

Adam Hain

CMU DET Program Cohort 4

Critical review of literature #2

Article Reviewed:

Ertmer, P. A., Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., ... Mong, C. (2007). Using peer feedback to enhance the quality of student online postings: An exploratory study. *Journal of Computer-Mediated Communication*, 12(2), 412–433.

Problem

1. *Identify the clarity with which this article states a specific problem to be explored.*

Ertmer et al. (2007) identify the research problem by stating the purpose of the study in a clearly delineated section of the article. This section concludes the review of literature and starts by reiterating how feedback has been shown to be beneficial. This statement is likewise addressed in the preceding literature review. From this premise the authors explore how feedback can be best implemented.

Next, Ertmer et al. (2007) identify two gaps in the literature. First, use of peer feedback online has not been addressed thoroughly. Second, research on the impact of peer feedback on discussion quality is sparse. The purpose of the study aims to address this by examining student perceptions of feedback as well as measuring quality of discussion postings. Three specific research questions, addressing these aims are clearly stated in this same section.

This section could not be structured with much additional clarity. In one concise segment, the authors summarize the literature review, identify the gaps in literature, then state the specific research problem and questions. A sole critique is that the initial sentence in this section contains a few different ideas and might benefit from being rewritten.

2. *Comment on the need for this study and its educational significance as it relates to this problem.*

While the authors identify gaps in the literature, this alone does not warrant investigation. However, in the literature review, the importance and prominence of online discussion is addressed in the introductory section. Another problem, not stated specifically in the research questions, is that grading of online discussion is particularly onerous on instructors. So, an implied over-arching research question is, “How can we encourage high-quality online discussion, while decreasing instructor workload?” Peer feedback is suggested explicitly as one possible solution. The literature review then appropriately discusses advantages and challenges of peer-review.

The prominence of online discussion is verifiable in the literature, as well as anecdotally, through the experience of most instructors in online environments. Additionally, the difficulty involved in grading discussion, especially in large classrooms might be described as inherent in the pedagogical technique. I would argue the problem does exist and is significant enough to the broader field to merit further research.

3. *Comment on whether the problem “researchable”? That is, can it be investigated through the collection and analysis of data?*

The specific research questions stated in the article aim to examine student perceptions of peer vs. instructor feedback and quality of discourse in online discussion. These address the stated problem, three research questions, as well as the more general problem of instructor overload discussed above. Student perceptions and quality of discourse are measurable in a variety of ways, as is demonstrated in the article.

In addition to the techniques used to operationalize and measure these concepts, others could have been chosen and be equally appropriate. Student perceptions could be measured using numerous high-quality validated survey instruments and could additionally be measured longitudinally, through the course of the study. Quality of discourse would most likely be measured using a rubric - as was done in the study using a rubric based on Bloom's Taxonomy. However, other rubrics could be used as well. This technique provides the advantage of quantitative measures applied to qualitative data in terms of numeric rubric scores.

In sum, student perceptions and quality of discourse are measurable phenomena. Aside from the techniques chosen, others could also be used.

Theoretical Perspective and Literature Review

4. Critique the author's conceptual framework.

To effectively critique an author's conceptual framework, criteria must first be established for this examination. Antonenko (2015) offered a discussion of conceptual frameworks in educational technology research, citing definitions from literature and a critique of the varied terminology used to describe conceptual frameworks. For example, the terms "conceptual framework" and "theoretical framework" are often used interchangeably. Antonenko (2015) argues this is imprecise, that theories are collections of related concepts and distinguished from conceptual frameworks which include the assumptions, interests, and beliefs of the researcher. Theories are frameworks which contain concepts, while conceptual frameworks are more varied, built of a combination of these theoretical frameworks, individual concepts, and observations that "are custom designed by researchers based on personal assumptions and epistemological beliefs, experiential knowledge, and existing

(formal) theories for each individual study” (Antonenko, 2015, p. 57). While Ertmer et al. (2007) did not define a conceptual framework explicitly, in adopting this perspective one can examine the conceptual framework presented in the literature review.

Four primary strands can be identified in the literature review. These are: discussion in online and face-to-face (f2f) environments, the role of feedback in instruction, feedback in online environments, and the advantages and challenges of peer feedback. All four strands are directly related to the research problem and questions.

A critique I would offer is that the introductory section is in fact a strand related to online and f2f discussion and not delineated as a section of the review. However, it is a major, distinct theme. Instead it is listed under the “Introduction” heading. I might instead create a broader introduction and provide this strand its own section in the literature review. Additionally, since the topic is so key to the research problem, I would also offer more detail on it generally in the review.

