BRIEF

Inter-Rater Reliability of Web-Based Calibrated Peer Review within a Pharmacy Curriculum

Alex N. Isaacs, PharmD, Monica L. Miller, PharmD, MSc, Tianyang Hu, Bailey Johnson, PharmD, Zach A. Weber, PharmD

Purdue University College of Pharmacy, West Lafayette, Indiana
Submitted March 3, 2019; accepted September 30, 2019; published April 2020.

Objective. To assess the inter-rater reliability of student evaluations submitted through a web-based calibrated peer review system versus faculty evaluations in a Doctor of Pharmacy (PharmD) curriculum.

Methods. Two calibrated peer review assignments were integrated into two sequential pharmacotherapy-based courses in the first professional year of the PharmD curriculum. To assess the inter-rater reliability of student peer and self-evaluations compared to faculty evaluations, 50 student submissions from each assignment were randomly selected for faculty evaluation, using the same evaluation criteria used by students. Statistical analysis was performed to assess the difference between the scores given by students versus faculty members and the correlation between the combined and individual assignment scores given by students versus faculty members.

Results. Mean faculty scores did not significantly differ from student peer or self-evaluation scores. When the assignments were combined, there was a significant fair correlation between scores assigned by peers and those assigned by faculty members, and between scores assigned by students (self-evaluation) and faculty members. For individual assignments, one showed significant positive correlations for both peer and self-assessment compared to faculty, while the other assignment did not.

Conclusion. Incorporating longitudinal calibrated peer review activities into the PharmD curriculum demonstrated fair inter-rater reliability between scores assigned by pharmacy student and faculty members. Further research on calibrated peer review is needed in pharmacy education to evaluate its reliability and utility. However, the current study suggests that the use of a web-based calibrated peer review system within a pharmacy curriculum could provide additional opportunities for critical evaluation and student self-assessment, as well as encourage more open-ended communication activities without increasing the grading burden on faculty members.

Keywords: calibrated peer review, assessment, pharmacy education, technology

INTRODUCTION
Pharmacy education aims to develop learners with critical evaluation and self-assessment skills not only for use in patient care delivery, but also in continued professional development. While self-assessment is an important skill, multiple studies within pharmacy education have found that some student self-assessments are less accurate than faculty evaluations of students’ work. Within higher education, peer and self-evaluation practices with training and calibration have been adopted to address these evaluation deficiencies, but has also been shown to be a beneficial learning experience for students. Peer assessment also supports student preparation for professional work environments where pharmacists may provide performance evaluations to peers or direct reports. Thus, pharmacy student participation in peer assessment allows opportunities for them to practice providing appropriate feedback in a structured learning environment, preparing them for their future careers.

In addition to enhancing student evaluation skills, peer assessment activities can be a way of incorporating open-ended written communication within a course, including within large enrollment courses because preparation and review of these activities require less faculty time. Conventional assessment methods such as multiple-choice questions have limitations compared to open-ended assessment activities. Constructing multiple-choice questions can be challenging due to the
existence of cueing which can lead students to an answer or the potential misalignment with clinical practice, in which students are not given only limited multiple-choice options, but have extensive options to consider.\textsuperscript{17,18} While open-ended questions can better mirror the decision-making processes students will use in future practice and provide the instructor with an enhanced view of a student’s thought process, the increased faculty time commitment and timeline for returning useful feedback to students are often seen as obstacles.\textsuperscript{18} However, peer review allows student pharmacists to provide timely and individualized feedback to one another for open-ended questions, which students are often seen as obstacles.\textsuperscript{18} While open-ended questions can better mirror the decision-making processes students will use in future practice and provide the instructor with an enhanced view of a student’s thought process, the increased faculty time commitment and timeline for returning useful feedback to students are often seen as obstacles.\textsuperscript{18}

