

Work in Progress: Reviewing engineering education scholarship in a mentored reviewer program: Trends based on geographic region

Lisa Benson

*Dept. of Engineering and Science
Education
Clemson University
Clemson, SC USA
lbenson@clemson.edu*

Kelsey Watts

*Dept. of Bioengineering
Clemson University
Clemson, SC USA
kwatts@clemson.edu*

Gary Lichtenstein

*Quality Evaluation Designs
Bluff, UT USA
gary@qualityevaluationdesigns.com*

Karin Jensen

*Dept. of Biomedical Engineering
University of Michigan
Ann Arbor, MI USA
kjens@umich.edu*

Evan Ko

*Dept. of Bioengineering
University of Illinois Urbana-
Champaign
Urbana, IL USA
mko5@illinois.edu*

Rebecca Bates

*Dept. of Integrated Engineering
Minnesota State University,
Mankato
Mankato, MN USA
bates@mnsu.edu*

Abstract— As a field, engineering education research (EER) draws from multiple disciplines and geographic regions. This presents challenges in the process of peer reviewing EER manuscripts, as peer reviewers typically do not receive formal training in critiquing research publications. We investigated how EER scholars develop an understanding of how to conduct peer reviews of manuscripts related to engineering education. We draw from theories of cognition and schema development to address the following question: What aspects of EER manuscripts are reviewers paying attention to, and how do these aspects differ for reviewers from different geographic regions?

The context for this study was a mentored peer reviewer program, which pairs less experienced reviewers (mentees) with experienced reviewers (mentors) to review a set of manuscripts. Participants ($n=27$: nine mentors and 18 mentees) from six countries and varied levels of expertise were organized into triads (one mentor and two mentees). Each triad collaboratively evaluated each manuscript, crafted feedback to authors and formulated recommendations to journal editors. Data collected from participants included Structured Peer Reviews (SPR) that prompted participants to comment on the main strengths and weaknesses of a short EER manuscript and formulate a recommendation to the editor. All participants completed SPRs for one manuscript prior to the program (Pre-SPR) and a second manuscript at the end of the program (Post-SPR). SPRs were coded and categorized into six themes: context, methods, results, discussion, mechanics, and EER relevance. Themes were categorized into codes related to positive or negative assessments.

The preliminary analysis presented in this paper focuses on the extent to which reviewers converged in the breadth of what they commented on in the pre- and post-SPR forms. Results were compared based on where participants earned their doctorate. Differences between participants from different geographic regions emerged in the Pre-SPR data but were minimal in the post-SPR. This research provides a foundation for considering peer review as a form of professional development as findings suggest that training affects reviewers' understanding of EER scholarship.

Keywords—engineering education research, peer review, academic publishing, mental models

I. INTRODUCTION

Advancing knowledge in any discipline relies on peer review of academic publications, wherein every aspect of a manuscript is critiqued by scholars with relevant expertise. In addition, the outcomes of peer review affect nearly every aspect of an academic's career, including funding, awards, promotion, tenure, and performance evaluations [1]. Yet scholars receive little or no training in effective and constructive peer review, leading to uneven quality, inappropriate personal remarks, bias, and lack of constructive feedback [2]. Harsh reviews result in exclusion of individuals from the academic community, particularly for scholars from marginalized groups [3]. Thus, there is a need to provide training to peer reviewers. To provide effective professional development in peer review, it is important to understand how scholars develop peer review skills, including their understanding of how to critique scholarship. This is especially important in the field of engineering education, where researchers hail from a variety of disciplines and cultures [4], which may influence strengths and weaknesses they focus on when reading journal manuscripts.

This paper presents preliminary findings that are part of a larger study of how participants in a peer reviewer training program [5] build mental models, or schema [6] as they develop reviewing skills, and what aspects of those schema are shared between reviewers from different disciplinary and cultural backgrounds, in particular as identified by geographic region of doctoral training.

II. LITERATURE REVIEW

A. Cognition and Schema Development

Cognition is the process of thinking and building knowledge; schema, or models of concepts stored in memory, are the building blocks of cognition [6]. Schema can also be defined as the process of building those mental models [7]. Specific to cognition during reading, schema are considered the

“fundamental elements upon which all information processing depends” (p. 33). Information processing starts with sensing incoming information through selective attention [8]; what individuals notice and pay attention to is the start of cognition and or building knowledge and understanding. Thus to understand how individuals are developing mental models about a process such as peer reviewing, an important first step is identifying what they are paying attention to as they go through that process.

