

Student peer assessment in evidence-based medicine (EBM) searching skills training: an experiment

Jonathan D. Eldredge, PhD, AHIP; David G. Bear, PhD; Sharon J. Wayne, MPH; Paul P. Perea, MA

See end of article for authors' affiliations.

DOI: <http://dx.doi.org/10.3163/1536-5050.101.4.003>

Background: Student peer assessment (SPA) has been used intermittently in medical education for more than four decades, particularly in connection with skills training. SPA generally has not been rigorously tested, so medical educators have limited evidence about SPA effectiveness.

Methods: Experimental design: Seventy-one first-year medical students were stratified by previous test scores into problem-based learning tutorial groups, and then these assigned groups were randomized further into intervention and control groups. All students received evidence-based medicine (EBM) training. Only the intervention group members received SPA training, practice with assessment rubrics, and then application of anonymous SPA to assignments submitted by other members of the intervention group.

Results: Students in the intervention group had higher mean scores on the formative test with a

potential maximum score of 49 points than did students in the control group, 45.7 and 43.5, respectively ($P=0.06$).

Conclusions: SPA training and the application of these skills by the intervention group resulted in higher scores on formative tests compared to those in the control group, a difference approaching statistical significance. The extra effort expended by librarians, other personnel, and medical students must be factored into the decision to use SPA in any specific educational context.

Implications: SPA has not been rigorously tested, particularly in medical education. Future, similarly rigorous studies could further validate use of SPA so that librarians can optimally make use of limited contact time for information skills training in medical school curricula.

STUDENT PEER ASSESSMENT (SPA)

Librarians have provided information skills training for medical students for at least 8 decades. Postell reported on survey results from the 1930s indicating that 78% of the medical schools in the United States included library instruction in their curricula [1]. Six decades later, Earl reported that 75% of the respondents in a nationwide survey similarly offered library instruction in medical schools [2]. The appearance of the 1998 Association of American Medical Colleges informatics competencies [3] and the rising popularity of teaching evidence-based medicine (EBM) in medical school curricula have increased the need for teaching information skills to medical students. Yet, librarians currently have to compete for time in medical school curricula, with numerous other teaching faculty members also needing to teach important competencies. This problem, revolving around what Smith and others refer to as the "crowded curriculum dilemma" [4], has led librarians to search for pedagogic techniques that will maximize the effectiveness of their limited face-to-face time with students [5-7].

SPA pedagogy

Student peer assessment (SPA) involves students evaluating the work of other students, oftentimes in highly structured environments. SPA often measures unambiguous observable behavior or completion of course assignments. SPA can be used for students to rate their fellow students on their performance of skills, oral presentations, or written compositions, to cite a few examples [8]. Students' evaluation of their peers' performance normally represents a formative rather than a summative assessment activity for these peers [9]. Formative assessment focuses primarily on providing feedback intended to enhance student learning. Formative assessment relies mainly on low-stakes quizzes or tests, representing only a small percentage of the overall course grade [10]. Could SPA possibly help solve the "crowded curriculum dilemma" during an era when medical educators are condensing more and more content into limited face-to-face classroom or lab contact time? SPA has been hypothesized to produce more comprehensive and internalized learning, allowing faculty and students to more effectively use limited class contact time. SPA thereby may enable students to retain and apply learned skills to new situations [11, 12]. Faculty members representing a vast variety of subject disciplines in higher education have been reported to use SPA techniques. These SPA techniques have been applied to assessment of many types of educational achievements, such as testing knowledge, skills, or expected behavior [13].



A supplemental appendix is available with the online version of this journal.

Falchikov notes that “Peer assessment requires students to provide either feedback or grades (or both) to their peers on a product, process, or performance, based on the criteria of excellence for that product or event” [14]. Topping defines SPA as “an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners” [15]. Arnold and Stern describe “peers” as “individuals who have attained the same level of training or expertise, exercise no formal authority over each other, and share the same hierarchical status in an institution” [16].

SPA often employs rubrics to ensure consistent and transparent feedback between students. Rubrics isolate discrete components of educational activities for ease of assessment. Rubrics define and describe these components and link them to a range of student performance levels on the discrete tasks. The four elements in rubrics consist of: (1) a task description; (2) scales of potential student achievement levels (examples: does not meet expectations, meets expectations, exceeds expectations, or simply “yes” or “no” responses); (3) component parts of the activities; and (4) descriptions of achievement levels that are reached through incremental stages [17].

