Selecting the Best Clinical Vignettes for Academic Meetings: Should the Scoring Tool Criteria be Modified?

Jeremiah Newsom, MD^{1,2}, Carlos A. Estrada, MD MS^{1,2}, Danny Panisko, MD FRCP³, and Lisa Willett, MD¹

¹The University of Alabama at Birmingham, Birmingham, AL, USA; ²Birmingham VA Medical Center, Veterans Affairs National Quality Scholars Program, Birmingham, AL, USA; ³University of Toronto, Toronto, ON, Canada.

BACKGROUND: The performance of scoring tools to select clinical vignettes for presentation at academic meetings has never been assessed.

OBJECTIVE: To measure the psychometric properties of two scoring tools used to select clinical vignettes and to determine which elements are most helpful.

DESIGN: Prospective observational study.

PARTICIPANTS: Participants submitting clinical vignette abstracts, Society of General Internal Medicine annual meetings (2006–2007).

MAIN MEASURES: The 2006 scoring tool had three criteria (clarity, significance, and relevance) with brief general descriptors. The 2007 modified tool had five criteria (clarity, significance, relevance, teaching value, and overall assessment) with more detailed descriptors.

KEY RESULTS: A total of 938 clinical vignette abstracts were submitted (484 in 2006; 454 in 2007); 59.5% (n=288) were accepted for presentation. Cronbach's alpha was 0.81 for the 2006 three-item tool and 0.95 for the 2007 modified five-item tool. Simplifying the five-item 2007 tool to three items (relevance, teaching value, overall assessment) yielded a Cronbach's alpha of 0.95. The agreement between the number of clinical vignettes accepted for presentation (2007) using the average score of the five items with the number that would have been accepted using the simplified three items (relevance, teaching value, overall assessment) was almost perfect, with kappa 0.89 (95% confidence interval, 0.85 to 0.93).

CONCLUSIONS: Both scoring tools performed well, but a simplified tool with three items (relevance, teaching value, and overall assessment) and detailed descriptors was optimal; the simplified tool could improve the reviewer efficiency and quality of clinical vignettes presented at national meetings.

KEY WORDS: professional competence; internship and residency; education medical/methods; educational measurement/methods/standards; congresses as topic; reproducibility of results.

J Gen Intern Med 27(2):202–6

DOI: 10.1007/s11606-011-1879-2

© Society of General Internal Medicine 2011

BACKGROUND

Trainees and faculty present clinical vignettes at annual academic meetings such as the American College of Physicians (ACP)¹, American Federation for Medical Research (AFMR)², Society of General Internal Medicine (SGIM)³, and the Society of Hospital Medicine (SHM)⁴. Presenting clinical vignettes can provide proof of scholarly activity for residents^{5,6} and may fulfill academic promotion requirements for faculty. Furthermore, presenting vignettes at meetings has other benefits and may lead to subsequent publications^{7–10}.

Although no data are available, our experience is that clinical vignette submissions are on the rise, and a process to ensure those with the highest teaching value are selected for presentation is increasingly more challenging. Once submitted, vignettes undergo a peer-review process whereby volunteer faculty score the submissions using internally developed criteria^{1–4}. Depending on the number and definitions of the scoring criteria, the process can be time consuming and unclear for reviewers. However, we are unaware of any studies examining the psychometric properties for any of the various scoring methods used to select a clinical vignette for presentation.

A standardized method to evaluate clinical vignette submissions could improve the selection process. Identifying the key elements for a high quality vignette would allow organizations to better develop scoring criteria to ensure the best vignettes are accepted for presentation. Therefore, the aim of our study is to measure the psychometric properties of two different clinical vignette scoring tools and to determine which elements of the tools are most helpful in selecting vignettes for presentation.

METHODS

Design, Setting, and Participants

We conducted a 2-year prospective observational study of two tools used to score clinical vignette submissions for a national

Received May 25, 2010

Revised August 24, 2010

Accepted September 1, 2011

Published online September 17, 2011

academic general internal medicine meeting, the Society of General Internal Medicine (SGIM). We evaluated the performance of the scoring tool used for the 2006 meeting (29th Annual Meeting, Los Angeles) and a modified scoring tool used for the 2007 meeting (30th Annual Meeting, Toronto). SGIM is the premier academic society for general internists and has approximately 3,000 members; its main goal is to promote improved patient care, research, and education in primary care and general internal medicine. The annual meeting draws approximately 1,700 participants per year presenting research abstracts, innovations in clinical practice and education, workshops, and case reports or clinical vignettes. The present study focuses on the scoring tool to select clinical vignettes. As part of the peer-review process to select clinical vignettes for presentation, each clinical vignette was scored by four to five participants. Participants are SGIM members, and every year, 70–100 are invited to review 20–30 clinical vignette submissions unaware of the authors' identity or institution. Participation was voluntary and without incentives.