I also found the first sentence difficult to defend: “Student discussion has been identified as a key component of interactive online learning environments; both instructors and researchers agree that this is where the “real” learning takes place” (p. 412). While this statement is cited, I feel it is far too broad for even an experienced researcher to make and suggests preconception in this opening sentence. While I agree that online discussion is important, I would suggest that “real” learning takes place in many different forms. I would also be surprised if there is anything researchers and instructors *all* agree on. Even if this could be shown in the literature, changes in online pedagogies as well as changes to the varying implementations of online discussion would make this claim would be difficult to support over time. This is a surprising opening to such a well-executed article.

5. *How effectively does the author tie the study to relevant theory and prior research? Are all cited references relevant to the problem under investigation?*

As a whole, the literature cited is relevant to the problem investigated. I'll discuss each section briefly.

The introductory (discussion) section cited literature current at the time. Citations were not particularly varied, with citations from the same Black (2005) article several places. However, the Vygotsky (1978) reference showed the researchers' familiarity with the broader implications of social cognitivism as a basis for discussion in educational environments.

Section two (feedback) was concise and relevant. It summarized previous literature for the unfamiliar reader offering a useful list of the role feedback should play in educational environments. I might add a bit more detail about how feedback functions as formative assessment and if particular types of feedback are shown to be more effective in the literature.

Section three (peer feedback advantages) provided an interesting set of arguments for the use of peer feedback, that are referenced appropriately. There is one statement, "If used effectively, both instructor and peer feedback have the potential to increase the quality of discourse, and thus the quality of learning, in the online environment" (p. 415). While I might agree with this statement generally, it may be more appropriately stated as an element of the research questions in the study. This statement appeared to refer to results of the research before stating evidence from the study. If there is evidence of this in previous literature, then it should be cited since it is clearly related to the research questions.

Section four (peer feedback challenges) provided an informed portrait of the challenges and complexities of peer feedback. I found it helpful to break the two sections apart as well. I have no critiques for this section.

All sections were directly relevant, informing the research problem which followed in the next section. They were concise and free of needless jargon. The review was clearly structured beginning with a broad topic and narrowing scope quickly with relevant discussion.

6. *Does the literature review conclude with a brief summary of the literature and its implications for the problem investigated?*

While not stated as such, the Purpose section provided a brief summary of the literature review as it relates to the problem. The first two sentences touch on each of the topics discussed previously.

This works well in this case, but I think a brief summary could have been added. These two sentences function more to identify a gap in the literature than to summarize the review. However, the strongest point of this section is not just that a gap is identified, but that relevance was established using the literature review as a backdrop.

7. *Evaluate the clarity and appropriateness of the research questions or hypotheses.*

There are three research questions, clearly stated in a list just after the purpose statement. This seems the ideal format to state research questions in terms of structure. Regarding the questions' content, they are clearly written, specifically measurable and appropriate to the study and background. These three questions were measured and discussed in the remainder of the article. Further, they addressed the stated gap in the literature.

Research Design and Analysis

8. Critique the appropriateness and adequacy of the study's design in relation to the research questions or hypotheses.

The research questions were concerned with two things: quality of discourse and student perceptions of peer feedback. So, the reader would likely expect to see survey data or interviews to gauge perceptions and various qualitative methodologies to code quality of discussion postings.

The design the authors chose employed a combination of these. There were three sources of data: surveys, interviews, and the scoring of quality of discourse. In terms of study design, the scoring of discourse seemed the most complex. Scores were based on Bloom's Taxonomy, a tool familiar to both the researchers and the students, as it is well-known to most in the field of education. A rubric based on Bloom's was adapted from previous literature and then applied to score discussion questions. This was modeled first by the instructor to provide examples, then implemented midway through the semester by the students. Qualitative feedback on rubric scores was also provided. All data were processed by the instructor before being distributed to individual students.

This scoring data served to answer the components of research question one. These components addressed quality of discourse. I felt this was a creative way to quickly teach students to essentially code qualitative data, using a somewhat universally recognized tool. Additionally, it provided an opportunity to capture the peer feedback data simultaneously. Overall, it was a practical way to measure quality of discourse.

As to whether Bloom's Taxonomy is the appropriate instrument to use to measure quality of discourse, especially in online environments, I am not as sure. While I agree that some

rubric must be used to objectively judge the discourse, Bloom's might be a bit vague for this application. Additionally, individuals have varying prior experience with the tool, which might color its application. While using Bloom's had the advantage of being familiar to students, any rubric could have been integrated into the study in the same way. I would argue that other rubrics might be more effective at judging quality of online discourse.

Survey and interview data were collected to answer the remaining two questions both having to do with perception of peer feedback. I feel these were appropriate in this case. The use of both methods allowed for triangulation of data overall, but little information is provided on the instruments used.

9. Critique the adequacy of the study's sampling methods (e.g., choice of participants) and their implications for generalizability.