Medical education tends to utilize SPA primarily when assessing professional skill competencies, particularly in the area of assessing professional behavior [18]. Norcini provides a synthesis of current practices of SPA in medical education. He notes that student participants need to be informed about the purposes of SPA, participate anonymously, be “graded” at low stakes when first introduced to SPA, and be trained in the criteria for defining excellent performance in those to be assessed [19]. A survey completed by students at four US medical schools on the delicate subject of students assessing one another’s professional behavior using SPA by Arnold et al. validates Norcini’s points about anonymity and clearly defined standardized criteria. That study also notes that SPA feedback must be delivered to the assessed students in a timely manner [20]. Modern experiences with SPA in medical education date back to at least Kubany’s 1957 sociometric study [21] and the subsequent, though intermittent, work of other SPA pioneers [22–25]. Yet, the uneven focus in the research literature on SPA in medical education still does not seem to match the potential of SPA to facilitate deep learning for medical students [26].

The underlying pedagogical principles supporting SPA are well grounded in mainstream theoretical constructs and conceptual frameworks in the field of education. SPA consistently adheres to experiential education theories as explored by Dewey [27–29] and Bruner [30], and later elaborated upon and codified by Kolb’s theory of learning cycles [31]. Kolb conceptualized experiential learning as consisting of concrete experience, reflective observations about the experience, abstraction of concepts from the experience, and active experimentation with knowledge gained from the experience to foster in-depth rather

than superficial learning. SPA clearly incorporates these four elements. SPA also has a theoretical basis in learning processes and curriculum design, as examined by Gagné [32–36]. SPA demonstrates early constructivist theory [37] and later aligned with constructivist theorists, such as Biggs and Tang, who seek to include learners in designing, implementing, and assessing learning [38]. As already noted, deep-learning theorists such as Marton and Säljö [39] likely anticipated current SPA practices. Finally, SPA additionally can rely on the social learning theories of Vygotsky [40–42] when employed within certain non-anonymous, face-to-face assessment contexts.

SPA evidence

Several systematic literature reviews [13, 43, 44] and a meta-analysis [45] have synthesized the available evidence on the effectiveness of SPA practices. Speyer et al. have conducted a systematic review specifically related to medical education, although it focuses narrowly on psychometric instruments with an investigation of their validity and reliability [18]. The existing evidence generally points to the effectiveness of SPA for enhancing deep learning.

There are very few experimental or quasi-experimental studies [46] involving SPA approaches, particularly in the area of student performance on subsequent tests. Instead, the highest forms of evidence to support SPA tend to be cohort (pretest, then posttest) studies. The majority of research studies also have concentrated upon either the agreement between student peer assessors and faculty assessments, or on student perceptions of SPA rather than actual performance outcomes [8].

Empirical research tends to form a consensus around certain practices such as using SPA anonymously and primarily for formative, low-stakes assessment events [26, 47–52]. Any system for enacting SPA needs to ensure fairness and guarantee to all students that those students in the peer reviewing role will neither under-score nor over-score their fellow students [53]. Perhaps counter intuitively, the evidence supports the common practice of using only one student peer rather than multiple student peers to evaluate the work of each individual student. Some authors have synthesized the best available evidence on SPA and applied this evidence to large class environments. The authors of the study reported here developed the protocols based on the aforementioned evidence to assure the integrity of the SPA process through use of explicit, unambiguous, and finite criteria for SPA rubrics [54]. With respect to these last points—concerning the need for explicit, unambiguous, and finite criteria—it should be noted that students also resist SPA when it entails a forced distribution of grades [55, 56].

RESEARCH QUESTION AND HYPOTHESIS

The present experimental study applied SPA to medical students learning evidence-based medicine

(EBM) searching skills using PubMed. Past efforts by the first author (Eldredge) to gauge the effectiveness of SPA—with smaller groups of five to twenty graduate students in clinical research, health policy, physician assistant, and public health curricula—while students learned evidence-based practice skills had suggested promisingly that SPA might enhance student learning. A medical school curriculum must cover many competencies within limited in-class time, so any opportunity to teach these EBM searching skills in such a way that these skills are retained and applied effectively throughout the remainder of the crowded curriculum seemed to be a prudent use of the extra effort required to implement an SPA process.

Norcini [19], Falchikov [8], and others have noted that SPA excels in teaching students new skills. This study focused on EBM informatics searching skills, since this skill set had been taught for a number of years in the medical school curriculum [57, 58]. The authors could thus use past experience to anticipate possible obstacles to implementation. If applying SPA to EBM literature searching yielded the expected pedagogical benefits, the authors reasoned that SPA might then prove similarly helpful elsewhere in their own and in other medical school curricula. The working hypothesis for this study proposed that medical students in an intervention group using SPA techniques would master a specific set of EBM PubMed searching skills better than students in the control group who had otherwise received identical training in EBM PubMed searching skills.

