2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering
Education for 130 Years
ASEE
Paper ID #38420

Work in Progress: A Systematic Literature Review of Person-Centered Approaches and Data-Driven Methods in Engineering Education Research

Mr. Jiafu Niu, University of Cincinnati

Jiafu Niu is a M.S. Student in Engineering Education at the University of Cincinnati. He holds a B.S. in Applied Statistics and Actuarial Science from Purdue University.

Dr. David Reeping, University of Cincinnati

Dr. David Reeping is an Assistant Professor in the Department of Engineering and Computing Education at the University of Cincinnati. He earned his Ph.D. in Engineering Education from Virginia Tech and was a National Science Foundation Graduate Research Fellow. He received his B.S. in Engineering Education with a Mathematics minor from Ohio Northern University. His main research interests include transfer student information asymmetries, threshold concepts, curricular complexity, and advancing quantitative and fully integrated mixed methods.

Work in Progress: A Systematic Literature Review of Person-Centered Approaches and Data-Driven Methods in Engineering Education Research

Abstract

In this Work in Progress (WIP) paper, we describe preliminary results from a systematic literature review (SLR) examining how engineering education researchers humanize data-driven quantitative methods by using "person-centered" approaches that emphasize heterogeneity and latent diversity. We systematically review the steps researchers took in their studies from four popular engineering education journals to categorize design elements as person- or variable-centered to elicit themes of how engineering education researchers make analytical decisions when applying data-driven methods. Moreover, we are engaging with the QuantCrit framework to synergize with person-centered approaches and ultimately provide the community with methods to engage with its principles in quantitative research. We anticipate that this SLR could help the field understand how researchers use data-driven quantitative methods and offset the issues within such methods by using person-centered approaches.

Introduction

Engineering education researchers commonly use quantitative methods to seek causal or correlational relationships between observed phenomena of interest. However, Godwin et al. (2021) claim that even though traditional quantitative methods can be used to address crucial questions in engineering education research, they also have the potential to exacerbate issues with equity; for example, removing nuance between subgroups we often call *underrepresented* or *minoritized* when they are aggregated into a monolithic, homogeneous category or excluding them altogether as outliers due to a small sample size. As a result, these methods can incorrectly generalize the findings based on the dominant group to the entire engineering student body.

Pawley (2017) argues that studies in engineering education often make assertions about students broadly when the participants are primarily white males. She explicitly articulates the necessity and urgency for the engineering education research community to reveal this default whiteness and maleness in their research. Further, Gillborn et al. (2018) have argued that statistics have been used to blur, cover, and even legitimize racism and inequity. They elaborate on this issue by showing an example of a British government agency and three leading newspapers advancing a narrative of white students as race-victims by highlighting the statistics of a lower university attendance rate for white students compared to other ethnic groups aggregated into a single non-white group. However, at the same time, a more deliberate analysis shows that white students in Britain form the highest proportion of learners entering elite universities and achieving higher grades when compared to each minoritized group individually.

With these challenges in mind, thoughtfully applied data-driven methods can potentially incorporate and expand on the experiences of minoritized individuals. Data-driven methods adopt a bottom-up framework focusing on the relationships rooted in the data themselves without researchers' presumptions (in theory). This contrasts with traditional statistical methods that adopt a top-down approach and seek causality (Qiu et al., 2018). The inductive nature of data-driven approaches often goes hand-in-hand with the idea that such methods allow "the numbers to speak for themselves" (Anderson, 2008, p. 2). However, this interpretation does not hold when we think more critically about where numbers originate. We choose how we measure, what we measure, and how we analyze the data; even if the analysis is supposedly value-free – meaning there is no model a priori – humans choose what went into the data-driven approach and their normative values with those inputs. Accordingly, Gillborn et al. (2018) warn that the pre-assumed objectivity associated with data-driven research could bring the risk of reinforcing racist stereotypes and systems of power because "numbers are social constructs and likely to embody the dominant (racist) assumptions that shape contemporary society" (p. 173). These issues can manifest in various contexts, which Cathy O'Neil showcases in her book, *Weapons of Math Destruction* (2017), by detailing how data-driven methods can be used for such nefarious purposes – both intentionally and unintentionally – in various contexts, such as the justice system, job applications, and online advertising.

