

2005

# Using Calibrated Peer Review<sup>â™™</sup> to Assess and Improve the Quality of Student Documentation of Clinical Encounters at the University of New Mexico School of Medicine (UNMSOM).

Teresita McCarty

Marie Parkes

Kim Persinger

Craig Timm

M Martinez

*See next page for additional authors*

Follow this and additional works at: [https://digitalrepository.unm.edu/ume\\_edu\\_pubs](https://digitalrepository.unm.edu/ume_edu_pubs)

---

## Recommended Citation

McCarty, Teresita; Marie Parkes; Kim Persinger; Craig Timm; M Martinez; Toby Palley; Greg Franchini; Sheila Hickey; Tom Howdieshell; Eve Espey; Summers Kalishman; and Jan Mines. "Using Calibrated Peer Review<sup>â™™</sup> to Assess and Improve the Quality of Student Documentation of Clinical Encounters at the University of New Mexico School of Medicine (UNMSOM)." (2005). [https://digitalrepository.unm.edu/ume\\_edu\\_pubs/13](https://digitalrepository.unm.edu/ume_edu_pubs/13)

This Presentation is brought to you for free and open access by the Undergraduate Medical Education at UNM Digital Repository. It has been accepted for inclusion in UME Faculty and Staff Educational Scholarship by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

---

**Authors**

Teresita McCarty, Marie Parkes, Kim Persinger, Craig Timm, M Martinez, Toby Palley, Greg Franchini, Sheila Hickey, Tom Howdieshell, Eve Espey, Summers Kalishman, and Jan Mines

Moderator: Michael Minzter, MD  
 Discussant: Jim Gordon, MD

## Improved Patient Notes from Medical Students during Web-Based Teaching Using Faculty-Calibrated Peer Review and Self-Assessment

Teresita McCarty, Marie V. Parkes, Teresa T. Anderson, Jan Mines, Betty J. Skipper, and James Grebosky

### Abstract

#### Background

This study examines the effectiveness of Calibrated Peer Review™ (CPR), a Web-based writing development program, to teach and assess medical students' patient note-writing skills in a standardized fashion.

#### Method

At the end of the clerkship year, 67 medical students were divided into three groups, introduced to CPR, and

instructed in patient note-writing. Students then wrote notes for three clinical cases, presented in different order to each group. After training on faculty-calibrated standards, students evaluated their peers' notes and their own notes. Trained faculty, blinded to author, order, and group, also graded student notes.

#### Results

Faculty gave lower scores than students, but both groups found students' scores

improved significantly from the first to the third note written.

#### Conclusions

Student-written patient notes improved in quality while using CPR. The program uses approaches valued in medicine (accurate peer review and self-reflection) to enhance performance.

Acad Med. 2005;80(10 suppl):S67–S70.

**M**edical students must develop patient note-writing skills in order to document clinical findings and to demonstrate clinical reasoning. Medical educators face financial, regulatory, and time obstacles<sup>1,2</sup> in providing students with the feedback needed for patient note improvement. Learning the writing and reasoning skills manifest in patient notes remains clinically important, and is specifically measured in the licensing examination. We studied the utility of Calibrated Peer Review™ (CPR), a Web-based writing development program, to optimize the learning opportunities associated with practice and feedback in order to help teach and assess medical students' note-writing skills.

Recent changes in the practice environment have affected medical students' patient note-writing training. Documentation regulations have decreased the attention given to student notes; at the same time, changes at the

national licensing level have increased the importance of patient notes to medical students. After each of the ten to 12 standardized patient encounters students see during the United States Medical Licensing Examination (USMLE)™ Step 2 Clinical Skills, students have ten minutes to write a patient note. These notes are scored by trained physician raters, and together with data-gathering checklists, comprise one of the examination's three components.<sup>3</sup> In spite of obstacles, it is important for medical educators to ensure students are able to write well-organized, concise, yet complete patient notes that reflect coherent medical reasoning.<sup>4</sup> At the University of New Mexico, the patient note feedback given to students was often idiosyncratic and inconsistent and scoring was usually analytic rather than holistic. In meetings of various faculty groups who teach and provide feedback to students, a decision was made to adopt a consistent, holistic approach for scoring patient notes. CPR has been demonstrated to have an effect on learning<sup>5</sup> and was chosen as one method to disseminate consistent patient-note feedback to students.