METHODS

Procedures

All first-year medical students who were beginning their second organ system block on genetics and neoplasia at the University of New Mexico School of Medicine were enrolled in this study. This study was approved by the University of New Mexico Institutional Review Board (HRRC 09-423). The chair of the block (Bear) allocated all students into problem-based learning tutorial groups using their average score on two tests in the previous block. In this way, students were distributed evenly in terms of previous test performance across eleven tutorial groups. The authors used a random number generator to assign each of the eleven tutorial groups into either an intervention or a control group; in other words, randomization was applied at the tutorial group level rather than at the individual level. Scheduling and room size constraints meant that more tutorial groups (seven) were allocated to the intervention group than the control group (four).

All students attended a one-hour introductory lecture to EBM presented by the librarian coauthor (Eldredge) on a Monday, during which they were given an overview of the EBM PubMed searching skills that they would learn during an EBM lab as well as a description of three closely related rubrics that would be used for grading their mastery of EBM

searching skills. Later that day, students participated in training labs in EBM searching skills. All students in every lab received instruction in how to conduct two basic types of PubMed searches: (1) a Medical Subject Headings (MeSH) term combined with one subheading, and (2) combination of two or more MeSH terms using a Boolean operator.

After the instructional portion of the lab, all students were given an assignment to complete and turn in to a proctor (Perea) before completion of the lab. The assignment required all students to conduct a diagnosis, treatment, and prognosis EBM search in PubMed on one of twenty-four listed subjects, such as alcoholism, breast cancer, colon cancer, cystic fibrosis, leukemia, or melanoma. The librarian instructor directed all students to copy, paste, and submit to the proctor only their best search strategies in each three of the categories as documented by PubMed histories. All students were given a physical copy of, and an explanation of, the assessment rubric (Appendix, online only) to be used for both the assignment and the subsequent test on the Friday at the end of that same week. Within forty-eight hours, the instructor also graded and returned the initial assignments to all students using the three rubrics.

Only during the two labs comprising students in the intervention group did the librarian instructor explain the mechanics of the SPA process. The librarian instructor reviewed the rubric criteria and modeled the way that students should assess their assigned fellow students over the next forty-eight hours. He stated explicitly that he would monitor their assessments of their fellow students for accuracy and completeness. Students in the control group did not receive these specific instructions. Following the labs, only the intervention group students received an assignment completed by another student in the intervention group whose identity was concealed (“blinded”). They were reminded to use the three rubrics to assess the other student’s assignment. Neither assessing nor assessed students were ever told the identity of their anonymous student peer partners. Over the next forty-eight hours, students in the intervention group anonymously assessed their assigned peers’ searches. These peer assessments followed the series of steps and decisions that students had to make in the EBM PubMed searching techniques learned during the lecture and lab. As assessors, student were prompted to evaluate search maneuvers such as accessing the MeSH database, attaching subheadings, combining two or more MeSH terms (i.e., Boolean logic), and applying filters (i.e., limits) to search results as detailed in the rubrics in the online only appendix.

Thus, students in the intervention group received and reviewed the peer assessment of their searches. Students in the control group received no similar feedback. The instructor graded *all* students’ assignments and provided feedback seventy-two hours after the labs using the same three rubrics with all students’ identities concealed. The peer rating [43] rubrics used in the present study evolved out of the grading

process used over the years in previous teaching experiences that employed SPA in other degree program curricula.

Measurement

At the end of the same week, all students took a low-stakes, formative skills test worth 5% of their course grade. The test gauged their EBM PubMed searching competency by asking them to search PubMed for relevant references on one of the following 3 clinical vignettes. Assignments were made according to the first letter of their first names:

A–E: You are a family practice physician examining a series of patients with what appears to be bacterial pneumonia. Search PubMed using all of the skills and guidelines given in the lecture and lab: diagnosis, treatment, and prognosis.

F–Q: You are a pediatrician specializing in adolescents who are at high-risk for contracting HIV infections. Search PubMed using all of the skills and guidelines given in the lecture and lab: diagnosis, treatment, and prognosis.

R–Z: You are an internist specializing in a geriatric population in an 80–100 year-old age cohort. A significant number of your patients in this population appear to suffer from depression. Search PubMed using all of the skills and guidelines given in the lecture and lab: diagnosis, treatment, and prognosis.