B. Analyzing Peer Reviews

In a study of reviews submitted to American Psychological Association journals [9], researchers characterized the weaknesses reviewers articulated in their recommendations, specifically in terms of the planning and execution of the study (conceptualization, design, method, analysis, interpretations, and conclusions) and presentation (quality of expression). Two-thirds of the weaknesses identified were related to planning and execution of the study, with the remaining one-third focusing on presentation. There was minimal consensus across reviewers as to whether to Accept, Revise or Reject, although there were few disagreements across reviewers about specific issues in the paper, or what aspects of the papers they paid attention to.

While it is likely that disciplinary background would affect what individual reviewers might focus on when critiquing a manuscript, it is less clear what effect, if any, geographic and cultural differences might have on that process. For example, different countries have different academic training models and degree requirements for doctoral studies [10] that prepare graduates, particularly those in the social sciences, differently for conducting independent research [11] and potentially for critiquing literature. Thus, this study addresses the question: What aspects of EER manuscripts are reviewers paying attention to, and how do these aspects differ for reviewers from different geographic regions?

III. METHODS

A. Overview of Study

Data were collected from participants ($n=27$; 9 mentors and 18 mentees) in a peer reviewer training program [5] in which participants were formed into nine triads that each completed three peer reviews of manuscripts submitted to an engineering education journal. All recruitment and data collection procedures followed approved IRB protocols.

B. Participant Selection and Background

The 18 mentees were selected through a competitive, online application process that collected demographic information, professional background (their Ph.D. concentration, institution at which they earned their doctorate, and year earned), current position, relevant EER experience, confidence reviewing EER manuscripts, and the number of EER colleagues with whom they regularly interact. Participants were chosen based on a baseline level of experience and their desire to help advance EER through peer review. Special consideration was given to individuals deemed “lone wolves” who were not well connected to an EER network [12] and participants from

diverse demographic groups. The nine mentors were invited to participate based on their experience in EER and recommendations from journal editors. Mentors provided similar background information as mentees, with additional questions related to mentoring experience. In total, mentors and mentees represented 23 universities in six countries. Participants’ levels of experience included graduate students, postdoctoral researchers, and faculty. Participants’ disciplinary backgrounds were in social sciences, engineering, and engineering education.

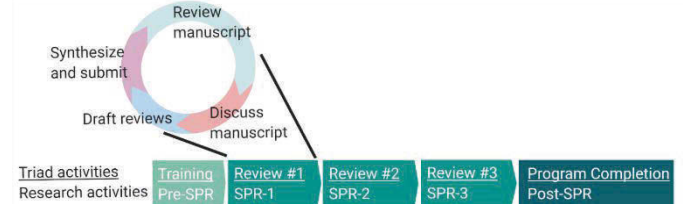


Fig. 1. Schematic of program activities of mentors and mentees in the mentored review program [5].

C. Data Collection and Analysis

Triads (one mentor and two mentees) were formed based on participants’ time zones and areas of expertise. After an orientation session and training in conducting peer reviews and the journal’s review criteria, each triad completed reviews of three articles submitted to an EER journal (Figure 1). Participants were asked to complete a Structured Peer Review (SPR) form prior to the triad’s first meeting (Pre-SPR), for each manuscript they reviewed as a triad (SPRs 1, 2, and 3), and after their final triad review was submitted (Post-SPR). The SPR is an online questionnaire that asked participants to describe the five most notable strengths and weaknesses of a manuscript, recommend a decision to the editor (Accept, Minor Revision, Major Revision, or Reject), and make a 200-word justification of their recommendation (Figure 2). Participants were instructed to fill out their SPRs individually and use them as a starting point in their discussions with their triads. Of the 27 participants, 26 consented to participate in the research, and their responses on the Pre-SPR and Post-SPR are the source of data for this paper. In addition, one mentor dropped out after collection of the Pre-SPR and was replaced by a second mentor for the rest of the program.