The sample size was a fairly small (n=15) convenience sample consisting of a one semester class of students. Differences in participants' background are well-detailed. A small sample size is common in qualitative research, because of the large amounts of data that are usually produced (McMillan, 2012). I felt the sample size was adequate, especially for an exploratory study. That said, sample size was mentioned as a limitation in the discussion.

However, I think the sample population choice limits generalizability. Quality of discourse and student perceptions are being measured for this very specific group - graduate students studying education. Undergraduate students in other fields would likely be less skilled or willing to apply Bloom's Taxonomy to their peers' discussion since it requires an educational background. To be fair, this is an exploratory study, as stated in the title, but it would have to be modified and repeated in various environments to find if the results truly generalize.

10. Critique the adequacy of the study's procedures and materials (e.g., interventions, interview protocols, data collection procedures).

As discussed above, there were three data measures collected. First, there was quality of discourse data as measured using the adapted Bloom's Taxonomy rubric. The first half of the semester was scored by the instructor, the second by the students. However, all posts were scored by the researchers to gauge the change in quality of discourse over the semester.

There were a couple of problems with this instrument discussed in the study. First, the scale did not provide much variation (0-2) making it hard to show change longitudinally. This may have caused a ceiling effect in the student scores, as most students were not willing to give scores of zero. This lack of fidelity in the instrument is recognized as a limitation of the study. Next, the instrument is somewhat vague and may be difficult to apply for students, hindering inter-rater reliability. This was mentioned in the qualitative results. While the researchers took standardized steps to ensure inter-rater reliability in their own analysis of postings, the same could not be done with students.

There was also a problem systematically with grading the quality of all postings. Two initial postings were required and graded for quality, but postings after that were not. Therefore, students had little motivation to maintain the same level of discourse. However, all postings were collected and rated by the researchers. The researchers were not able to sort which postings were the initial two graded submissions. So, postings of varying quality were mixed together.

Beside this scoring data, student perception data were collected using surveys and interviews. The inclusion of three types of data aided in triangulation. Various samples are provided of interview and survey questions. However, little information is provided on the

survey instrument or the interview questions, so it is hard to comment on them. Despite the lack of detail, in terms of procedures, the data appear to be collected using accepted research techniques.

11. Critique the appropriateness and quality (e.g., reliability, validity) of the measures used.

Validity is described by McMillan as “a judgment of the appropriateness of a measure for the specific inferences or decisions that result from the scores generated by the measure” (2012, p. 131). He argues “it is the inference that is valid or invalid, not the measure” (2012, p. 131). Using the measures above, the authors explore the perceptions of peer feedback and quality of discourse in online discussions.

Randolph (2007) defines threats to internal validity as “factors that can make it appear that the independent variable is causing a result when in fact it is a different variable or set of variables that is causing the result. In other words, threats to internal validity in experimental research are hidden causes” (p. 49). He defines 7 types: history, attrition, regression to the mean, maturation, instrumentation, testing and selection. None of these appear to be threats in this study. Additionally, since this is not an experimental study, with independent variables and controls, these types of experimental validity are less relevant. Validity is established in this study primarily using triangulation.

Additionally, little is said about the survey, suggesting it is not validated instrument. The authors simply state it was created collaboratively. Likewise, the rubric is not provided so neither can be verified by the reader.

However, validity and reliability issues are primarily addressed using triangulation of the data. In this way, perceptions stated in interviews and on surveys should match and could be verified by researchers. Interview data was also member checked with participants. It isn't

stated to what degree this took place or when longitudinally it occurred, but member checking is a common method for validating interview data (McMillan, 2012). The coding of qualitative data was performed by all researchers, in a robust and standardized way.

The authors did spend some time describing the process of ensuring inter-rater reliability in scoring of discussion posts among the researchers. The process described seems logical and also robust.

Validity of the scoring rubric, in terms of student use, is provided by instructor examples prior to beginning the course, and instructor feedback for the first half of the semester. It is also argued that since students and instructors were both familiar with Bloom's, face validity is ensured. Other measures of construct validity are not addressed, however I would argue that content validity may be an issue (Drost, 2011). While the use of Bloom's is easy to implement, the researchers might have spent more time arguing that this measure accurately represents quality of discourse. This, combined with the ceiling effect, created a somewhat low fidelity measure, that might be improved in future replications of the study.

External validity refers to issues of generalizability to a target population (Drost, 2011). As discussed previously, the generalizability of this sample to the larger population of higher education students, even of graduate students specifically, is questionable. However, this might be expected in an exploratory study of this type.

Overall, the researchers address issues of reliability and validity systematically, in a dedicated section. Accepted research techniques to ensure triangulation are used. However, inclusion of the survey instrument, rubric and interview questions would strengthen the article.