The aforementioned critiques can be more broadly cast into the analytical toolkit of QuantCrit, which adopts the lens of Critical Race Theory to interrogate conventional – but potentially harmful – practices in quantitative research. Gillborn et al. (2018) offer a full discussion of the five fundamental principles that embody QuantCrit: (1) the centrality of racism (this principle asserts that racism is a ubiquitous component of society, and some scholars do not believe it is quantifiable); (2) numbers are not neutral (e.g., using statistics to show deficits in minoritized groups); (3) Categories/groups are not natural nor given (i.e., race and gender as social constructs); (4) Data cannot speak for themselves (all data require interpretation); (5) Social justice and orientation (QuantCrit denies assumed objectivity and political neutrality when applying quantitative research).

Person-Centered and Variable Centered Approaches

To help distinguish between the underlying mechanisms of various quantitative approaches, Godwin (2021) introduced the concept of person-centered approaches to the engineering education community, which originated in the context of longitudinal analyses. A person-centered approach recognizes heterogeneity and attempts to identify latent groupings among individuals in the sample based on the relationships among variables which reflect the characteristics of individuals and their environment. In contrast, a variable-centered approach is focused on prediction and relationships between variables (Laursen & Hoff, 2006). Although person-centered approaches may use data-driven methods to fulfill these tasks, not all data-driven methods can be used in a person-centered fashion without more critical thought (Godwin et al., 2021). For example, Principal Component Analysis is a data-driven method used to reduce the dimensionality of a dataset, but it is not necessarily applied in a person-centered fashion because it consolidates *variables* into composite quantities (i.e., the principal components). Often it is not clear how to interpret the meaning of the resulting principal components, which is compounded by the loss of information during their formation.

Although person-centered approaches can offer more robust and detailed insights into groupings within the dataset, it does not imply that person-centered approaches should replace variable-centered methods completely (Godwin et al., 2021). Put another way, variable and person-centered approaches should not be considered as a "right" and "wrong" dichotomy; instead, researchers are encouraged to choose the quantitative methods based on the research question and the merits of different quantitative methods (Laursen & Hoff, 2006).

Given the new lens for framing quantitative methods in engineering education, there is potential to rethink how the community employs data-driven approaches. Moreover, with the fresh perspective of person-centered approaches, little is known about how engineering education researchers already employ these techniques– or if these considerations are made intentionally. By reviewing the extant literature for examples of thoughtful applications of data-driven methods using person-centered approaches, such manuscripts can serve as exemplars for future efforts across research areas in engineering education.

Research Aims

This work-in-progress addresses how engineering education researchers adopt data-driven methods and humanize the application of these methods using person-centered approaches. Our central research question is: "How do engineering education researchers use data-driven quantitative methods, and how do they humanize the methods by employing person-centered approaches." In particular, we seek to understand how engineering education researchers can leverage data-driven approaches while engaging with the principles of QuantCrit.

Method

To dig deeper into the use of data-driven methods and person-centered approaches in engineering education research, we conducted a Systematic Literature Review (SLR). SLRs have been used to form the evidence base to inform research, practice, and policy through systematically synthesizing, appraising, critiquing, and summarizing the existing literature on a topic (Borrego et al., 2014). We followed the conventional process for conducting a systematic literature review (i.e., Preferred Reporting Items for Systematic Reviews and Meta-Analyses, or PRISMA) to analyze publications from four journals within the last decade (2011-2021). Moreover, we cataloged the steps taken within each manuscript, including data collection, analysis, results, and conclusions drawn to understand how person-centered analyses were situated with data-driven approaches.