The Pre-SPR and Post-SPR were based on 1000-word manuscripts that had been submitted to a peer-reviewed EER journal (different from the journal to which the manuscripts they reviewed with their triads were submitted) and were used with permission from the authors for research purposes. Participants were told to review the Pre- and Post-SPR manuscripts assuming they had been submitted to an engineering education journal. Triads completed an SPR on the same manuscripts, one for the Pre-SPR and one for the Post-SPR.

The Pre-SPR article and SPR form (Figure 2) were introduced to participants at the start of the training program during orientation prior to the first triad meetings. Strengths and weaknesses from the Pre-SPR were used to develop the initial

list of codes, which were adjusted, added or combined throughout analysis to increase interrater reliability.

The final codes were organized into six themes: Context, Methods, Results, Discussion, Mechanics and Structure, and EER Relevance. Within each theme, codes were organized as strengths (positive attributes) and weaknesses (negative attributes) (Figure 3). Once codes were finalized, they were inserted into the SPR form as drop-down lists that respondents could select within strengths and weaknesses in the Post-SPR. The 200-word justification of the recommendation to the editor was kept as an open-ended response.

After each triad had completed three reviews, they were sent the manuscript to review for the Post-SPR. All of the 200-word justifications of the recommendations to the editors were coded independently by two researchers with an interrater reliability (IRR) of 70%. IRR was determined by having each coder independently assign SPR codes to relevant sections of the 200-word justification of the recommendation to the editor. In some responses, a single code appeared more than once; however, a single SPR code was only counted once in each 200-word response. IRR was calculated by dividing the total number of codes cited by each rater by the number of codes upon which both raters agreed. To ensure integrity of the analyses, only the codes that both independent coders agreed upon were used.

Structured Peer Review

Greetings JEE Mentored Review Program Mentee or Mentor, Use this template to complete Structured Peer Reviews (SPRs). SPRs are a way to begin thinking about a manuscript review. Your responses may be first impressions that may change; there are no right or wrong answers. We will not share these responses with your triads, but feel free to use them as a start to your discussions.

Thank you,
The JEE Mentored Review Program Research & Evaluation Team

1. Email
2. Your Name
3. Who is your mentor (choose yourself if you are the mentor)?
4. Manuscript Title
5. Which SPR are you completing?
 - a. Pre-SPR
 - b. Manuscript 1
 - c. Manuscript 2
 - d. Manuscript 3
 - e. Post-SPR
6. What are up to five significant strengths of this manuscript?
7. What are up to five significant weaknesses of this manuscript?
8. What would be your recommendation to the editor?
 - a. Accept As Is
 - b. Minor Revision (revise and resubmit; no need for another round of reviews)
 - c. Major Revision (revise and resubmit; needs another round of reviews)
 - d. Reject
9. Explain the basis of your recommendation for the editor in 200 words or less.

Fig. 2. Structured Peer Review (SPR) questionnaire used for the Pre-SPR.

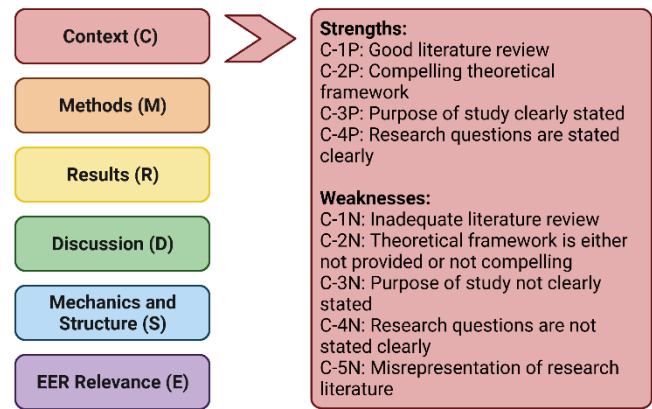


Fig. 3. Categories and codes used in analysis.

The Pre- and Post-SPR response rates were both 96% (25 of 26 participants). In the following analyses, we report results for codes that evidenced convergence by at least 50% of reviewers (which we defined as “convergence”) in strengths, weaknesses, or justification. It is important to note that some codes had 50% convergence in strengths or weaknesses *and* the 200-word justification to the editor. Our research team assumed that convergent codes reflected in the strengths or weaknesses *and* the justification were more important to reviewers than codes that were only identified within strengths or weaknesses and not mentioned in the justification. In Figure 4, the codes referenced in the strengths or weaknesses *and* in the justification are denoted with a superscript.