The librarian instructor scored all student formative tests with students' identities concealed, with the help of the proctor, using the three grading rubrics previously distributed to all students. The team statistician (Wayne) compared mean scores on the exam between the intervention and the control groups. The test results also were categorized (correctly answering 90% or more of the questions versus correctly answering less than 90% of the questions) and compared by group. The differences between the means were tested using a Wilcoxon rank sum test due to the non-normal distribution of the scores. The percentages were tested using a chi-square test. The analysis for this article was generated using SAS/Stat software (version 9.2, Cary NC).

RESULTS

During the fall, 80 students began the genetics and neoplasia block. EBM knowledge and skills training constituted a component of this block. Six students were absent during the EBM labs, and 3 did not take the formative test, leaving a total study size of 71. Of these, 47 (66%) students were in the intervention group and 24 (34%) students were in the control group. Table 1 indicates that there were no statistical differences between the groups for the demographic variables of gender, minority status, or age. The formative skills test scores ranged from a low of 0 to a high of 49.

Table 2 presents the mean scores by group. The mean score for the intervention group was higher

Table 1
Demographic profile and test score results by group

Demographics	Intervention group (n=47)		Control group (n=24)		P-value*
	(n)	(%)	(n)	(%)	
Gender					
Female	29	(62%)	13	(54%)	0.54
Male	18	(38%)	11	(46%)	
Minority†					
Yes	20	(43%)	11	(55%)	0.39
No	26	(57%)	9	(45%)	
Age at matriculation					
≤24	24	(51%)	17	(71%)	0.11
25+	23	(49%)	7	(29%)	

* P-value for percentages is from chi-square test.

† Minority status is not known for 1 student in the intervention group and 4 students in the control group.

than the control group (45.7 and 43.5, respectively, a difference of marginal statistical significance $P=0.06$). The percentage of students receiving scores of 45 (out of a possible 49 points) or higher was almost identical for the 2 groups (30% and 29%).

Pretest results were not available for these same students. Yet, the authors had observed over the preceding 8 years that prior to their taking this training, almost all medical students lacked knowledge of either EBM or the specific subset of EBM skills used in searching PubMed for the evidence that pertained to this study when they began the block. One author tested and confirmed this assumption by administering a pretest to a different cohort of first-year medical students. These students performed poorly on the pretest when graded with rubrics nearly identical to the rubrics used in the present study. That group of students earned an average score of 1.28 out of a possible 22 points, which would have been a score of only 6% on a grading scale (data not shown).

DISCUSSION

The researchers tested the hypothesis that SPA techniques would improve mastery of EBM PubMed searching skills, using three closely related scoring

Table 2
Results of formative exam

	Intervention group (n=47)		Control group (n=24)		P-value*
	Mean (SD)	(n) (%)	Mean (SD)	(n) (%)	
Test score	45.7 (4.0)		43.5 (9.5)		0.06
Score category					
Top 10% (45+)		14 (30%)		7 (29%)	0.96
Bottom 90% (≤44)		33 (70%)		17 (71%)	

* P-value comparing means is from Wilcoxon rank sum test.

rubrics to teach these EBM searching skills and to quantifiably assess student learning on a formative skills test. Both the control group and the intervention group received the same lecture-based instruction, the same post-lecture lab, and copies of the same three assessment rubrics, and then within seventy-two hours received timely post-lab session assessment feedback from the instructor on the required assignment. The only difference between the groups was that the intervention group received instruction in SPA techniques and applied this instruction to assess their peers' work anonymously over a forty-eight-hour period. This study was designed to implement Beckman and Cook's recommendation to use a true experiment in order to determine which of two or more courses of educational methods succeeds better in reaching a desired outcome [59].

The authors reasoned that students in both the intervention group and the control group would be motivated to utilize the EBM searching skills they learned in the lab to later address the complex learning issues introduced in their problem-based learning tutorials during phase I of their curriculum. The specific EBM techniques for searching PubMed for evidence, the second step in the EBM process, could be directly applied to acquire answers to questions related to the subjects covered during the genetics and neoplasia block. First-year medical students do not have occasion to apply most EBM skills and knowledge while still in the preclinical phase of the curriculum. Yet, students could apply the specific EBM PubMed searching skills taught in the lab immediately in their problem-based learning, so these skills were expected to have high relevance for all students.

