do these analyses with a disaggregated data set regarding race and gender, they can preserve different student experiences rather than reporting findings that homogenize student experiences.

Positionality

The last decision we want to highlight is the existence of researcher positionality in articles. While positionality is rare in quantitative research broadly, in engineering education research it has yet to be introduced. In a previous systematic literature review of the Journal of Engineering Education, European Journal of Engineering Education, and International Journal of Engineering Education from 2008 to 2020, we did not find positionality discussed in a single quantitative research article (Hampton, Reeping, & Ozkan, 2021). While this decision can also be an aspect of the peer review process and the journal's priorities, we note its absence in EER because positionality has appeared in other STEM education research articles.

In an article by Young and Cunningham in the journal, *Investigations of Mathematics Learning*, the authors "note [that] our positionalities, as Black female researchers who were once young high schoolers, provide us experiential knowledge that guided our analytical

https://doi.org/10.52202/066488-0085

understanding of the Black female learners in this study" (Young & Cunningham, 2021, p. 38). These authors provide their positionality at the end of the article.

We note that our positionalities, as qualitative, quantitative, visual, and mixed method researchers in engineering and medical education fields, have provided a broad range of research experiences each with countless opportunities to make and justify methodological decisions. Our positionalities as Black and White researchers, male and female researchers, straight and gay researchers, have led us to identify different assumptions with each other and in published articles. However, each author resides in the United States, and we acknowledge our US-centric scope in this review of the research.

Research is a series of decisions made by researchers with different positions in society. Generally, academic research has been produced by a non-representative demographic of the Global North. As we continue to rely on established norms and past research, we find it necessary to critique the tools that have been handed down to us. While quantitative research has opened up a vast array of possible insights, which is scaled further with introduction of machine learning research methods that further reify dominant narratives without critical contextualization. Lastly, in this paper we focus on examining researcher decisions in quantitative research in the field of engineering education, but many of these issues also exist in qualitative research as well (Holly, 2020).

Conclusion

In this work, we aimed to critically examine methodological decisions and subsequent justifications provided by quantitative researchers in engineering education research articles. To accomplish this aim, we reviewed engineering education literature to identify these patterns to provide examples of critical quantitative research methods for future quantitative researchers.

While research in the space of problematizing researcher decisions is not new, we situate this work within the engineering education research discipline to shed light on the power dynamics that exist in researcher-subject relations. Research is the culmination of a number of related human decisions, often justified by expertise, past research, standards, among other disciplinary practices. These decisions and their justifications have histories that require attention as seek to disrupt inequities that can stem from knowledge production that serves the dominant groups in power. In engineering education research, researchers seek out truths with respect to student learning, faculty learning, professional practice, institutional systems, among countless other aspects of the engineering environment. These countless decisions and their justifications are entangled in various histories that are important to acknowledge in the way we as researchers carry out education research.

References

Apple, M. W. 2001. *Educating the 'Right' Way: Markets, Standards, God, and Inequality.* New York: Routledge Falmer.

*Bowen, C. L., & Johnson, A. W., & Powell, K. G. (2021, July), Critical Analyses of Representation and Success Rates of Marginalized Undergraduate Students in Aerospace Engineering Paper presented at *2021 ASEE Virtual Annual Conference Content Access*, Virtual Conference. https://peer.asee.org/36878

Brown, N., & Leigh, J. (2020). *Ableism in Academia: Theorising experiences of disabilities and chronic illnesses in higher education*. UCL Press.

*Campbell, S. L. (2020). Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform. *Race Ethnicity and Education,* 1-19.

*Chen, B., West, M., & Zilles, C. (2019). Analyzing the decline of student scores over time in self-scheduled asynchronous exams. *Journal of Engineering Education,* 108(4), 574-594.

 https://doi.org/10.52202/066488-0085

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

D'ignazio, C., & Klein, L. F. (2020). *Data feminism.* MIT Press.

Douglas, K. A., Rynearson, A., Purzer, S., & Strobel, J. (2016). Reliability, validity, and fairness: A content analysis of assessment development publications in major engineering education journals. *The International journal of engineering education*, 32(5), 1960-1971.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology,* 105(3), 399-412.