While peer and self-evaluation have documented benefits, there are still some concerns regarding these types of assessments. One particular concern is the possibility that student self-assessment will be less accurate than faculty evaluation.\textsuperscript{2-6} This can be seen in cases where students provide identical or high scores for all peers due to fears of negatively impacting peer grades or due to feelings of inadequacy as an evaluator as a result of limited clinical experience.\textsuperscript{2-4,6} One means of facilitating appropriate peer and self-assessment is through calibrated peer review.\textsuperscript{5} The process of calibrated peer review includes assignment submission and calibration, followed by peer and self-evaluation.\textsuperscript{5} To ensure grading accuracy, after students submit their assignments, they complete a calibration phase. During this phase, students review sample student submissions and assess these based on set criteria. Once they demonstrate evaluation standards comparable to those of faculty members, they are allowed to proceed to conducting peer reviews and self-evaluations. The purpose of the calibration process is to train students in using the evaluation criteria to limit variance and enhance the objectivity of their assessments.\textsuperscript{5} By incorporating calibration or standardized review into peer assessment, data from higher education demonstrates the strong levels of inter-rater reliability and positive student perceptions of calibrated peer review as an assessment method.\textsuperscript{5-16}

As the Academy embraces and implements the Pharmacists Patient Care Process (PPCP) within their curriculum, a continuing challenge is assessing students’ true application and mastery of each PPCP phase in large-enrollment courses.\textsuperscript{17-19} Calibrated peer review allows the student to demonstrate application of the PPCP not only through the provision of the original submission, but also further opportunity for growth through critical evaluation in the review of calibrated, peer, and self-submissions. Notably, calibrated peer review can help accomplish this beneficial learning experience for students while not adding to the grading workload of faculty members.\textsuperscript{5-16} While other areas of education have evaluated the use of calibrated peer review, there have been few published studies within health professions education on this assessment technique.\textsuperscript{5-16} The purpose of this research was to assess the inter-rater reliability of evaluations for free-response assignments within a web-based calibrated peer review software in sequential pharmacotherapy-based courses within a Doctor of Pharmacy (PharmD) curriculum.

**METHODS**

Calibrated peer review assignments were incorporated into two sequential pharmacotherapy-based courses in the first professional (P1) year of the curriculum: Introduction to Patient Centered Care (fall of P1 year) and Integrated Pharmacotherapy 1 (spring of P1 year). Courses were primarily delivered through team-taught didactic instruction by faculty members in pharmacy practice and the basic sciences. Courses in these areas provide the foundation for the PPCP. However, because of the school’s large enrollment of approximately 150 students, the primary mode of assessment was multiple-choice examinations. Thus, calibrated peer review was implemented as a supplemental assessment activity.

For each calibrated peer review activity, faculty members provided a guidance document on the assignment that included the focus and length of the submission as well as the deadline for submission. For the first step of the assignment, the faculty member provided students with a patient case along with prompts or questions for which they were required to provide a written response. In other areas of higher education, essays are often used as the prompt for a submission that will be peer reviewed, but patient cases can be an appropriate alternative to essays for health professions education.\textsuperscript{5-14} Students were given instructions with each activity to respond to specific PPCP questions about a patient case. Each calibrated peer review activity included multiple, diverse PPCP components, but not necessarily every part of the PPCP depending on the focus of the activity and its timing within the curriculum. Students were given one week to submit a written response to the prompt of less than 1000 words. After reviewing the instructions and responding to the prompt, students submitted their response through the web-based calibrated peer review program, Gradient, developed by Purdue University’s information technology department in 2012.

In the second step of the process, students had their evaluation skills calibrated by independently assessing three faculty-provided sample submissions. These samples varied in quality (high, medium, and low) to enable students to distinguish between levels of performance.
To facilitate calibration and evaluation, faculty members clearly defined performance criteria within rubrics, which students were able to refer to when evaluating submissions. For the calibration component, students were required to demonstrate alignment with an established, predefined faculty score. If a student did not meet these expectations, they could repeat the calibration exercises until acceptable alignment was attained. The acceptable of 10% variation in the calibration scores between student and faculty evaluations was predetermined by faculty members.

After completing the calibration process, the web-based calibrated peer review software program randomized and anonymized three peer submissions for each student to assess. In the third step of the process, students evaluated peer assignments on the same predefined performance criteria used for the calibration process. Students were required to provide quantitative feedback along with qualitative feedback for each submission reviewed.