Scoring Tools: Baseline and Modified

The 2006 scoring tool had been used for 6 years at SGIM national meetings and had three items with brief descriptors: (1) clarity of presentation (brevity, pertinence of data, grammatically structured), (2) significance of learning objectives (unique teaching points that may not be generally known or understood), and (3) relevance of the topic to the clinical practice of general internal medicine. It used a Likert scale: 1=worst, 7=best. The final score was the average of the individual ratings.

In an effort to improve the 2006 scoring tool, it was modified following three steps: (1) review of the literature to identify domains of value for clinical vignettes, (2) review of the domains for face validity, and (3) cognitive interviewing with faculty experienced in clinical vignette selection. First, we found no guidance from the literature on published scoring tools for clinical vignette selection at academic meetings. We then reviewed the case report and clinical vignette literature,^{11–16} and reviewed the American College of Physicians scoring criteria for clinical vignette award selection¹⁷. From those, we identified general domains for a high quality clinical vignette. The first draft of the domains included items related to format and content. Format domains initially included 166 adjectives describing the quality of the title, background, objectives, case description, graphics, discussion, and overall assessment. Content domains included over 25 adjectives describing the quality of the case description and clinical or educational significance. We then conducted several cycles of cognitive interviewing at two institutions (D.P., Canada; C.A.E., US) among a convenience sample of clinician educators, house staff, and chief medical residents.

The final modified scoring tool had five items and included more detailed select adjectives and phrases to describe the intent of each of the five items: (1) clarity of presentation (concise, complete, organized, well-written, focused objectives), (2) significance (clinically important, contributes to scientific knowledge, unique, interesting), (3) relevance to general internal medicine (describes impact on clinical practice, teaching/education, or future research, places case in context), (4) teaching value (offers an important diagnosis, physical examination, or manage-

ment pearl), and (5) overall assessment (overall evidence of scholarship, potential for publication). It used a Likert scale: 1=worst, 7=best. The final score was the average of the individual ratings.

Analysis

We used standard descriptive statistics; we also analyzed inter-item correlation with Spearman's rho and explored the number of constructs among the items with factor analysis. We calculated Cronbach's alpha to determine the internal consistency of the scoring tools; furthermore, we conducted sensitivity analysis of the modified scoring tool to explore whether fewer items could be used. As Cronbach's alpha is directly associated with an increased number of items, we calculated alpha for several three-item combinations from the modified tool criteria.

Finally, to illustrate the potential implications of using a simplified version of the modified scoring tool, we compared the agreement of the actual number accepted for presentation at the 2007 meeting based on meeting capacity using the final score (average of 5 items) with the number that would have been accepted using the score from a simplified version of the modified tool with the Kappa (K) statistic. A K value of 0–0.19 is considered poor agreement, 0.20–0.39 fair, 0.40–0.59 moderate, 0.60–0.79 substantial, and 0.80–1 almost perfect¹⁸.

All analyses were performed using SPSS Version 16.0 (Chicago, IL), and a p value of <0.05 was considered to determine statistical significance.

Disclosures

The Institutional Review Board at the University of Alabama at Birmingham approved the study.

RESULTS

A total of 938 clinical vignettes were submitted for peer-review representing two countries and more than 40 states and provinces. Of the 484 vignettes submitted for the 2006 meeting, 288 were accepted for presentation (59.5%; 78 for oral presentation and 210 for poster presentation). Of the 454 submitted for the 2007 meeting, 239 were accepted for presentation (52.7%; 38 for oral presentation and 201 for poster presentation). Of note, the number accepted for oral presentation for 2006 and 2007 was based on space availability. The scores for the 2007 meeting using the modified scoring tool were lower than for the 2006 meeting using the original scoring tool ($P < 0.001$; Table 1).

The inter-item correlations for the individual scoring criteria for the 2006 scoring tool ranged from 0.63 to 0.70 and for the 2007 modified tool from 0.61 to 0.88 (Table 2). Factor analysis showed that one domain explained 80.2% of the variability in the 2006 tool (eigenvalue 2.41) and 83.0% of the variability of the 2007 modified tool (eigenvalue 4.15).