Data Collection and Preparation

The first step of collecting data in an SLR involves determining the inclusion criteria, forming the appropriate search strings, and choosing databases. Because of the exploratory nature of this SLR, we did not establish strict inclusion criteria. Our inclusion criteria are listed below:

- Must use a quantitative or mixed methods research design
- Must use at least one data-driven method in the research design
- Must be an empirical article
- Must be human subjects research

We combined terms for common data-driven methods, such as "cluster analysis" and "random forest,” with "OR" as the conjunction to form our search strings. More generic terms, such as "data mining," were also added. The entire search string can be found in Appendix 1.

We focused on four popular engineering education journals: the *Journal of Engineering Education (JEE), the European Journal of Engineering Education (EJEE), the International Journal of Engineering Education (IJEE), and IEEE Transactions on Education (TOE)*. We narrowed the publication period to the last decade (2011 – 2021) and queried Education Research Complete and the *Journal of Engineering Education*'s search engine to retrieve manuscripts. The PRISMA flowchart for this study can be found in Appendix 1.

The total number of articles retrieved from the databases was 236. After the search stage, we screened the abstracts, excluding 138 after applying the inclusion criteria. Three reasons led to a manuscript's removal, including (1) not employing a data-driven method, (2) employing a data-driven method, but the technique was not applied with human subjects, and (3) not being an empirical study. Thus, 98 articles remained and were grouped by their main quantitative method: Exploratory Factor Analysis (EFA), Logistic Regression, Cluster Analysis, Principal Component Analysis (PCA), Decision Tree, Random Forest, and Others. The "Others" group was temporary because, based on the abstracts of articles in this group, either the data-driven methods taken by researchers did not fit into the other six categories (e.g., using multiple approaches simultaneously) or authors used generic terms such as "data mining" or "machine learning" to describe the quantitative methods. These manuscripts needed further investigation to identify the specific data-driven methods they used.

Next, based on the groupings, we implemented a full-text review of the remaining articles to filter out manuscripts that did not meet our inclusion criteria when scrutinized further. One of our major decisions was tentatively excluding the EFA and Logistic Regression groups. We found that the logistic regression manuscripts tended to perform traditional statistical modeling rather than using an inductive data-driven framework. Regarding the EFA group, even though the technique itself is data-driven (Godwin et al., 2021), the focus of those articles involved constructing an instrument - not necessarily applying the technique to human subjects - which did not align with this SLR's scope. We removed one article from the PCA group, Chan and Fong (2018), because the results of PCA from other researchers' studies were discussed rather than applying the PCA method in their study. Pizard & Vallespir (2017) was also removed from the Cluster Analysis group because it was found not to involve human subjects upon further inspection. Thus, twenty-six articles were retained and moved to the synthesis stage.

Table 1. Summary of the composition of the final sample of selected articles

Data-Driven Method	Number of Articles	% Sample
Cluster Analysis	13	50%
PCA	2	7.6%
Decision Tree	3	11.6%
Random Forest	1	3.9%
Naïve Bayes	1	3.9%
Hidden-Markov Chain	2	7.6%
Topic Modeling	2	7.6%
Association Analysis	1	3.9%
Model Comparison	1	3.9%
Total	26	

Analyzing the Data

We conducted three rounds of coding to gain insight into how engineering education researchers humanize data-driven analyses using person-centered approaches. In the first round, each manuscript's methods and results/discussion sections were coded using descriptive and In Vivo codes to categorize and index the content (Saldaña, 2013). In the second round, within each group, all the content associated with the same codes was extracted for further analysis and summarization. Finally, the findings from the previous two coding cycles were aggregated at the method level to highlight the person-centeredness of each technique. These themes were aligned back with the QuantCrit framework to begin finding ways for engineering education researchers to engage with its principles.