Due to the limited size of our sample, to study differences in reviewers’ evaluations across geographic contexts we split our comparison into two geographic regions based on where participants earned their Ph.D. degrees: The United States (U.S.; 19 participants) and Africa, Australia, Canada, and Europe (AACE; 8 participants).

IV. PRELIMINARY RESULTS

The Venn diagram in Figure 4 illustrates the summary of what U.S. and AACE participants paid attention to in their SPR-Pre and SPR-Post. In their responses to the Pre-SPR, 17 U.S. reviewers converged at least 50% of the time in noticing three aspects of the manuscript within two categories (Context and Methods). The eight AACE participants converged only on *Inadequate literature review* (code C-1N) at least 50% of the time for the SPR-Pre. Both groups converged on *Relevant to EER and/or Timely* (e.g., COVID) (code E-3P) in reviews of the Pre-SPR. In their responses to the Post-SPR, the six AACE respondents and 19 U.S. respondents both converged at least 50% of the time in noticing C-1N and *Research questions not stated clearly* (code C-4N). The AACE respondents converged on C-1N and U.S. respondents converged on C-4N in their responses to both the weaknesses of the manuscript and their 200-word justifications to the editors (denoted by the superscripts in Figure 4). U.S. participants also converged on one additional code (E-3P). These results indicate possible differences in the tacit understanding of what reviewers construe as important indicators of value and quality of EER

scholarship based on the cultural norms of the geographic region where they received their Ph.D. training [11].

It is also interesting to note that the ratio of the of Major Revisions to Reject decisions remained consistent for both U.S. and AACE reviewers across the Pre- and Post-SPR. Both of the 1000-word articles used for research purposes for the Pre- and Post-SPR were chosen for their similarities in quality; both had received recommendations of Major Revisions and Reject from peer reviewers, with the editor ultimately deciding on Major Revisions. We are unsure if these differences are a result of cultural norms or an inherent predisposition of certain scholars to choose Major Revisions over Reject and vice versa.

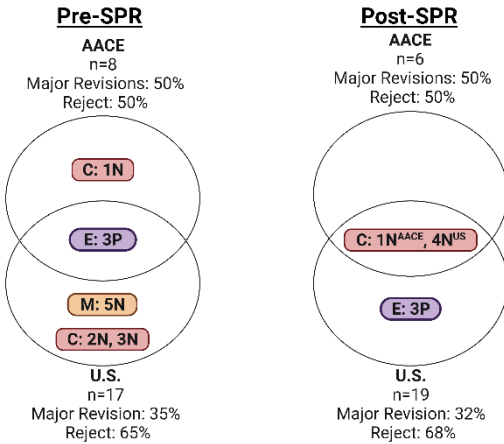


Fig. 4. Comparison of the codes used to characterize what reviewers from the U.S. and from Africa, Australia, Canada, and Europe (AACE) noticed at least 50% of the time. Codes for context (C) are defined in Figure 3; other codes are defined as Inadequate description of research methods (M-5N) and EER Relevance: Relevant to EER and/or Timely (e.g., COVID) (E-3P). The superscript by codes C:1N and C:4N indicates that these aspects were identified in the strength/weakness lists as well as the 200-word justification to the editor.

V.

VI. FUTURE DIRECTIONS AND CONCLUSION

The results from our preliminary study are from 19 U.S. and eight AACE scholars from the first cohort of the peer review training program. These results show possible differences between the criteria that reviewers from different geographic regions use to evaluate manuscripts; this will be further investigated with data from additional cohorts. We acknowledge that this is a relatively small sample, however, other studies that investigate peer review are also limited to a small sample, typically two to three reviews for a single manuscript [9]. The use of the abbreviated manuscripts allowed us to collect 25 responses for two separate manuscripts, a much larger sample in comparison. Additionally, we have data collected from two additional cohorts (49 new participants), 14 of whom received their Ph.D. degrees from a university outside of the U.S. This larger sample will allow us to expand our analysis and investigate trends based on continent or even country. Some of the participants in the AACE group in the first cohort received their Ph.D. training in Canada. However, to preserve the anonymity of the participants as well as aid in statistical analysis, we plan to group by continental region in

future analyses until we have collected an adequate sample from any one specific country.