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison.* Penguin Social Sciences.

Garcia, N. M., López, N., & Vélez, V. N. (2018). QuantCrit: rectifying quantitative methods through critical race theory, *Race Ethnicity and Education*, 21:2, 149-157, DOI: 10.1080/13613324.2017.1377675

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: education, policy,'Big Data' and principles for a critical race theory of statistics. *Race Ethnicity and Education*, 21(2), 158-179.

Godwin, A. (2020). Sitting in the tensions: Challenging whiteness in quantitative research. *Studies in Engineering Education,* 1(1).

Godwin, A., Benedict, B., Rohde, J., Thielmeyer, A., Perkins, H., Major, J., Clements, H, & Chen, Z. (2021). New epistemological perspectives on quantitative methods: An example using Topological Data Analysis. *Studies in Engineering Education*, 2(1).

Hampton, C., Reeping, D., & Ozkan, D. S. (2021). Positionality Statements in Engineering Education Research: A Look at the Hand that Guides the Methodological Tools. *Studies in Engineering Education,* 1(2).

Hendren, S. (2020). *What Can a Body Do?: How We Meet the Built World*. Penguin.

Holly,, J., Jr. (2020), Disentangling engineering education research's anti-Blackness. *Journal of Engineering Education*, 109: 629-635. https://doi.org/10.1002/jee.20364

*Jackson, A., Mentzer, N., & Kramer-Bottiglio, R. (2021). Increasing gender diversity in engineering using soft robotics. *Journal of Engineering Education,* 110(1), 143-160. https://onlinelibrary.wiley.com/doi/epdf/10.1002/jee.20378

Johnson, T. P., & Wislar, J. S. (2012). Response rates and nonresponse errors in surveys. *Jama*, 307(17), 1805-1806.

*Jensen, K. J., & Cross, K. J. (2021). Engineering stress culture: Relationships among mental health, engineering identity, and sense of inclusion. *Journal of Engineering Education*. 2021; 110: 371–392. https://doi.org/10.1002/jee.20391

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2, 255-273. doi:10.1080/10705519509540013

National Science Foundation (2020). About the Division of Engineering Education and Centers. <https://www.nsf.gov/eng/eec/about.jsp>, Accessed 7/15/2021

*Polmear, M., Chau, A. D., & Simmons, D. R. (2020). Ethics as an outcome of out-of-class engagement across diverse groups of engineering students. *Australasian Journal of Engineering Education*, 1-13.

Porter, T. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton, New Jersey: Princeton University Press. http://www.jstor.org/stable/j.ctt7sp8x

Seely, B. E. (1999). The other re-engineering of engineering education, 1900–1965. *Journal of Engineering Education*, 88(3), 285-294.

Shafer, D., Mahmood, M. S., & Stelzer, T. (2021). Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both Asian American and African American students. *Physical Review Physics Education Research,* 17(1), 1-13. https://doi.org/10.1103/PhysRevPhysEducRes.17.010113

Slaton, A. E. (2010). *Race, rigor, and selectivity in US engineering: The history of an occupational color line.* Harvard University Press.

Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it. *Journal of Psychology and Behavioral Science*, 2(2), 205-220.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education,* 2, 53.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, 60, 174-195. doi:10.1002/j.1556-6678.2002.tb00167.x

Walter, M., & Andersen, C. (2016*). Indigenous statistics: A quantitative research methodology.* Routledge.

Wong, A. (Ed.). (2020). *Disability visibility: First-person stories from the twenty-first century.* Vintage.

*Young, J., & Cunningham, J. A. (2021). Repositioning black girls in mathematics disposition research: New perspectives from QuantCrit. *Investigations in Mathematics Learning,* 13(1), 29-42.

Zuberi, T. (2008). Deracializing social statistics: Problems in the quantification of race. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, white methods: Racism and methodology,* (pp. 127-134). Rowman & Littlefield.

 https://doi.org/10.52202/066488-0085