The Cronbach's alpha for the 2006 scoring tool was 0.81 and for the 2007 modified tool was 0.95. To see how the change in descriptors affected the internal consistency of the tool, we

Table 1. Average Scores of the Clinical Vignette Abstract Submissions at Two National Meetings (2006 and 2007)

Item	2006 N=484	2007 N=454
Clarity*	4.9 (0.8)	4.7 (0.8)
Significance*	4.9 (0.7)	4.8 (0.7)
Relevance*	4.9 (0.8)	4.7 (0.7)
Teaching	–	4.7 (0.8)
Overall	–	4.6 (0.8)
Total*	4.9 (0.7)	4.6 (0.8)

Values are means (SD). Likert scale, 1=worst, 7=best

*A Student's *t*-test was used to compare the average scores between the 2 years and $p < 0.001$ for all comparisons

found the Cronbach's alpha to be 0.87 for the same three scoring criteria from the 2007 modified tool as those that made up the 2006 tool (clarity, significance, and relevance). To find the best three-item scoring tool, we calculated Cronbach's alpha for several combinations of the scoring criteria and found the Cronbach's alpha to be 0.95 for relevance, teaching value, and overall assessment. If either teaching value or overall assessment were removed from the five-item scoring tool, the Cronbach's alpha decreased to 0.93 and 0.92, respectively.

The agreement between the number of clinical vignettes accepted for presentation at the 2007 meeting using the average score of the five-item modified tool with the number that would have been accepted using the simplified three-item modified tool that included relevance, teaching value, and overall assessment was almost perfect, with kappa 0.89 (95% confidence interval, 0.85 to 0.93; $P < 0.001$); the overall agreement was 94.5% (Table 3). We found that of the 239 vignettes that were selected for presentation with the five-item modified tool, 226 would have still been accepted with the simplified three-item modified tool, and 13 would have instead been rejected. Also, 12 of those that had been rejected would have been accepted if the simplified three-item modified tool was used.

DISCUSSION

The internal consistency, or reliability, of the two scoring tools to select clinical vignettes for presentation at a national

Table 3. Agreement Between Accept/Reject Decisions for the 2007 Meeting Using the Simplified Three-Item Modified Tool as Compared to the Original Five-Item Modified Tool (n=454)

	Final score (average of five items*)		
Score decision†	Accept	Reject	Total
Simplified Score (average of three items‡)			
Accept	226 (49.8%)	12 (2.6%)	238 (52.4%)
Reject	13 (2.9%)	203 (44.7%)	216 (47.6%)
Total	239 (52.6%)	215 (47.4%)	454 (100%)

*Five items=clarity of presentation, significance, relevance to general internal medicine, teaching value, overall assessment

†Three items=relevance to general internal medicine, teaching value, overall assessment

‡Kappa 0.89 (95% confidence interval, 0.85 to 0.93; $P < 0.001$)

meeting was excellent. The 2007 modified tool was an improvement from the 2006 tool due to higher inter-item correlations and Cronbach's alphas. We showed that the improvement was due to the addition of the two new scoring criteria (teaching value and overall assessment) more than the change in the descriptors. We also demonstrated that the modified five-item scoring tool could be pared down to three items with their improved descriptors and would perform as well as the five-item tool.

Clinical vignettes are an important educational tool, and therefore efforts to improve their value are important. Trainees use them as an educational method to learn clinical reasoning from patient presentations and diagnostic processes¹⁹. Clinical vignettes identify system issues affecting patient care outcomes, allowing for targeted quality improvement and patient safety efforts^{20,21}. By presenting vignettes at academic meetings, fellows, residents and medical students develop writing, presentation, and communication skills,^{6,8,11} and gain an opportunity for mentorship by faculty members⁶. Therefore, having an evidence-based method to ensure the best vignettes are selected for academic meetings is needed. The scientific rigor used for selecting research abstracts should be applied for the selection of clinical vignettes.

Our study has some potentially important implications. First, as the peer-review process is important for the quality of academic presentations, reviewers are volunteers with many obligations competing for their time. Reviewers are often asked to rate multiple clinical vignette submissions within a short timeframe. Our modified three-item tool is an important contribution to the peer-review process. Although we did not

Table 2. Inter-Item Correlation for the Scores of the Clinical Vignette Abstract Submissions at Two National Meetings (2006 and 2007)

Group item	Clarity	Significance	Relevance	Teaching	Overall
2006 (n=484)					
Clarity	1.0	–	–	N/A	N/A
Significance	0.70	1.0	–	N/A	N/A
Relevance	0.63	0.68	1.0	N/A	N/A
2007 (n=454)					
Clarity	1.0	–	–	–	–
Significance	0.65	1.0	–	–	–
Relevance	0.61	0.71	1.0	–	–
Teaching	0.71	0.82	0.79	1.0	–
Overall	0.78	0.88	0.76	0.88	1.0

All correlations, $P < 0.001$. N/A, not applicable

measure the time spent scoring the vignette abstracts, by using only three criteria to identify the best vignettes, reviewers may be more efficient in their reviews of the submissions.