Selected Results

Our analyses are still in progress, but we will highlight examples of person-centeredness and variable-centeredness found in the study at this point. In this case, we focus on cluster analysis, which was used in half of the papers reviewed in this SLR. Lund & Ma (2021) describes cluster analysis as a popular data mining process to identify underlying patterns and groupings in a sample among the measured variables. Further, Godwin et al. (2021) suggest that cluster analysis is a data-driven method that can be person-centered and is a prime technique for uncovering latent diversity (see Godwin, 2017) in a sample by finding commonalities within individuals' non-cognitive attributes. Our first example is from Faber & Benson (2017), who sought to understand how engineering students solve open-ended assignment problems and the relationship between their epistemic motivation, engineering epistemic beliefs, and epistemic cognition. They used an explanatory sequential mixed methods design (Plano Clark & Creswell, 2018) with cluster analysis to identify the subgroups of students who shared similar engineering epistemic beliefs and epistemic motivations. Semi-structured interviews followed the clustering and statistical inter-cluster analysis, and the qualitative and quantitative data were overlaid to investigate the potential connections among measured constructs in the context of problem-solving.

Faber & Benson (2017) used cluster analysis as a person-centered technique focused on identifying latent groups by examining the patterns of individual responses within the dataset (Godwin et al., 2021). Adding a qualitative layer of context to these groups empowered them further to explore the nuanced differences within the clusters/subgroups to gain a deeper understanding of the individual responses. This implementation of cluster analysis manifests the descriptive element of person-centeredness and synergizes with the QuantCrit principle of "categories/groups are not natural nor given" by recognizing students' latent diversity within epistemic motivations and beliefs and identifying subgroups beyond demographics. Moreover, when investigating the intra-cluster similarity of one cluster during their qualitative analyses, they found one student with no epistemic aim. This finding contrasted with all other interviewed members in the same cluster. From a classic statistical point of view, the dissimilarity with other group members would lead a researcher to conclude that the student was an outlier. However, instead of ignoring the student's experiences, her responses were carefully examined to explain why she differed from other group members. This approach to analysis embodied the "numbers are not neutral" QuantCrit principle, moving beyond the static description of the "average" for the cluster and exploring contrarian variation within groups that share common traits.

Second, we share an example of how engineering education researchers could employ PCA - a data-driven technique that is not often applied in a person-centered manner (Godwin et al., 2021). PCA is one of the most frequently used dimension reduction methods, which aims to identify a subset of variables to represent a dataset in a lower dimension without losing significant information (Kherif & Latypova, 2020). In this example, Martin & Sorhaindo (2019) compared intrinsic and extrinsic motivational factors as predictors of academic achievement for civil engineering students. They employed PCA to consolidate motivational factors into intrinsic and extrinsic groups. The original twenty-two motivation variables were ultimately grouped into five principal components, which accounted for 66% percent of the variance. This implementation of PCA to aggregate variables showcases elements of variable-centeredness, which can be interpreted as working against the QuantCrint principle of “numbers are not neutral” and “categories/groups are not natural nor given.” To elaborate, after the retained factors were grouped, researchers examined if any mean difference existed for each principal component among participant groups categorized by their demographic information, such as gender (male vs. female), origin (native vs international) and age. This implementation of comparing groups categorized by dichotomized social constructs can work against the QuantCrit principle “categories are neither natural nor given,” especially when monolithic categories are used. Trends can reverse or disappear at different levels of aggregation (e.g., Shafer et al., 2021), so care must be taken when defining reference and comparison groups. Moreover, despite PCA forming aggregate variables to perform data analysis in higher dimensional datasets, variation is washed out across latent constructs to favor simplying the analytical process. In other words, PCA focuses on creating composite variables that can make it difficult to understand relationships among individuals. The composite variable may or may not be theoretically meangingful, and when used in decision-making can be lead one down an incorrect path – hence the application working against the “numbers are not neutral” principle.