#### Calibrated Peer Review

Calibrated Peer Review™ is a Web-based program that allows students to develop writing and critical evaluation skills.<sup>6</sup> CPR has been used at all educational levels, from elementary through graduate school, and in a variety of disciplines including initial development in chemistry.

Students are typically given several days or weeks in which to complete CPR assignments; students complete the following steps for any CPR assignment regardless of time frame, discipline, or educational level. Provided with a context—for our purposes a patient encounter—students complete an assignment by:

writing a focused patient note into the CPR program,

evaluating three faculty-written "calibration" patient notes of varying quality on the same patient–clinician encounter,

evaluating three randomly selected, anonymous, peer-written patient notes on the same patient–clinician encounter, and

performing a self-evaluation on the student's own patient note.

Correspondence: Teresita McCarty, MD, Undergraduate Medical Education – Assessment & Learning, MSC08 4710, 1 University of New Mexico, Albuquerque, NM 87131; e-mail: (tmccarty@salud.unm.edu).

For the seven evaluations associated with each patient case, the student answers a series of questions relating to both the content and the style of the patient note, assigns the patient note a global rating on a scale of 1–10 (1–3 represents unacceptable notes, 4–7 represents acceptable notes, and 8–10 represents outstanding notes), and provides feedback to the note-writer. Both the content and style questions follow a faculty-generated standard “key.” The student receives a grade based on:

- the quality of the student’s note,
- the competence of the student in reviewing the calibration patient notes,
- the student’s ratings of the three peer-reviewed patient notes, and
- the student’s rating of his or her own patient note.

Supported by funding from the Howard Hughes Medical Institute, there is presently no charge for the use of CPR. The program provides general and case-specific feedback, bases assessment on clear criteria, and uses concepts that are valued in medicine: peer review<sup>7</sup> and self-assessment.<sup>8</sup>

The purpose of this article is to evaluate whether medical students near the end of their third year improve their note-writing skills through repeated CPR cases and whether note-writing skills are associated with particular clinical cases or the clinical rotations the students are completing.

## Method

Human Research Review Committee approval was obtained for this study. Three clinical cases—a young woman with dysuria, an older man with diarrhea, and a nine-year-old boy with a rash—were written for use with CPR. For each

case a seven- to 12-minute video documenting a clinician interacting with a simulated patient was produced. Each video contained the elements of the history and physical examination pertinent to the case. During three expert pilot-testing sessions, faculty participants watched the videos, wrote patient notes, and completed each step of the CPR program as if they were students. These pilot sessions were used to further define the essential elements of focused patient notes, improve the scoring standards, and refine the format of the student CPR introductory workshop.

At the end of the clerkship year, all 67 students in the Class of 2005 were divided into three groups based on the last clerkship completed. Students attended a mandatory workshop where they were introduced to CPR and received instruction in writing focused patient notes. After viewing a clinician–patient video, students were given ten minutes to write a focused patient note in the CPR program. Each student then completed the remaining steps in the CPR process. The steps were repeated for each video, but the order in which the cases were presented was different for each cohort (Table 1). Two of the resulting 201 patient notes were discarded because of student entry errors, leaving 199 patient notes for analysis. Of the multiple scores generated from CPR, this study focuses on the student text score, which is a weighted average of the three peer scores. The weight assigned to peers’ note scores is determined by the peer’s ability to accurately score the calibration notes.

Two faculty members were trained to consistently apply the ten-point global rating scale by participating in a four-hour training session in which they scored and discussed notes written about the same cases by students who were not part of this study. The faculty members

then scored the study notes, blinded to student identity, order in which the student viewed the case, and clerkship cohort. Faculty scores were more than two points apart for 17% of the notes; these were scored by the faculty trainer, who then met with the two faculty raters to resolve the differences. The faculty score used for analysis was the average of the two faculty raters’ final scores.

The generalized estimating equation (GEE) for longitudinal data<sup>9</sup> was used since each student completed a sequence of three notes. The dependent variables were the score given by the student’s peers and the average faculty score; the independent variables were the order in which the clinical note was written, the case, and the final clerkship group. The GEE parameters and least squares means were calculated using PROC GENMOD in Statistical Analysis Software (SAS Institute, Cary, NC).