Second, we have not formally tested the proposed and shortened three-item tool independently (relevance, teaching value, overall assessment); testing a single item independently (overall assessment) may be warranted in a future study. The factor analysis confirmed a single domain construct; we interpreted the single domain as quality. In addition, the meeting planners (D.P. and C.A.E.) also used the revised scoring tool to rank the clinical vignette submissions, top ranked ones were selected for oral presentations, medium ranked ones were selected for poster presentations, and lower ranked ones were not accepted. This further helps ensure the best vignettes are presented at the meeting and could improve the overall quality of the vignettes and the meeting itself.

Third, residency program directors could use the revised scoring tool to internally select the best clinical vignettes for submission to academic meetings. For example, at our institution, we have held an internal competition over the past 10 years to select the best clinical vignettes to submit to annual meetings. We have used the revised tool for the past 2 years for the internal competition.

Finally, while the psychometric properties of scoring tools have been examined for research abstracts submitted to national meetings²²⁻³³, the same evidence does not exist for clinical vignettes. We provide evidence of the psychometric properties of a scoring tool to select a clinical vignette for presentation at academic meetings.

Our study has some limitations. First, we examined the experience at one academic society and our findings may not be applicable to other specialties. However, among academic general internal medicine meetings, our findings are applicable to meetings such as ACP, SHM, and AFMR. Second, we have not formally tested the proposed and shortened three-item tool independently (relevance, teaching value, overall assessment). Third, the reviewers participating in the peer-review process over the 2-year study were likely different, and we were unable to calculate inter-rater reliability.

In conclusion, our study is the first to measure the psychometric properties of a clinical vignette scoring tool, specifically to compare the relative value of two clinical vignette scoring tools. We identified that the most important criteria are: (1) relevance to general internal medicine (describes impact on clinical practice, teaching/education, or future research, places case in context), (2) teaching value (offers an important diagnosis, physical examination, or management pearl), and (3) overall assessment (overall evidence of scholarship, potential for publication), and that a three-item tool performed as well as a longer, five-item tool. We plan to continue this evaluation of our scoring tool with continued refinement and further study of inter-rater reliabilities as well as other psychometric properties. Future research may even compare a one-item overall score with the three-item tool. We ultimately hope to ensure the best clinical vignettes are chosen with the scoring tools and the valuable time of volunteer faculty reviewers is used most efficiently.

Contributors: None

Funders: None

Prior Presentations: The main findings of this paper were presented in part as a poster at the Internal Medicine 2010, American College of Physicians Annual Meeting (22–24 April 2010), and as oral presentations at the Southern Society of General Internal Medicine Annual Meeting (25–27 February 2010) and at the 33rd Society of General Internal Medicine Annual Meeting (28 April–1 May 2010).

Conflict of Interest: None disclosed.

Corresponding Author: Lisa Willett, MD; BDB 341, 1530 3rd Avenue South, Birmingham, AL 35294-0012, USA (e-mail: lwillett@uab.edu).