This study adds to the limited experimental evidence supporting the effectiveness of SPA, as measured by a formative test following EBM searching skills instruction. Results from this study suggest that students who learn about, and participate in, a SPA process perform marginally better on subsequent formative skills assessments than do similar students not using the SPA process. Although the difference was not statistically significant, students in the intervention group had higher mean scores than did students in the control group when tested on their EBM searching skills. The medical school's assessment unit requires students to evaluate every block and course they have taken in the curriculum using a 5-point Likert-scale instrument, with "5" as the highest score. Students in this specific cohort year continued to rate the EBM skills component of the block highly, with 4.0 scores on the 5-point Likert scale consistent with past years, suggesting that neither control nor intervention group students perceived themselves as adversely affected by this experiment.

The limited evidence from research in higher education suggests that SPA stimulates deep learning and improves student performance outcomes. This study lends limited support to this body of evidence. Could it be that because medical students are

dramatically more motivated to succeed academically, they will do better than most other higher education students on tests, regardless of the presence of SPA? In this context, SPA might contribute to marginally better student performances but not markedly so.

The librarian instructor (Eldredge) has taught graduate students in clinical research, dental, health policy, nursing, pharmacy, physician assistant, and public health graduate programs for a number of years. He tends to agree with the general observation that medical students are more highly motivated to succeed than students in these other degree programs. In the case of highly motivated medical students, perhaps they markedly improve their test scores instead when notified in advance about and guided by well-constructed rubrics. One can only speculate about this possibility at this juncture, although it does offer a tantalizing testable hypothesis for the future.

Practical considerations

The literature nearly universally reports that implementing SPA requires more of instructors' time and incrementally more of student time. Two of the authors (Eldredge and Perea) determined from their logs and calendars that concealment of the identities of students from one another and from the faculty member required about six more hours for the instructor and approximately ten hours for the proctor who was charged with devising a system to conceal identities to ensure the integrity of this study. These two authors estimate that by automating future SPA arrangements, they might reduce this administrative effort to only four hours' effort combined. The current educational assessment technology does not yet support an easy way to automate the specific SPA applications described in this article [60]. Each student in the intervention group probably involved an estimated twenty to thirty minutes more in anonymous peer review outside of their in-class time to complete their SPA assignments than their control group counterparts, based on anecdotal student accounts. While SPA outside of class contact time might lead to strengthening essential EBM skills in a crowded first-year medical school curriculum, will many instructors be willing to invest the extra time to make it happen out of their own already overbooked professional work lives? Ultimately, in the authors' view, the instructor will need to make that decision. Falchikov also has noted many of these practical considerations in her book on SPA [8].

Limitations

The design and context for this study present certain limitations. These first-year medical students, who were still within their first four months of school, had no prior experience with SPA within the medical school curriculum. A number of researchers who have studied SPA recommend that students be trained on SPA prior to practicing it [13], whereas this study

accelerated training and application because of the crowded curriculum.

Second, practical considerations revolving around medical students' frequent interactions across the cohort of their class in medical school prevented extending this experiment over a time period longer than a week to measure sustained learning of searching skills. Students in the intervention group during a experiment of longer duration could have possibly taught their hypothetically superior EBM PubMed searching skills to students in the control group in subsequent weeks, thereby contaminating individual members of the control group for the study. A randomized controlled trial in medical education at Oxford University acknowledged the possibility of such a contamination effect [61]. Thus, in this type of course-long design, it was not possible to truly distinguish between groups. Yet, a prospective cohort study could probably determine if all students using SPA retain high levels of competence beyond a single week of a block compared to a similar group of students lacking training in and application of SPA.

Third, no test data are available to evaluate how test scores might fluctuate over time for other students participating in a conventional learning and assessment arrangement.

Fourth, this study could not determine if the observed differences resulted from the fact that students in the intervention group simply spent more time using the skills set or using the rubrics. SPA takes longer not only to learn, but also to practice. The argument might be made that more time invested in any meaningful activity related to learning, particularly when coupled with timely and consistent rubric-based feedback, will result in better scores. The SPA methodology furthermore inherently involves students spending incrementally more time interacting with the skills set. In this version of SPA, however, most of the activity occurred outside class or lab contact hours, a key advantage of SPA in a crowded curriculum.

Fifth, the presence of one of the authors (Perea) who was known to all students as a proctor for major medical school exams might have caused some degree of inadvertent anxiety for all students. This anxiety might have prompted students to take the formative test more seriously as a mild form of the Hawthorne effect [62]. If all students were anxious, this might have motivated them all to perform better than they would have in a more typical, formative SPA setting. This, in turn might have elevated the grades for both groups.