Next, students they used the same predefined performance criteria they had used to conduct peer reviews to evaluate their own work. Students had one week to complete the calibration, peer, and self-evaluation steps. Finally, after the review process was completed for peers and self, students were able to view their scores in addition to peer feedback via the calibrated peer review software. The final score for each individual assignment was calculated based on the quality of the submission (response to the prompt) as evaluated by peers (50%), calibration performance (20%), peer review completion (20%), and accuracy of self-evaluation compared to peers (10%). Students received full credit if their self-evaluation score was within 10% of the median peer review score.

To assess the longitudinal use of calibrated peer review in pharmacotherapy-based courses, we sought to determine the inter-rater reliability of student peer and self-evaluations compared to faculty evaluations. Two academic years of data, 2015-2016 and 2016-2017, were compiled from the calibrated peer review program. Assignments from each course for which each step had been completed were eligible for study inclusion. Assignment submissions meeting this inclusion criteria were anonymized and then 50 assignments were randomly selected to be evaluated by two separate faculty evaluators. For faculty evaluation, two faculty members graded blinded submissions using the same predefined performance criteria used by the students at the time assignment was completed.

With the objective of the research being to assess the inter-rater reliability of the review process, the average score from each cohort was used to assess the correlation between faculty members and peer and faculty members vs students’ self-assessment scores. As the data were nonparametric, a Kendall’s tau test was performed to evaluate the correlation.20,21 The strength of correlation for Kendall’s tau was categorized as: little or none (<0.25), fair (0.26-0.5), moderate to good (0.51-0.75), and good to excellent (0.76-1).20,21 For the study, a \( p \) value of <.05 was used to assess significance. The study was deemed exempt by the Purdue University Institutional Review Board.

RESULTS

For the 2015-2016 academic year, 150 of 159 assignments (94.3%) met inclusion criteria, with the 50 randomly selected assignments representing 33.3% of the eligible cohort. For the 2016-2017 academic year, 145 of 151 assignments (96.0%) met inclusion criteria, with the 50 selected assignments representing 34.5% of the eligible cohort. When combining both assignments, the mean faculty score of 9.1 (SD=1.4) did not significantly differ from the peer evaluation score of 9.2 (SD=0.9) \( (p=.73) \). There was also no significant difference between the mean faculty score and the student self-evaluation score of 9.4 (SD=0.7) \( (p=.07) \).

Inter-rater reliability results are provided in Table 1. For student peer evaluation scores compared to faculty scores and for self-evaluation scores compared to faculty scores, there were significant correlations for the combined assignments and for one individual assignment.

DISCUSSION

This study found fair correlation between calibrated peer review by students and faculty evaluations, suggesting that calibrated peer review may be a useful assessment technique to incorporate in pharmacy courses with large enrollments. This is the first study within pharmacy education to highlight the potential use of calibrated peer review to supplement traditional assessment methods. Previous studies have shown discordance between student self-evaluation and faculty evaluation.2-4 However, the current study found no significant difference between faculty scores and student peer and self-evaluations, which could have resulted from the calibration phase.

Previous literature in higher education has demonstrated the improvement of students’ writing and evaluation skills as the result of completing calibrated peer review activities.6-16 Studies have also consistently demonstrated both the inter-rater reliability and positive student perceptions of calibrated peer review.6-16 However, there is limited literature regarding the use of calibrated peer review within health professions education.
McCarty and colleagues implemented calibrated peer review for third-year medical students to provide learning opportunities in documentation and evaluation of patient care notes using three longitudinal activities. The study results showed that medical students’ clinical writing improved with each calibrated peer review activity. In addition to enhanced clinical documentation, students demonstrated a significant improvement in evaluating clinical writing, with more student evaluations aligning with the faculty scores for the third and final assignment. Although significant, the current study only found minimal alignment between student peer evaluations and faculty evaluations for combined assignments and one individual assignment, but did not demonstrate improved alignment over completion of multiple individual assignments as the study by McCarty and colleagues did. In the current study, the lack of significant correlations for both individual assignments could have been because of the increased complexity of the second assignment.

