BOARD # 358: ECR: Facilitating change in undergraduate STEM: A multidisciplinary, multimethod metasynthesis mapping a decade of growth

Dr. Ying Wang, FHI 360

Ying Wang is an education researcher in the U.S. Education at FHI 360, a nonprofit human development organization. Formerly, she was a Postdoctoral Fellow in the College of Engineering at the Georgia Institute of Technology since Fall 2021. She earned her Ph.D. in Educational Psychology at The Pennsylvania State University and her Master of Education at Temple University. Ying is an educational psychologist who applies educational psychology science in STEM educational contexts.

Emily Bolger
Rachel L Renbarger
Taylor Boyd, Western Michigan University
Prof. Noah D Finkelstein, University of Colorado Boulder

Noah Finkelstein is a Professor and Vice Chair in the department of Physics at the University of Colorado Boulder. He conducts research is in physics education, specifically studying the conditions that support students' identities, engagement and outcomes in physics - developing models of context. In parallel, he conducts research on how educational transformations get taken up, spread, and sustained. He is a PI in the Physics Education Research (PER) group and was founding co-director of CU's Center for STEM Learning. He co-directs the national Network of STEM Education Centers, is building the STEM DBER-Alliance, and coalitions advancing undergraduate education transformation. He is involved in education policy serving on many national boards, sits on a National Academies' STEM education roundtable, is a Trustee of the Higher Learning Commission, is a Fellow of both the American Physical Society and the American Association for the Advancement of Science, and a Presidential Teaching Scholar and the inaugural Timmerhaus Teaching Ambassador for the University of Colorado system.

Dr. Charles Henderson, Western Michigan University

Charles Henderson is a Professor at Western Michigan University (WMU), with a joint appointment between the Physics Department and the WMU Mallinson Institute for Science Education. He is the Director of the Mallinson Institute and co-Founder and co-Direc

Andrea L Beach, Western Michigan University
Scott P. Simkins, North Carolina A&T State University

Scott P. Simkins is an Associate Professor in the Department of Economics at North Carolina Agricultural and Technical State University. He conducts research in economic education, emphasizing the use and diffusion of evidence-based teaching practices within economics and across disciplines.

Marcos Caballero

ECR: Facilitating Change in Undergraduate STEM: A Multidisciplinary, Multimethod Metasynthesis Mapping a Decade of Growth

Abstract

The aim of this NSF ECR project is to perform an extensive multi-method metasynthesis of literature published between 2011 and 2023 on strategies for enhancing undergraduate STEM instruction. Specifically, we update the previous review and examine the change strategies implemented after a decade of research. We present an updated methodology with the innovative application of machine learning methods to select and analyze articles. From initially determined potentially relevant articles (n = 9,262) from keyword search, 253 articles were included after the title and abstract and full-text screening. Subsequently, we conducted both human qualitative coding and quantitative machine learning analyses to examine the themes of the included articles. Preliminary findings from the qualitatively coding showed that most articles implemented a dissemination change strategy focusing on telling or teaching individuals about new teaching practices; the predominant target for disseminating pedagogy was individual faculty and developing reflective teachers-focused strategies, whereas departments and institutions tended to be the target for developing a policy or a shared vision. Additionally, preliminary findings from the quantitative machine-learning clustering analyses showed groupings related to specific science disciplines (e.g. engineering, chemistry). Next steps of the project are discussed.

Introduction

The previous review requires an update and expansion to address both the progress made in the last decade and the current context of U.S. higher education and how new technologies can support systematic reviews. For example, the NASEM (2024)[1] suggests that more research is needed to understand how early STEM education innovations can be scaled and sustained so that students can succeed academically in STEM learning through early to postsecondary education. As such, given numerous educational reforms have occurred in the past decade, it is important to understand what change strategies have been employed in these reforms and analyze how these changes took place. The aim of the overall NSF ECR project is to perform an extensive multi-method metasynthesis of literature published between 2011 and 2023 on strategies for enhancing undergraduate STEM instruction. Specifically, the project's overarching research questions include:

1. How have the change strategies published between 2010 and 2021 compared to those previously? Do the four categories stand up? Do the more recent change efforts work across or draw from multiple categories more than previous efforts?
2. What broader categories are now seen in literature? Are there new approaches or categories of change? How have specific social, cultural matters shown up to change efforts in US STEM education, specifically, to what degree do change efforts attend to the COVID-19 pandemic and/or issues of diversity equity inclusion?
3. How might machine learning (ML) approaches support and apply to systematic reviews of this kind?

This project update will share progress toward these goals and research questions. We share updated methodological approaches, which include new screening criteria, an updated framework and rubrics for identifying both the approach to and quality of systemic change work, and the design and preliminary outcomes of machine learning approaches.

Methods

In metasyntheses, researchers systematically search, review, and integrate the research on a specific topic [2]. Here, we focus on undergraduate STEM instructional change empirical literature. However, rather than simply describe or summarize a body of literature, metasyntheses integrate findings from qualitative, quantitative, mixed method, and theoretical work to create new understandings of the field [3], [4]. For this metasythesis of the change literature, we followed the PRISMA 2020 systematic review guide [5].