Finally, this study only included first-year medical students in one block at a single medical school, and so any generalization will remain limited until others elsewhere replicate this study.

CONCLUSION

This study suggests that students in the intervention group using SPA performed better than students in

the control group, although not significantly so, on a formative skills test a few days later. Faculty or staff time invested on implementing a SPA program might be minimized through yet-to-be-developed new information technology. Further experimental studies need to either confirm or dispute the findings in this study, particularly in light of the limited amount of rigorous evidence of the effectiveness of SPA. Cohort studies in the future also could evaluate the effect of SPA on long-term retention of EBM PubMed searching skills. In the meantime, this experiment offers suggestive evidence that SPA can be applied to teaching skills that will require predicted, repeated application throughout the medical school curriculum, such as the EBM skills evaluated here.

ACKNOWLEDGMENTS

The authors thank mentors and facilitators Craig Timm, MD, and Summers Kalishman, PhD, in the University of New Mexico School of Medicine's Medical Education Scholars Program for their continuous enthusiasm for and ongoing support of this project. The first author thanks Holly S. Buchanan, AHIP, FMLA, and Deb LaPointe for their support of this faculty training program and the research project itself. The authors thank Leslie Sandoval from the Learning Design Center for assisting in labs and certifying that both intervention and control groups received the exact same training, except for the intended interventions. We also thank Nancy Sinclair and Teresita McCarty for their personnel support. Finally, the authors thank their faculty colleagues in the School of Medicine's Medical Education Scholars for their three peer-review sessions of this study during the proposal stage.

REFERENCES

1. Postell WD. Further notes on the instruction of medical school students in medical bibliography. *Bull Med Lib Assoc.* 1944 Apr;32(2):217-20.
2. Earl MF. Library instruction in the medical school curriculum: a survey of medical college libraries. *Bull Med Lib Assoc.* 1996 Apr;84(2):191-5.
3. Association of American Medical Colleges. Medical School Objectives Project: contemporary issues in medical education: medical informatics and population health [Internet]. Washington, DC: The Association; 1998 [cited 10 Mar 2013]. <https://members.aamc.org/eweb/DynamicPage.aspx?Action=Add&ObjectKeyFrom=1A83491A-9853-4C87-86A4-F7D95601C2E2&WebCode=PubDetailAdd&DoNotSave=yes&ParentObject=CentralizedOrderEntry&ParentDataObject=Invoice_Detail&ivd_formkey=69202792-63d7-4ba2-bf4e-a0da41270555&ivd_prc_prd_key=4F099E49-F328-4EEC-BB7B-AE073EA04F6B>.
4. Smith HS. A course director's perspectives on problem-based learning curricula in biochemistry. *Acad Med.* 2002 Dec;77(12):1189-98.
5. Just ML. Is literature search training for medical students and residents effective? a literature review. *J Med Lib Assoc.* 2012 Oct;100(4):270-6. DOI: <http://dx.doi.org/10.3163/1536-5050.100.4.008>.