Conclusion and Future Work

Based on our preliminary review and coding, we found that engineering education researchers have used a wide range of data-driven methods, as shown in Table 1 – cluster analysis is the most popular. We provided two examples from our sample to contextualize the applications of data-driven methods and their respective person and variable-centeredness. The next steps for this study involve aggregating our findings to the method level with respect to the principles of QuantCrit. Moreover, we will extract the sequence of methods as a network and use ego-network analysis to find common pairings of methods and how person-centered approaches are situated in the research designs (Reeping, 2022).

Acknowledgments

This research was supported by a grant from the American Educational Research Association Division D.

References

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired, 16*(7), 1-2

Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic literature reviews in engineering education and other developing interdisciplinary fields: Systematic literature reviews in engineering education. *Journal of Engineering Education*, *103*(1), 45–76. https://doi.org/10.1002/jee.20038

Chan, C. K. Y., & Fong, E. T. Y. (2018). Disciplinary differences and implications for the development of generic skills: A study of engineering and business students' perceptions of generic skills. *European Journal of Engineering Education*, *43*(6), 927–949. https://doi.org/10.1080/03043797.2018.1462766

Faber, C., & Benson, L. C. (2017). Engineering students' epistemic cognition in the context of problem-solving. *Journal of Engineering Education*, *106*(4), 677–709. https://doi.org/10.1002/jee.20183

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. https://doi.org/10.1080/13613324.2017.1377417

Godwin, A. (2017). *Unpacking Latent Diversity.* 2017 ASEE Annual Conference & Exposition Proceedings, Columbus, OH. https://peer.asee.org/29062

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), 16. https://doi.org/10.21061/see.18

Kherif, F., & Latypova, A. (2020). Principal component analysis. In A. Mechelli & S. Vieira (Eds.), *Machine Learning* (pp. 209–225). Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00012-2

Laursen, B. P., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*(3), 377–389. https://doi.org/10.1353/mpq.2006.0029

Lund, B., & Ma, J. (2021). A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering. *Performance Measurement and Metrics*, *22*(3), 161-173. https://doi.org/10.1108/PMM-05-2021-0026

Martin, H., & Sorhaindo, C. (2019). A comparison of intrinsic and extrinsic motivational factors as predictors of civil engineering students' academic success. *International Journal of Engineering Education*, *35*(2), 458–472.

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pizard, S., & Vallespir, D. (2017). Towards a controlled vocabulary on software engineering education. *European Journal of Engineering Education*, *42*(6), 927–943. https://doi.org/10.1080/03043797.2016.1235139

Qiu, L., Chan, S. H. M., & Chan, D. (2018). Big data in social and psychological science: Theoretical and methodological issues. *Journal of Computational Social Science*, *1*(1), 59–66. https://doi.org/10.1007/s42001-017-0013-6

