Type: Research

Poster Reviews – Can We Do Better?

Michael Spiess
California State University, Chico

College of Agriculture
California State University, Chico
Chico, CA 95929-0310
Telephone: (530) 898-4554
E-mail: mspiess@csuchico.edu

Type: Research

Poster Reviews – Can We Do Better?

Poster abstracts submitted for presentation at the meetings of the American Association for Agricultural Education (AAAE) are reviewed by faculty and graduate students of the Association. Authors and poster session chairs often comment on the lack of agreement between reviewers. Reviewers are asked to give a numerical score to the abstract based on a list of criteria and to provide a recommendation for acceptance. Posters provide an important entry point for sharing professional activities. The review process not only determines which posters will be accepted, but should provide important feedback to the author.

Rubrics that address individual review criteria have been developed for poster reviews (AAAE, 2019) but these are not readily available to reviewers. No rubric has been developed for the acceptance scoring. Poster chairs are tasked with accepting posters to each conference. The number of accepted posters will vary somewhat with the limitations of the venue. General guidelines for acceptance are provided in the Poster Session Guidelines (AAAE, 2019). Generally, a blend of the acceptance scores and poster (numeric) scores are used to choose what posters are accepted. For many years a ranking by the mean Z score method developed by Franklin and Elliot (Franklin, 2003) has also been used as a means of evaluating posters for acceptance. This method converts a raw score to a Z score for each reviewer. The mean of the Z scores is then used to rank submissions. The purpose of the study was to quantify the issue of inconsistency. Objectives of this study were to: 1) report the number and types of poster reviews of national conferences 2) quantify the variation in criteria scores 3) quantify the disagreement between reviewers, and 4) compare criteria scores with accept scores.

Conceptual Framework

Literature on peer review suggests that peer reviews are important to many professions (Daniel, 1993) (Baethge, 2013) (British Academy, 2007). Inter-rater reliability has been explored in the AAAE Conference paper submissions recently and found that there are large variances in peer review (Shoulders, 2015). Rothwell found "agreement between reviewers in clinical neuroscience was little greater than would be expected by chance alone" when researching paper submissions to journals and conferences (2000). Inconsistency appears to be a common issue in peer review.

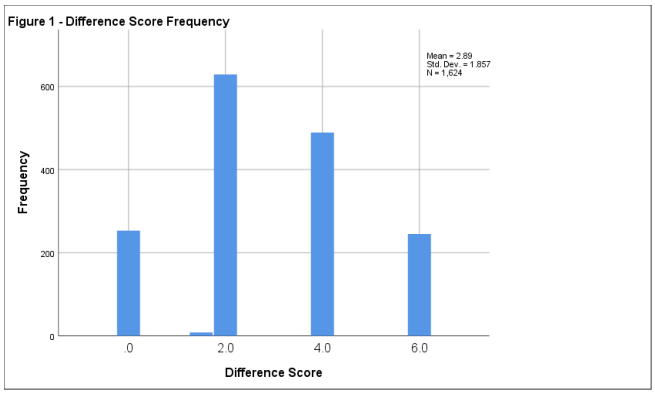
Methodology

Poster reviews from 2011 to 2020 national AAAE meetings were extracted from the AAAE poster submission system. The reviews are composed of scores given to each of a list of criteria and an "accept" score where reviewer recommends acceptance on a four point Likert scale: accept (4), possibly accept (3), possibly reject (2), and reject (1). A coefficient of variation (CV) was calculated for the reviews of each poster as suggested by Freund & Wilson (1993) and used by Shoulders, et. al. (2015). A simple agreement score can be used to measure agreement between reviewers. In this method if reviewers scored a poster 4 3 1 or 4 3 2 the agreement score (average number of agreements) is 0. If reviewers scored a poster 4 4 4 or 111 the agreement score is 1. To further quantify the discrepancy in agreement a discrepancy score (DS) was derived by adding the absolute value of accept score differences between reviewers. Almost all posters were reviewed by three reviewers, however this is not always true so the discrepancy scores were normalized by dividing by the number of reviews then multiplying by 3 (the normal number of reviewers). A DS of 0 indicates all reviewers agree. If reviewers scores 4 3 1 then the DS of 6 is computed. Analysis for this study was done using MS-Access and SPSS (V25).