Searching

We identified key terms through integrating the search terms used in the original review [6], manual searches within STEM education journals, and feedback from a panel of systemic change, higher education, and systematic review researchers. As a result, we developed the following search queries: 1) change strategy query (e.g., "change" or "reform"); 2) instruction query (e.g., "teaching" or "instruction"); 3) undergraduate education query (e.g., "college" or "higher education"); and 4) STEM query (e.g., "STEM" or "physics"). We also used non-higher education (e.g., "high school" or "secondary") and non-U.S. country terms (e.g., "Turkey" or "Spain") to filter out irrelevant articles. We searched for relevant articles using these search queries in library databases, including Academic Search Complete, ERIC, Education Research Complete, EconLit, and Sciecne & Technology Collection. In this paper, we only report the searching results from library databases although we have other sources of data (e.g., Google Scholar), which will be reported after the project is done. As a result of the library databases, we identified 9,262 potentially relevant articles.

Screening

We established inclusion criteria to include peer-reviewed articles that describe a systemic change in undergraduate STEM education with the ultimate goal of changing the instructional practices of instructors and improving students' STEM learning. Articles that met these criteria also needed to be written and published in English and the studies must have been conducted in the United States. As we target articles focusing on systemic change and scaling, we did not include articles that described a classroom intervention or curriculum development to change students' non-academic outcomes (e.g., students' wellbeing) nor intervention articles that did not discuss how to scale up the intervention components to other and wider contexts. The research team conducted multiple training sessions until the team reached a substantial agreement on judging articles for inclusion (Cohen's Kappa = .72). After the title and abstract and full-text screening stages, we identified 253 articles for coding.

Human Qualitative Coding

In the human qualitative coding process, four coders first coded a set of articles collectively and held multiple training sessions to reach a consensus. Coding categories include discipline of the journal, first author's affiliation, funding mechanism, institution type, primary stated purpose, change strategy, target of change, and agent of change. While these coding categories are consistent with the original review[6], we also added categories to mark articles focusing on online instruction, COVID, technology, DEI, and adjunct faculty to identify new trends in the analyses.

Quality Measures

We also independently developed and applied a coding rubric that captured the quality of scholarship around systemic change in undergraduate STEM education as the "coding of study quality is essential"[7]. The core components of the rubric include relevance (i.e., to what degree the topic is related to systemic change), richness (i.e., connections to systemic change literature), and rigor (i.e., whether the data, analysis, and conclusions are connected to each other with systemic change).

Machine Learning Approach

In addition to human coding, we have employed Natural Language Processing (NLP) techniques in the analysis pipeline to assess the utility and efficacy of data science and machine learning methods in systematic literature reviews, particularly in information extraction [8], [9]. We used both traditional approaches such as Latent Dirichlet Allocation (LDA) [10] and approaches that leverage large language models (LLMs) and unsupervised clustering techniques [11], [12], [13]. It is important to note that the purpose of clustering is to identify patterns common across the entire corpus of data. Our goal of the machine learning approach is to validate the human qualitative coding and identify new themes that might not be discovered through qualitative coding.

Results

Qualitative Coding

Preliminary findings from the qualitatively coded 100 articles showed that many articles (n = 45) implemented a dissemination change strategy focusing on telling or teaching individuals about new teaching practices versus the other strategies focusing on developing reflective teachers, policy, or shared vision; the predominant target for disseminating pedagogy was individual faculty- and developing reflective teachers-focused strategies, whereas departments and institutions tended to be the target for developing a policy or a shared vision. We also found that articles published in a science education research journal tended to focus on the dissemination strategy, whereas policy development articles tended to appear in higher education and other journals.

Quality Measures

We evaluated quality for a set of 50 randomly selected articles. Overall, we found that the bulk (60%) of the articles were highly relevant. However, less than half the articles were considered highly rich (connected to the literature) and similarly fewer were considered of high rigor. Additionally, we found that articles with high relevance tended to have high richness. We also observed that half of highly relevant articles were highly rigorous and only 10% low rigor. Low relevance articles papers were more likely to be of low rigor. Among the findings that these results suggest is that there are different paper types focusing on systemic and institutional change. Some papers are deeply focused on the scholarship and models of change, and these are more likely to be richly connected to the literature on change and provide compelling arguments (empirically or theoretically) about their claims. Another class of articles seems to be about issues or approaches that abut, but do not deeply focus on systemic change – articles such as describing a course reform or pieces that argue about the role of technologies in learning.