## Results

Forty of the 67 medical students were women (60%) and 27 (40%) were men. Although this intervention occurred at the end of the third year, two students (3%) had two and eight students (12%) had a single clerkship still to complete. Students in the three groups were academically similar: their clerkship clinical grade point averages, average scores on the seven National Board of Medical Examiners’ Clinical Science Subject Examinations, and average scores on the Comprehensive Clinical Skills Examination were not significantly different.

Student score analysis found the adjusted mean score of the first note written by students was significantly lower than the second ( $p = .03$ ) and third ( $p = .0002$ ) notes, with adjusted mean scores of 7.09, 7.45, and 7.58 respectively (Table 2). Student scores were not significantly

**Table 1**  
**Structure of CPR Workshop Intervention and Order of Patient Cases**

Final clerkship group	n	Introduction	First case	Second case	Third case
Family medicine, internal medicine	22	Online CPR tour Online pretest Practice case Note-writing instruction	Young woman with dysuria	Boy with rash	Older man with diarrhea
Neurology, pediatrics, psychiatry	22		Older man with diarrhea	Young woman with dysuria	Boy with rash
Obstetrics–gynecology, surgery	23		Boy with rash	Older man with diarrhea	Young woman with dysuria

different across the three cases. However, the adjusted mean score of students in the family medicine/internal medicine final clerkship group was significantly higher than the neurology/pediatrics/psychiatry group ( $p = .002$ ) and the obstetrics–gynecology/surgery group ( $p = .03$ ).

Faculty score analysis showed the adjusted mean score of the first note written by students was significantly lower than the third note (4.11 and 4.59 respectively,  $p = .004$ ). Although the faculty scores were lower than student scores, the pattern of improvement was consistent. The adjusted mean score of notes written about the young woman with dysuria was significantly higher than the adjusted mean score of notes written about the older man with diarrhea ( $p = .0003$ ) and the boy with the rash ( $p = .0001$ ). There was no statistically significant difference in faculty scores of the notes of students in the three final clerkship groups.

## Discussion

The results of this study support our hypothesis that using CPR improves the quality of medical student patient notes. The scores trended upward for both student and faculty scoring of student notes and were significantly different between the first and third note for both faculty and student raters.

Although student scores did not vary significantly across the three cases, faculty raters gave significantly higher scores for the dysuria case; during the scoring session, faculty raters commented that dysuria is a common patient presentation and seemed to have the easiest differential. Faculty raters also commented that the boy with the rash seemed relatively harder for students and faculty score averages ultimately reflected these views. While case-specific information is provided as part of CPR, this information is not available until after students have written their patient notes. This information can affect the rigor of students' peer review. In this instance, faculty knowledge and experience outweighed the students' ability to use what they had learned about specific patient presentations to evaluate the case-specific aspects of the work of their peers as effectively as did the faculty. Students did learn to more rigorously evaluate the global patient-note characteristics—those less dependent on case content—and the learning effect was demonstrated independent of the order in which the cases were presented.

Although students who had just completed family or internal medicine clerkships were more likely to write notes that were scored higher by their peers, this group did not differ academically from the other groups and faculty did not award them higher scores. It is unlikely

the notes were actually better and more likely the students were less rigorous in evaluating one another's notes. Higher CPR scores have been observed in other settings with smaller groups of students who spend more time with one another (M. V. Parkes, personal communication, 2004), which may be the case for students on the two medicine clerkships.

As seen in Table 2, on average, patient-note scores given by students were three points higher than scores given by faculty. Three factors could have influenced this difference. First, there were note-scoring training differences: faculty training was more extensive and focused on the behaviorally anchored, paper version of the global rating scale, while student training was brief with as much emphasis on answering content and style questions as on the global rating scale. Although students were trained to use the anchored scale, they had more exposure to the unanchored electronic version of the scale. Second, the faculty raters have much more experience reviewing patient notes, most of which are written without time limits, causing them to judge notes constrained to ten minutes more harshly than the students did. Third, the students' camaraderie may have affected their evaluation of their peers' notes. The shared history of students at the end of their third year may have resulted in evaluating their peers generously during a workshop held on an examination day.