Within pharmacy education, there is no existing published literature evaluating the use of web-based calibrated peer review. There is one study within pharmacy education highlighting peer review in a workshop series to build writing, critical thinking, and constructive feedback skills. The study by Davis found positive perceptions of student self-rated evaluation, but other endpoints were not assessed in this study. Unlike that study, in the current study, the calibrated peer review software provided enhanced functionality, allowing for randomization and anonymization of student responses.

Table 1. Comparison of Pharmacy Students’ Peer and Self-Evaluations Versus Faculty Evaluations

| Comparison Group | Kendall’s Tau | p Value |
|------------------|---------------|---------|
| Student peer evaluation vs faculty evaluation | | |
| Combined | 0.16 | <.01 |
| Assignment 1 | 0.21 | .01 |
| Assignment 2 | 0.11 | .13 |
| Student self-evaluation vs faculty evaluation | | |
| Combined | 0.28 | <.01 |
| Assignment 1 | 0.44 | <.01 |
| Assignment 2 | 0.13 | .16 |

Students were unaware of who they were evaluating, which potentially improved their objectivity when completing their evaluation. Also, a major feature missing from the study on peer review by Davis was the calibration step. Calibration is a vital component of the web-based peer review platform as it assists in standardizing assessment by requiring students to demonstrate a caliber of critical evaluation skills before progressing to peer and self-evaluation. Students are aware that their peers have been trained regarding evaluation criteria before grading their submissions. Outside of evaluation competencies associated with calibrated peer review, this assessment process enables students to provide free response answers to elaborate on the PPCP or other pharmacy concepts.

Open-ended communication activities are vital in preparing student pharmacists for future clinical experiences. Given the drawbacks to using multiple-choice questions and the faculty time required for traditional grading of free-response activities, calibrated peer review allows faculty members to require students to submit written communication without increasing time spent on grading.

This study has limitations. First, a random convenience sample of students was selected for this study rather than including the entire class. Additionally, the study focused on assessing the inter-rater reliability of overall student and faculty evaluations, rather than on assessing individual student achievement and documentation of the PPCP. However, this could be tracked longitudinally for each student through the calibrated peer review software and incorporated into programmatic assessment data. An additional limitation is the limited strength of the correlations. While significant correlations were identified, the strengths of the correlation were mostly fair with one moderate level of agreement, and not all reached statistical significance. Finally, the current study examined only two calibrated peer review assignments rather than reviews longitudinally across throughout a curriculum. Nevertheless, the study does have strengths in that it is the first study to evaluate the incorporation of web-based calibrated peer review assignments within pharmacy education. It is among the first uses of web-based calibrated peer review with patient-based activities within health professions education. While the strength of the correlations was fair between student and faculty assessments of assignments within pharmacy education. It is among the first uses of web-based calibrated peer review with patient-based activities within health professions education. While the strength of the correlations was fair between student and faculty assessments of assignments within pharmacy education.
CONCLUSION

Incorporation of longitudinal calibrated peer review activities throughout pharmacotheapy-based courses demonstrated fair inter-rater reliability for pharmacy student evaluation as compared to faculty assessments. Given that the correlations were only fair, further studies regarding calibrated peer review within pharmacy education are needed to evaluate the reliability and utility of this tool. This study introduces and demonstrates the potential of calibrated peer review within a pharmacy curriculum to increase opportunities not only for critical evaluation and self-assessment, but also for open-ended written communication activities without increasing faculty-grading burden.