6. Dorsch JL, Perry GJ. Evidence-based medicine at the intersection of research interests between academic health sciences librarians and medical educators: a review of the literature. *J Med Lib Assoc.* 2012 Oct;100(4):251–7. DOI: <http://dx.doi.org/10.3163/1536-5050.100.4.006>.
7. Gagliardi JP, Stinnett SS, Schardt C. Innovation in evidence-based medicine education and assessment: an interactive class for third- and fourth-year medical students. *J Med Lib Assoc.* 2012 Oct;100(4):306–9. DOI: <http://dx.doi.org/10.3163/1536-5050.100.4.014>.
8. Falchikov N. Improving assessment through student involvement: practical solutions for aiding learning in higher and further education. New York, NY: Routledge Falmer; 2005. p. 83–116.
9. Patton C. ‘Some kind of weird, evil experiment’: student perceptions of peer assessment. *Assess Eval High Educ.* 2012 Sep;37(6):19–31.
10. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud Higher Educ.* 2006 Apr;31(2):199–218.
11. Basheti IA, Ryan G, Woulfe J, Bartimote-Aufflick K. Anonymous peer assessment of medication management reviews. *Am J Pharm Educ.* 2010;74(5):1–8.
12. Liu LLX, Steckelberg AL. Assessor or assessee: how student learning improves by giving and receiving peer feedback. *Br J Educ Tech.* 2010;41(3):525–36.
13. Topping K. Peer assessment between students in colleges and universities. *Rev Educ Res.* 1998 Fall;68(3):249–76. (p. 250).
14. Falchikov N. The place of peers in learning and assessment. In: Boud D, Falchikov N, eds. *Rethinking assessment in higher education: learning for the longer term.* New York, NY: Routledge; 2007. p. 128–43. (p. 132).
15. Topping KJ. Peer assessment. *Theory Pract.* 2009 Winter;48(1):20–7.
16. Arnold L, Stern D. Content and context of peer assessment. In: Stern DT, ed. *Measuring medical professionalism.* New York, NY: Oxford University Press; 2006. p. 175–94. (p. 175).
17. Stevens DD, Levi AJ. Introduction to rubrics: an assessment tool to save grading time. Sterling, VA: Stylus; 2012.
18. Speyer R, Pilz W, van der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. *Med Teach.* 2011;33:e572–85.
19. Norcini JJ. Peer assessment of competence. *Med Educ.* 2003 Jun;37(6):539–43.
20. Arnold L, Shue CK, Kalishman S, Prislun M, Pohl C, Pohl H, Stern DT. Can there be a single system for peer assessment of professionalism among medical students? a multi-institutional study. *Acad Med.* 2007 Jun;82(6):578–86.
21. Kubany AJ. Use of sociometric peer nominations in medical education research. *J Appl Psychol.* 1957 Dec, 41(6):389–94.
22. Korman M, Stubblefield RL. Medical school evaluation and internship performance. *J Med Educ.* 1971 Aug;46(8):670–3.
23. Linn BS, Arostegui M, Zeppa R. Performance rating scale for peer and self assessment. *Br J Med Educ.* 1975 Jun;9(2):98–101.
24. Arnold L, Willoughby L, Calkins V, Gammon L, Eberhart G. Use of peer evaluation in the assessment of medical students. *J Med Educ.* 1981 Jan;56(1):35–42.
25. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA.* 1993 Apr 7;269(13):1655–60.
26. Finn GM, Garner J. Twelve tips for implementing a successful peer assessment. *Med Teach.* 2011 Jun;33(6):443–6.
27. Dewey J. My pedagogic creed. *School J [Internet].* 1897 Jan 16;54(3):77–80. [cited 10 Mar 2013]. <<http://www.dewey.pragmatism.org/creed.htm>>.
28. Dewey J. *Experience and education.* New York, NY: Collier Books; 1938.
29. Soltis JF. John Dewey, 1859–1952. In: Guthrie JW, ed. *Encyclopedia of education.* 2nd ed. Vol. 4. New York, NY: Macmillan Company; 2003. p. 577–82.
30. Bruner JS. *The process of education.* Cambridge, MA: Harvard University Press; 1962.
31. Kolb DA. *Experiential learning: experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall; 1984.
32. Gagné RM, Scriven M. Curriculum research and the promotion of learning. In: Tyler RW, ed. *Perspectives of curriculum evaluation.* Chicago, IL: Rand McNally; 1967. p. 19–38.
33. Gagné RM. Learning: transfer. In: Sills DL, ed. *International encyclopedia of the social sciences.* Vol. 9. New York, NY: Macmillan Company; 1968. p. 168–73.
34. Gagné RM. *Conditions of learning and theory of instruction.* 4th ed. New York, NY: Holt, Rinehart and Winston; 1985.
35. Gagné RM, Briggs LJ, Wager WW. *Principles of instructional design.* Belmont, CA: Wadsworth/Thomson Learning; 1992.
36. Strauss S. Learning theories of Gagne and Piaget: implications for curriculum development. *Teachers Coll Record.* 1972 Sep;74(1):81–102.
37. Lamon M. Constructivist approach. In: Guthrie JW, ed. *Encyclopedia of education.* Vol. 4. New York, NY: MacMillan Reference; 2003. p. 1463–7.
38. Biggs J, Tang C. *Teaching for quality learning at university; what the student does.* 3rd ed. Buckingham, UK: Open University Press; 2007.
39. Marton F, Säljö R. On quantitative differences in learning: I – outcome and process. *Br J Educ Psychol.* 1976 Feb;46(1):4–11.
40. Vygotsky LS. *Thought and language.* Hanfmann E, Vakar G, eds. and translators. Cambridge, MA: MIT Press; 1962.
41. Burns MS, Bodrova E, Leong DJ. Vygotskian theory. In: Guthrie JW. *Encyclopedia of education.* 2nd ed. Vol. 4. New York, NY: MacMillan Reference USA; 2003. p. 574–7.
42. Gredler ME. Lev Vygotsky 1896–1934. In: Guthrie JW. *Encyclopedia of education.* 2nd ed. Vol. 7. New York, NY: MacMillan Reference USA; 2003. p. 2658–60.
43. Douchy F, Segers M, Sluijsmans D. The use of self-, peer, and co-assessment in higher education: a review. *Stud High Educ.* 1999 Oct;24(3):331–50.
44. Van Zundert M, Sluijsmans D, van Merriënboer J. Effective peer assessment processes: research findings and future directions. *Learn Instr.* 2010 Aug;20(4):270–9.
45. Falchikov N, Goldfinch J. Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Rev Educ Res.* 2000 Fall;70(3):287–322.
46. Strijbos JW, Sluijsman D. Unravelling peer assessment: methodological, functional, and conceptual developments. *Learn Instr.* 2010 Aug;20(4):265–9.
47. Trautmann NM. Interactive learning through web-mediated peer review of student science reports. *Educ Technol Res Dev.* 2009 Oct;57(5):685–704.
48. Dannefer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, Barclay C, Epstein RM. Peer