Machine Learning Approach

In our initial analysis, we found clusterings that were focused on different STEM disciplines (e.g. engineering, math, physics), even across a variety of parameter combinations tested. Based on the theory behind text embeddings, the initial results were not necessarily surprising; however, they were unhelpful in identifying instructional change strategies in the corpus of articles. Thus, we attempted to shift the focus of the model to the text that is relevant to an analysis focused on uncovering themes of institutional change in a variety of STEM domains. Repeating the NLP analysis pipeline, the resulting clusters focused on the change strategies that we sought, rather than the disciplines in which these strategies were implemented. Most notably, we noticed how the clustering on the dataset with the removed disciplinary words achieves topics from multiple disciplines with a focus on a specific change strategy. For example, we see articles situated in disciplines such as math, engineering, and biology be placed in a single cluster focused on equity and inclusion change strategies. This data shift speaks to the challenges of integrating NLP tasks in traditional human analysis techniques, especially in the information extraction stage of literature reviews.

Conclusion & Discussion

From these results, the systematic change literature continues to reflect a diversity of foci that take place across STEM fields. Compared to the previous synthesis[6], more research includes a theoretical framing regarding systemic change; indeed, many of the articles referenced the previous synthesis as a major framework for their change work. These findings also suggest more nuance regarding the success of their interventions. The results suggest that change research could improve; issues of quality abound in the literature and both authors and editors must work to ensure change research contains methodological rigor – specifically, data collection and analyses should focus on systemic change methods rather than anecdotal or descriptive data. The field should move beyond small-scale change as the only method of choice and engage in methodological diversity to facilitate more large-scale change. The ML results can help with categorizing, such as through identifying disciplinary clusters or broad strokes strategies, although without a deeper read of the articles, ML results cannot be used as the only source of themes.

In Spring 2025, the human coding team will increase the sample size of articles and confirm the findings shared here, examine correlations between these quality codes and the change strategies used, and compare between these quality codes and the ML findings below. The next stages of the ML work include creating labels for these clusters using both human expert knowledge and NLP techniques, identifying other potential data cleaning ideas to strengthen the clustering results, and performing a similar analysis with the full text all of the included articles.

Acknowledgment and NSF Disclaimer
This material is based upon work supported by the National Science Foundation under Grant Numbers (2201792, 2201793, 2201794, and 2201795). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

References

[1] C. M. Massey and A. Stephens, Eds., *Scaling and sustaining Pre-K-12 STEM education innovations: Systemic challenges, systemic responses*. Washington, D.C.: National Academies Press, 2024. doi: 10.17226/27950.
[2] J. Lachal, A. Revah-Levy, M. Orri, and M. R. Moro, "Metasynthesis: An original method to synthesize qualitative literature in psychiatry," *Front. Psychiatry*, vol. 8, Dec. 2017, doi: 10.3389/fpsyt.2017.00269.
[3] E. J. Erwin, M. J. Brotherson, and J. A. Summers, "Understanding qualitative metasynthesis: Issues and opportunities in early childhood intervention research," *J. Early Interv.*, vol. 33, no. 3, pp. 186–200, Sep. 2011, doi: 10.1177/1053815111425493.
[4] K. Thunder and R. Q. Berry, "Research commentary: The promise of qualitative metasynthesis for mathematics education," *Journal for Research in Mathematics Education*, vol. 47, no. 4Jul. 2016, doi: 10.5951/jresematheduc.47.4.0318.
[5] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Int. J. Surg.*, vol. 88, p. 105906, Apr. 2021, doi: 10.1016/j.ijsu.2021.105906.
[6] C. Henderson, A. Beach, and N. Finkelstein, "Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature," *J. Res. Sci. Teach.*, vol. 48, no. 8, pp. 952–984, 2011, doi: 10.1002/tea.20439.
[7] T. D. Pigott and J. R. Polanin, "Methodological guidance paper: High-quality meta-analysis in a systematic review," *Rev. Educ. Res.*, vol. 90, no. 1, pp. 24–46, Feb. 2020, doi: 10.3102/0034654319877153.
[8] E. Popoff, M. Besada, J. P. Jansen, S. Cope, and S. Kanters, "Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews," *Syst. Rev.*, vol. 9, no. 1, p. 293, Dec. 2020, doi: 10.1186/s13643-020-01520-5.
[9] R. van Dinter, B. Tekinerdogan, and C. Catal, "Automation of systematic literature reviews: A systematic literature review," *Inf. Softw. Technol.*, vol. 136, p. 106589, Aug. 2021, doi: 10.1016/j.infsof.2021.106589.

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.

[11] L. K. Nelson, “Computational grounded theory: A methodological framework,” *Sociological Methods & Research,* vol. 49, no. 1, pp. 3–42, Feb. 2020.

[12] R. J. G. B. Campello, D. Moulavi, and J. Sander, “Density-based clustering based on hierarchical density estimates,” in *Advances in Knowledge Discovery and Data Mining*, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds., Berlin, Heidelberg: Springer, 2013, pp. 160–172.

[13] P. Tschisgale, P. Wulff, and M. Kubsch, “Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory," *Phys. Rev. Phys. Educ. Res*, vol. 19, no.2, Sep. 2023. [Online]. Available: https://journals.aps.org/prper/abstract/10.1103/PhysRevPhysEducRes.19.020123 [Accessed Jan. 15, 2025].