REFERENCES

1. American College of Physicians. Associate Abstract Competitions. Available at: http://www.acponline.org/residents_fellows/competitions/abstract. Accessed August 11, 2011.
2. American Federation for Medical Research. Available at: <http://www.afmr.org>. Accessed August 11, 2011.
3. Society of General Internal Medicine. Clinical Vignette Submission Info. Available at: <http://www.sgm.org/index.cfm?pageid=1138>. Accessed August 11, 2011.
4. Society of Hospital Medicine. SHM RIV Competition Abstracts. Available at: <http://www.hospitalmedicine.org/Content/NavigationMenu/Education/Abstracts/Abstracts.htm>. Accessed August 11, 2011.
5. Accreditation Council on Graduate Medical Education (ACGME). Educational Program, Residents' Scholarly Activities. Available at: http://www.acgme.org/acWebsite/navPages/commonpr_documents/IVB123_EducationalProgram_ResidentScholarlyActivity_Explanation.pdf. Accessed August 11, 2011.
6. Rivera JA, Levine RB, Wright SM. Completing a Scholarly Project During Residency Training. Perspectives of residents who have been successful. *J Gen Intern Med*. 2005;20:366–69.
7. Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001;134:330–4.
8. Sridhar ARM, Willett LL, Castiglioni A, et al. Scholarship opportunities for trainees and clinician educators: learning outcomes from a case report writing workshop. *J Gen Intern Med*. 2008;24(3):398–401.
9. Drenth JPH, Smits P, Thien T, Stalenhoef AFH. The case for case reports in the Netherlands Journal of Medicine. *Neth J Med*. 2006;64:262–4.
10. Estrada CA, Heudebert GR, Centor RM. JGIM new section: case reports and clinical vignettes. *J Gen Intern Med*. 2005;20:971.
11. Willett LL, Paranjape A, Estrada CA. Identifying key components for an effective case report poster: an observational study. *J Gen Intern Med*. 2008;24(3):393–7.
12. Wright SM, Kouroukis C. Capturing zebras: what to do with a reportable case. *CMAJ*. 2000;163:429–31.
13. Brodell RT. Do more than discuss that unusual case: write it up. *Postgrad Med*. 2000;108:19–20, 23.
14. Bignall J, Horton R. Learning from stories—The Lancet's case reports. *Lancet*. 1995;346:1246.
15. Huston P, Squires BP. Case reports: information for authors and peer reviewers. *CMAJ*. 1996;154:43–4.
16. Sorinola O, Olufowobi O, Coomarasamy A, Khan KS. Instructions to authors for case reporting are limited: a review of a core journal list. *BMC Med Educ*. 2004; 4: 4. Available at: <http://www.biomedcentral.com/1472-6920/4/4>. Accessed August 11, 2011.
17. American College of Physicians. Criteria and Grading System Sheet Associates Clinical Vignette Competition: Poster Division Finals. Available at: http://www.acponline.org/residents_fellows/competitions/abstract/prepare/judge_criteria.pdf. Accessed August 11, 2011.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
19. Gross CP, Donnelly GB, Reisman AB, Sepkowitz KA, Callahan MA. Resident expectations of morning report: a multi-institutional study. *Arch Intern Med*. 1999;159:1910–4.
20. Aboumatar HJ, Blackledge CG, Dickson C, Heitmiller E, Freischlag J, Pronovost PJ. A descriptive study of morbidity and mortality conferences and their conformity to medical incident analysis models:

- results of the Morbidity and Mortality Conference Improvement Project, Phase 1. *Am J Med Qual*. 2007;22(4):232–8.
21. **Aeby TC**. The role of morbidity and mortality (M&M) conferences in medical education. *Hawaii Med J*. 2011;70(2):39–40.
22. **Poolman RW, Keijser LC, de Waal Malefijt MC, Blankevoort L, Farrokhyar F, Bhandari M**. Reviewer agreement in scoring 419 abstracts for scientific orthopedics meetings. *Acta Orthop*. 2007;78(2):278–84.
23. **van der Steen LP, Hage JJ, Kon M, Mazzola R**. Reliability of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg*. 2003;111(7):2215–22.
24. **van der Steen LP, Hage JJ, Kon M, Monstrey SJ**. Validity of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg*. 2004;113(1):353–9.
25. **Bydder S, Marion K, Taylor M, Semmens J**. Assessment of abstracts submitted to the annual scientific meeting of the Royal Australian and New Zealand College of Radiologists. *Australas Radiol*. 2006;50(4):355–9.
26. **Glover GH, Henkelman RM**. Abstract scoring for the annual SMR program: significance of reviewer score normalization. *Magn Reson Med*. 1994;32(4):435–9.
27. **Kemper KJ, McCarthy PL, Cicchetti DV**. Improving participation and interrater agreement in scoring Ambulatory Pediatric Association abstracts. How well have we succeeded? *Arch Pediatr Adolesc Med*. 1996;150(4):380–3.
28. **Montgomery AA, Graham A, Evans PH, Fahey T**. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Serv Res*. 2002;2(1):8.
29. **Rothwell PM, Martyn CN**. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain*. 2000;123(Pt 9):1964–9.
30. **Rowe BH, Strome TL, Spooner C, Blitz S, Grafstein E, Worster A**. Reviewer agreement trends from four years of electronic submissions of conference abstract. *BMC Med Res Methodol*. 2006;6:14.
31. **Rubin HR, Redelmeier DA, Wu AW, Steinberg EP**. How reliable is peer review of scientific abstracts? Looking back at the 1991 Annual Meeting of the Society of General Internal Medicine. *J Gen Intern Med*. 1993;8(5):255–8.
32. **Schmidt AH, Zhao G, Turkelson C**. Levels of evidence at the AAOS meeting: can authors rate their own submissions, and do other raters agree? *J Bone Joint Surg Am*. 2009;91(4):867–3.
33. **Timmer A, Sutherland LR, Hilsden RJ**. Development and evaluation of a quality score for abstracts. *BMC Med Res Methodol*. 2003;3:2.