Results

For the 10 national events 4961 reviews of 1663 posters were analyzed. 771 posters were innovative idea and 892 were research. For purposes of analysis posters with a missing review or those with one or more disqualification recommendations were omitted from the analysis. Posters are not reviewed or disqualified because they do not meet the submission requirements. 4847 reviews for 1620 posters were included in the analysis. Since the two types have different criteria they were generally separated for analysis. It is interesting to note that there has been a steady trend of submission of more research posters than idea posters. In 2011 research submissions were 46% of the total and in 2020 research submissions were 66% of the total. 295 reviewers were used during this period. The greatest number of posters reviewed by a reviewer was 74 and the least was 5. The mean number of reviews by reviewer was 17. The average number of events reviewed by reviewers was 2.4. 75% of reviews contained a comment to the author. More comments were provided for posters with low accept scores (reject-87%, possibly reject - 80%, possibly accept-73%, and accept -- 68%).

Accept scores were highly correlated with criteria total scoring; research posters $p_{(x,y)}$=.820 and for idea posters $p_{(x,y)}$=.824. The majority of individual criteria also had strong correlations; only criteria for references and grammar had weak correlations (.245-.397). The mean CV for idea posters was 15.39% and for research posters 16.82%. These values are less that the 19.65% reported by Shoulders (2015) study of 336 AAAE Conference paper reviews. Data was examined to see if the reviewers scored posters (typically 5-8) for an event in the same order as the accept recommendation. 3% of scores were found to be inconsistent with the accept recommendation order (a lower scoring poster was given a higher accept recommendation). This finding is supported by the high correlation between poster scores and accept recommendations. Data was examined to see if the mean Z score ranking agreed with the ranking by mean accept recommendation. 36% of the reviewed posters did not agree. Perhaps the most telling statistic is that overall only 15.6% of reviewers agree and 45.2% strongly disagree (DS>2) see Figure 1.



Conclusions and Recommendations

Unlike the review process for journal papers the review process for posters involves a large number of reviewers and a small window of time. For many reviewers this is their first review experience. Given the strong association of score to accept recommendation simply encouraging reviewers to use the existing scoring rubrics may improve the consistency of the acceptance recommendation. Consideration should be given to training, possibly by video. Rubrics should be more accessible reviewers. Incorporating the rubrics into the online scoring system (similar to rubric scoring in a Learning Management System) where the reviewer could choose directly from the rubric might help to provide more consistent scoring. If using a rubric produces more consistent numeric scores between reviewers perhaps the acceptance recommendation would not be necessary. If changes are made in the review process (training, rubrics) this study should repeated to see of these efforts improve consistency.

References

AAAE. (2019). *AAAE Poster Session Guidelines.* Retrieved from aaaeonline.org: http://aaaeonline.org/resources/Documents/AAAEPosterSessionGuidelines_UPDATEDJune2019.pdf 1/3/2020

Baethge, C. F. (2013). Substantial agreement of referee recommendations at a general medical journal - a peer review evaluation at Deutsches Ärzteblatt International. *PLOS One, 8*(5), 1-7. Retrieved from http://europepmc.org/articles/PMC3642182

British Academy. (2007). Peer review: The challenges for the humanities and social sciences. London: The Academy.

Daniel, H. D. (1993). Guardians of science: Fairness and reliability in peer review. Weinheim: VCH Publishers.

Franklin, E. (2003). Personal Communication.

Freund, R. J. (1993). *Statistical methods.* San Diego, CA: Academic Press.

Rothwell, P. M. (2000). Reproducibility of Peer Review in Clinical Neuroscience. Is Agreement Between Reviewers Any Greater Than Would Be Expected by Chance Alone? *Brain, 123*(9), 1964-1969. Retrieved from https://academic.oup.com/brain/article-pdf/123/9/1964/794263/1231964.pdf

Shoulders, C. J. (2015). Peer Review in Agricultural Education: Interrater Reliability of Manuscript Reviews for the 2014 National Agricultural Education Research Conference. *Journal of Agricultural Education*, *56*(3).