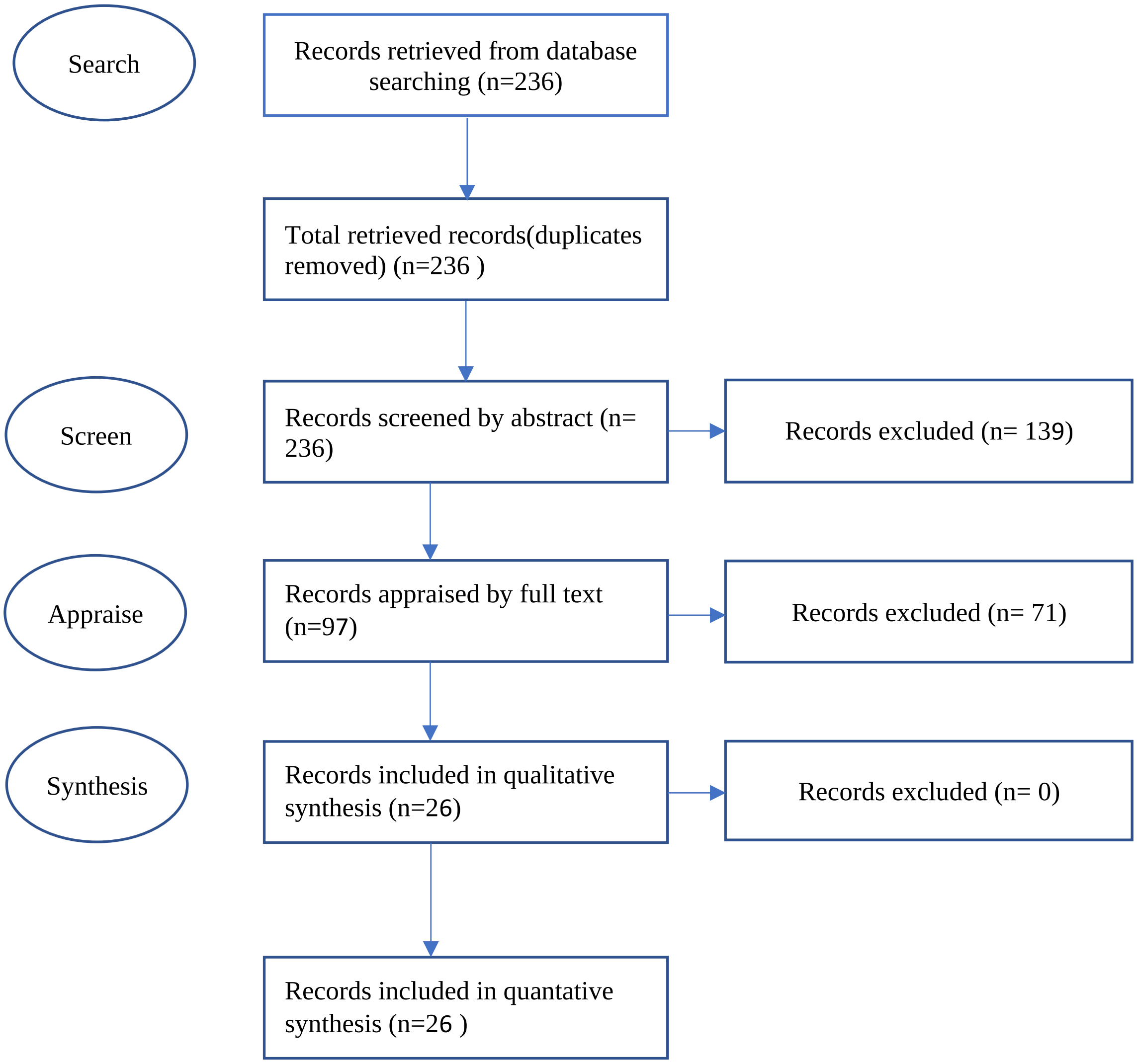
Reeping, D. (2022). *Work in progress: Using ego network analysis to analyze how engineering education researchers construct mixed methods designs.* Proceedings of the 2022 ASEE Annual Conference, Minneapolis, MN. https://peer.asee.org/40829

Saldaña, J. (2013). *The coding manual for qualitative researchers* (3rd Ed). Sage.

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

Appendix 1: Flowchart of Articles

Full Search String: "cluster analysis" OR "association analysis" OR "item-set" OR "item set" OR "rule based" OR "rule-based" OR "classification" OR "random forest" OR "machine learning" OR "data mining" OR "principal component" OR "decision tree" OR "KNN" OR "nearest-neighbors" OR "nearest neighbors" OR "support vector " OR "exploratory factor analysis" OR "EFA" OR "logistic regression" OR "Bayes"


Search
Records retrieved from database searching (n=236)
Total retrieved records(duplicates removed) (n=236)
Screen
Records screened by abstract (n= 236)
Records excluded (n= 139)
Appraise
Records appraised by full text (n=97)
Records excluded (n= 71)
Synthesis
Records included in qualitative synthesis (n=26)
Records excluded (n= 0)
Records included in quantative synthesis (n=26)