C. Interpretation and Implications of Results

13. Critique the author's discussion of the methodological and/or conceptual limitations of the results.

Much of the discussion of limitations is related to validity and reliability and covered above. However, there were also several limitations addressed in the discussion portion of the article. The first limitation involved problems with the implementation of the scoring rubric. Its low fidelity produced a ceiling effect since student scores were somewhat inflated. Therefore, growth could not be measured since most scores began fairly high. The design feature of having only the first two postings scored also confounded the results, as the researchers were unable to locate graded and non-graded posts after the fact.

The use of Bloom's Taxonomy in relation to the discussion questions used was also noted as a limitation. Some discussion questions may not have inherent need for answers at the higher levels of Bloom's. Thus, postings might have been of high quality, yet scored low in terms of the Taxonomy.

Next, the channeling of all comments through the instructor was seen as a limitation. This design decision negated the increased speed in feedback usually seen as an advantage of the peer-feedback process.

All of the limitations above were noted in the discussion section, but the authors did also provide a specific section examining limitations. These included small sample size, short duration of study, and again low fidelity of the rubric instrument. Lack of robust student training with the rubric was mentioned in this section as well.

Overall, the authors provided an honest portrayal of the limitations of the results. They did not make unfounded claims and attempted to explain how the limitations affected the results

and conclusions. For most of the limitations discussed, the authors provide suggested solutions for future research to avoid the same problems. These aspects reflect a high level of objectivity which likely results from experienced researchers directing the study.

14. How consistent and comprehensive are the author's conclusions with the reported results?

The authors expected to find increased quality of discourse throughout the semester, as a result of peer feedback, but did not. They point out that this conflicts with previous literature. Since essentially this hypothesis was incorrect, the authors would have difficulty making unfounded conclusions. Instead, numerous possible causes for these findings are discussed. A key related finding is that the quality of discourse did not decrease with peer feedback. So, once a level of discourse is created, it was maintained through peer feedback.

Another conclusion was that students perceived a greater value in instructor feedback over peer feedback, both before and after the study. This is reflected in the results, both in the reported survey data and the interview data. While students perceived instructor feedback as more valuable, they also found value in giving feedback. This conclusion is also reflected in the stated results.

Overall, the conclusions are consistent with the identified results. These conclusions include discussions of limitations as well, strengthening the appearance of objectivity. A way the discussion section might have been improved is by referencing specific results in the section. In this way, the results that led to these conclusions would be more directly accessible.

15. How well did the author relate the results to the study's theoretical base?

The article references several of the same authors in the discussion and in the review of literature. These references are connected in a meaningful way. In terms of topics, the primary strands of the literature review are addressed in the discussion of results. These stands:

discussion in online environments, perceptions of peer vs. instructor feedback, and benefits of peer feedback are all discussed in light of the stated results and conclusions. There are no unexpected or added topics in the discussion. I would not change the integration of results with the theoretical base as presented in the literature review.

16. In your view, what is the significance of the study, and what are its primary implications for theory, future research, and practice?

As an exploratory study, the authors provide several valuable contributions to the literature on peer feedback and discourse quality – and provide opportunity for further, more refined exploration. We see that students value giving peer feedback, but find less value in the authenticity of the peer feedback received. It seems that students prefer instructor feedback but will accept peer feedback and perform at consistent levels with both. This is an important and practical finding. This is also an important distinction discussed in the article. Further study might explore how peer-feedback can be better legitimized in the perceptions of students receiving it.

The issue of instructor workload was stated as an initial problem in the literature review. In fact, I have incorporated both peer feedback and discussion in my course – in very different ways. So, I am very interested in ways to provide high-quality feedback to students while maintaining a realistic time commitment. Incorporating peer feedback may provide a part of the solution. Conversely, peer feedback may function better as a graded assignment for the giver, than a way to more easily provide grades for students overall. Further research may provide answers and better models for implementations.

In terms of quality of discourse, I feel the design of the study provided a good model for judging quality. There were several limitations to the design, but all could be easily modified.

Peer-rating using a shared rubric is also a creative idea, and variations of this rubric, perhaps not based on Bloom's Taxonomy might provide more specific results in various domains.

References

- Antonenko, P. D. (2015). The instrumental value of conceptual frameworks in educational technology research. *Educational Technology Research and Development*, 63(1), 53–71.
<https://doi.org/10.1007/s11423-014-9363-4>
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105.
- Ertmer, P. A., Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., ... Mong, C. (2007). Using peer feedback to enhance the quality of student online postings: An exploratory study. *Journal of Computer-Mediated Communication*, 12(2), 412–433.
- McMillan, J. H. (2012). *Educational Research: Fundamentals for the Consumer*. Pearson.
- Randolph, J. J. (2007). *Multidisciplinary Methods in Educational Technology Research and Development by Justus J. Randolph*. HAMK University of Applied Sciences, Digital Learning Lab.