REFERENCES
1. Accreditation Council for Pharmacy Education. Accreditation standards and key elements for the professional program in pharmacy leading to the doctor of pharmacy degree (Standards 2016). https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf. Published February 2015. Accessed April 11, 2020.
2. Austin Z, Gregory PA. Evaluating the accuracy of pharmacy students’ self-assessment skills. Am J Pharm Educ. 2007;71(5):Article 89.
3. Lundquist LM, Shogbon AO, Momary KM, Rogers HK. A comparison of students’ self-assessments with faculty evaluations of their communication skills. Am J Pharm Educ. 2013;77(4):Article 72.
4. Poirier TI, Pailden J, Jhala R, Ronald K, Wilhelm M, Fan J. Student self-assessment and faculty assessment of performance in an interprofessional error disclosure simulation training program. Am J Pharm Educ. 2017;81(3):Article 54.
5. Chapman OL, Fiore MA. Calibrated Peer Review: A writing and critical thinking instructional tool. The White Paper: A Description of CPR. http://myeducationchannel.weebly.com/uploads/8/0/9/7/80970730/cpr-aed-how_to_steps_-_emt_white_paper.pdf. Published February 2001. Accessed April 11, 2020.
6. Li H, Xiong Y, Hunger CV, Guo X, Tywoniw R. Does peer assessment promote student learning? A meta-analysis. Assess Eval High Educ. 2020;45(2):193-211.
7. Topping KJ. Peer assessment between students in colleges and universities. Rev Educ Res. 1998;68(3):249-276.
8. Balfour SP. Assessing writing in MOOCs: automated essay scoring and calibrated peer review. Res Pract Assess. 2013;8(1):40-48.
9. McCarty T, Parkes MV, Anderson TT, Mines J, Skipper BJ, Grebosky J. Improved patient notes from medical students during web-based teaching using faculty-calibrated peer review and self-assessment. Acad Med. 2005;80(10):S67-70.
10. Peleaz NJ. Problem-based writing with peer review improves academic performance in physiology. Adv Physiol Educ. 2002;26(1):174-184.
11. Walvoord ME, Hoefnagels MH, Gaffin DD, Chumchal MM, Long DA. An analysis of calibrated peer review (CPR) in a science lecture classroom. J Coll Sci Teach. 2008;37(4):66-73.
12. Gunersel AB, Simpson N. Improvement in writing and reviewing skills with calibrated peer review. Int J Schol Teach Learn. 2009;3(2):Article 15.
13. Gunersel AB, Simpson NJ, Auferheide KJ, Wang L. Effectiveness of calibrated peer review for improving writing and critical thinking skills in biology undergraduate students. J Schol Teach Learn. 2008;8(2):25-37.
14. Mynlieff M, Manogaran AL, St. Maurice M, Eddinger TJ. Writing assignments with a metacognitive component enhance learning in a large introductory biology course. Biochem Mol Biol Educ. 2010;38(5):290-295.
15. Likkel L. Calibrated peer review essays increase student confidence in assessing their own writing. J Coll Sci Teach. 2012;41(3):42-47.
16. Clase KL, Gundlach E, Peleaz NJ. Calibrated peer review for computer-assisted learning of biological research competencies. Biochem Mol Biol Educ. 2017;80(10):S67-70.
17. Joint Commission of Pharmacy Practitioners. Pharmacists’ Patient Care Process. https://jcpp.net/wp-content/uploads/2016/03/PatientCareProcess-with-supportingorganizations.pdf. Published May 2014. Accessed April 11, 2020.
18. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357(9260):945-949.
19. Hift R. Should essays and other “open-ended”-type questions retain a place in written summative assessment in clinical medicine? BMC Med Educ. 2014;14(1):249.
20. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Upper Saddle River, NJ: Prentice Hall Health; 2000.
21. Overholser BR, Sowinski KM. Biostatistics primer: part 2. Nutr Clin Pract. 2008;23(1):76-84.
22. Blommel ML, Abate MA. A rubric to assess critical literature evaluation skills. Am J Pharm Educ. 2007;71(4):Article 63.
23. Walton AM, Isaacs AN, McFarland AT, Czoskowski LM, Nisly SA. Design of a drug-induced diseases elective utilizing active learning. Curr Pharm Teach Learn. 2016;8(1):90-95.
24. Stevens DD, Levi AJ. Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning. 2nd ed. Sterling, VA: Stylus Publishing; 2013.
25. Davis LE. A workshop series using peer-grading to build drug information, writing, critical thinking, and constructive feedback skills. Am J Pharm Educ. 2014;78(10):Article 191.