Figure 5 shows the current breakdown of the continents where participants received their Ph.D. in our combined cohorts, with 56 from North America and six each from Africa, Australia and Europe. In addition to the SPR data that we shared in this paper, we have also conducted think aloud protocol (TAP) or cognitive interviews with participants. These interviews will allow for a deeper analysis on the similarities and differences about what these individuals are taking note of as they conduct their reviews. It will also allow for exploring the tacit understandings about value and quality of EER scholarship that are shared by scholars trained from around the world, and how those understandings are shared between mentors and mentees. We further plan to analyze how the ratio of positive to negative comments may vary by continental region.

We acknowledge that engineering education is a global enterprise and that scholarly review, particularly in doctoral credentialing, varies by geographic region. We are thus motivated to further explore differences seen in this preliminary data with the expectation that understanding and valuing these differences can lead to better peer reviewer training as well as a more inclusive global community of scholars.



Fig. 5. Summary of the continents where the participants received their Ph.D. training for three cohorts of the program combined ($n=72$ total).

ACKNOWLEDGMENT

We thank our study participants for their time in completing reviews, surveys, and interviews. This material is based on work supported by the National Science Foundation under Grant #s EEC-2037807, 2037788, 2037797. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] M. Hojat, J.S. Gonnella & A.S. Caelleigh, "Impartial judgment by the "gatekeepers" of science: fallibility and accountability in the peer review process," *Advances in Health Sciences Education*, 8(1), 75-96, 2003.
- [2] A. E. Carroll, "Peer Review: The Worst Way to Judge Research, Except for All the Others." *NY Times*, November 6, 2018. Retrieved from <https://nyti.ms/2yRcClr>.

- [3] N. J. Silbiger & A. D. Stubler, "Unprofessional peer reviews disproportionately harm underrepresented groups in STEM." *PeerJ*, 7:e8247. 2019. <https://doi.org/10.7717/peerj.8247>
- [4] B. K. Jesiek, L. K. Newswander & M. Borrego, "Engineering education research: Discipline, community, or field?" *Journal of Engineering Education*, 98(1), 39-52. 2009. <https://doi.org/10.1002/j.2168-9830.2009.tb01004.x>
- [5] L. Benson, R. Bates, K. Jensen, G. Lichtenstein, K. Watts, E. Ko and B. Albayati, "Building Research Skills Through Being a Peer Reviewer," Paper presented at the 2021 Annual American Society of Engineering Education (ASEE) Conference and Exposition, Virtual Conference, July 26 - 29, 2021.
- [6] D. Rumelhart, "Schemata: The building blocks of cognition," In: R. J. Spiro, B. C. Bruce & W. F. Brewer. (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum, 1980.
- [7] J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science*, 12: 257-285, 1988. https://doi.org/10.1207/s15516709cog1202_4
- [8] R. Axelrod, "Schema Theory: An Information Processing Model of Perception and Cognition," *The American Political Science Review*, 67(4), pp. 1248-1266, 1973. <https://doi.org/10.2307/1956546>
- [9] D. W. Fiske and L. F. Fogg, "But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments," *American Psychologist*, vol. 45, no. 5, pp. 591-598, 1990, <https://doi.org/10.1037/0003-066X.45.5.591>
- [10] B.R. Clark (Ed.) *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan*. Berkeley, CA: California University Press. 1994.
- [11] S. Delamont, P. Atkinson & O. Parry (1997) Critical mass and doctoral research: Reflections on the Harris report, *Studies in Higher Education*, 22:3, 319-331. <https://doi.org/10.1080/03075079712331380926>
- [12] D. Riley, J. Karlin, J. Pratt and S. Quiles-Ramos, "Board # 127: Building Social Infrastructure for Achieving Change at Scale." Paper presented at 2107 ASEE Annual Conference & Exposition, Columbus, Ohio. <https://doi.org/10.18260/1-2-27722>.