Table 2  
Analysis of Patient Note Scores Assigned by Students and Faculty

Variables	Student scores Adj. mean (SE)	Faculty scores Adj. mean (SE)
<b>Order completed</b>		
First case	7.09 (.12)*	4.11 (.13) <sup>†</sup>
Second case	7.45 (.13)	4.28 (.11)
Third case	7.58 (.12)	4.59 (.13)
<b>Case description</b>		
Young woman with dysuria	7.48 (.13)	4.76 (.13) <sup>‡</sup>
Older man with diarrhea	7.38 (.12)	4.13 (.13)
Boy with rash	7.26 (.13)	4.09 (.12)
<b>Final clerkship group</b>		
Family medicine, internal medicine	7.73 (.10) <sup>§</sup>	4.47 (.13)
Neurology, pediatrics, psychiatry	7.13 (.16)	4.46 (.11)
Obstetrics–gynecology, surgery	7.27 (.18)	4.06 (.16)

\* Student scores differ significantly from second case ( $p = .03$ ) and third case ( $p = .0002$ ).

<sup>†</sup> Faculty scores differ significantly from third case ( $p = .004$ ).

<sup>‡</sup> Faculty scores differ significantly from diarrhea case ( $p = .0003$ ) and rash case ( $p = .0001$ ).

<sup>§</sup> Student scores differ significantly from neurology/pediatrics/psychiatry ( $p = .002$ ) and obstetrics–gynecology/surgery ( $p = .03$ ).

There are a number of limitations to this study. The CPR workshop allowed students only a limited amount of time to review the case-specific and general note-writing material. Since the feedback imbedded in the assignments is an important mechanism for learning, students could presumably have learned more if they had more time to review the feedback. Secondly, the timing of the intervention may have reduced the effect of the note-writing improvement. Students in this study had completed their clerkship year and were already proficient at writing clinical notes; had the intervention occurred earlier in their medical training, there might have been a greater improvement in the quality of written notes. Additionally, the training of the faculty and students in assigning scores differed: the two faculty reviewers had less experience answering the content and style questions associated with each

assignment and more training in the application of the 1–10 rating scale than the students did. Finally, we do not know if the improvement in note quality seen in this study will apply to the notes written in actual patient situations. We are presently studying notes written after standardized patient encounters with and without the CPR intervention to determine whether the patient note-writing improvement transfers to the simulated clinical setting.

Calibrated Peer Review™ is a well-developed educational tool with potential for many applications in medical education. Students have commented that they feel the program was helpful in preparing for the note-writing portion of USMLE Step 2 Clinical Skills. CPR is well suited to content areas such as evidence-based medicine and to settings where students are at different and distant sites,

or when large groups of students make individual faculty feedback difficult. Emphasizing evaluation, the CPR program standardizes faculty effort while reinforcing informed and accurate peer- and self-review of patient notes.

Thanks to William Anderson, PhD, S. Scott Obenshain, MD, Miriam Friedman, PhD, Summers Kalishman, PhD, Craig Timm, MD, Lisa Serna, and Eve Espey, MD.

## References

- 1 Brodkey AC, Sierles FS, Spertus IL, Weiner CL, McCurdy FA. Clerkship directors' perceptions of the effects of managed care on medical students' education. *Acad Med.* 2002;77: 1112–20.
- 2 Jones MD Jr, Boat T, Adler R, et al. Final report of the FOPE II financing of pediatric education workgroup. *Pediatrics.* 2000;106: 1256–69.
- 3 United States Medical Licensing Examination Step 2 Clinical Skills (CS) Content Description and General Information (<http://www.usmle.org/step2/Step2CS/Step2CS2005G1/TOC.asp>). Federation of State Medical Boards of the United States, Inc, and the National Board of Medical Examiners, 2004. Accessed 2 August, 2005.
- 4 Barzansky B. Educational programs in US medical schools, 2003–2004. *JAMA.* 2004;292: 1025–31.
- 5 Pelaez NJ. Problem-based writing with peer review improves academic performance in physiology. *Adv Physiol Educ.* 2002;26:174–84.
- 6 Chapman OL. The white paper: a description of CPR (<http://cpr.molsci.ucla.edu/>). University of California, Los Angeles, 2001. Accessed 2 August 2005.
- 7 Ramsey PG, Wenrich MD. Peer ratings: an assessment tool whose time has come. *J Gen Intern Med.* 1999;14:581–82.
- 8 Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. *Adv Health Sci Educ.* 2002;7:63–80.
- 9 Zeger SL, Liang KY. Longitudinal data-analysis for discrete and continuous outcomes. *Biometrics.* 1986;42:121–30.