- assessment of professional competence. *Med Educ.* 2005 Jul;39(7):13–22.
49. Wen ML, Tsai CC. University students' perceptions and attitudes toward (online) peer assessment. *High Educ.* 2006 Jan;51(1):27–44.
50. Lu R, Bol L. A comparison of anonymous versus identifiable e-peer review of college student writing performance and the extent of critical feedback. *J Interact Online Learn.* 2007 Summer;6(2):100–15.
51. Pelaez NJ. Problem-based writing with peer review improves academic performance in physiology. *Adv Physiol Educ.* 2002 Sep;26(3):174–84.
52. Bangert AW. An exploratory study of the effects of peer assessment activities on student motivational variables that impact learning. *J Stud Cent Learn.* 2003;1(2):69–76.
53. Hughes C, Toohey S, Velan G. eMed teamwork: a self-moderating system to gather peer feedback for developing and assessing teamwork skills. *Med Teach.* 2008 Feb;30(1):5–9.
54. Ballantyne R, Hughes K, Mylonas A. Developing procedures for implementing peer assessment in large classes using an action research process. *Assess Eval High Educ.* 2002;27(5):427–41.
55. Ryan GJ, Marshall LL, Porter K, Jia H. Peer, professor and self-evaluation of class participation. *Active Learn High Educ.* 2007;8(1):49–61.
56. Liu NF, Carless D. Peer feedback: the learning element of peer assessment. *Teach High Educ.* 2006 Jul, 11(3):279–90.
57. Eldredge JD. EBM informatics component of the genetics & neoplasia block. In: Sewell RR, Brown JF, Hannigan GG, eds. *Informatics in health sciences curricula [MLA DocKit]*. Rev. ed. Chicago, IL: Medical Library Association; 2005.
58. Eldredge JD. Introducing evidence-based medicine and library skills. In: Brown JF, Hannigan GG, comps. *Informatics in health sciences curricula [MLA DocKit]*. Chicago, IL: Medical Library Association; 1999. p. 313–27.
59. Beckman TJ, Cook DA. Developing scholarly projects in education: a primer for medical teachers. *Med Teach.* 2007 Mar;29(2):210–8.
60. Luxton-Reilly A. A systematic review of tools that support peer assessment. *Comp Sci Educ.* 2009 Dec;19(4):209–32.
61. Rosenberg WMC, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *J R Coll Physicians Lond.* 1998 Nov–Dec;32(6):557–63.
62. Roethlisberger FJ, Dickson WJ. *Management and the worker: an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago.* Cambridge, MA: Harvard University Press; 1939. p. 194–9, 227.

AUTHORS' AFFILIATIONS

Jonathan D. Eldredge, PhD, AHIP, jeldredge@salud.unm.edu, Associate Professor, School of Medicine, Health Sciences Library and Informatics Center, MSC09 5100; **David G. Bear, PhD**, dbear@salud.unm.edu, Professor of Cell Biology and Physiology, and Biochemistry and Molecular Biology, and former Dean of Admissions, MSC08 4750; **Sharon J. Wayne, MPH**, Swayne@salud.unm.edu, Health Sciences Associate Scientist (project statistician), School of Medicine Program Evaluation Program, MSC08 4550; **Paul P. Perea, MA**, ppperea@salud.unm.edu, Research and Information Manager, School of Medicine, Educational Tech Support, MSC08 4710; University of New Mexico, Albuquerque, NM 87131-0001

Received January 2013; accepted May 2013