

Rubric Evaluation of Pediatric Emergency Medicine Fellows

DEBORAH C. HSU, MD, MEd
CHARLES G. MACIAS, MD, MPH

Abstract

Objectives To develop and validate a rubric assessment instrument for use by pediatric emergency medicine (PEM) faculty to evaluate PEM fellows and for fellows to use to self-assess.

Methods This is a prospective study at a PEM fellowship program. The assessment instrument was developed through a multistep process: (1) development of rubric format items, scaled on the modified Dreyfus model proficiency levels, corresponding to the 6 Accreditation Council for Graduate Medical Education core competencies; (2) determination of content and construct validity of the items through structured input and item refinement by subject matter experts and focus group review; (3) collection of data using a 61-item form; (4) evaluation of psychometrics; (5) selection of items for use in the final instrument.

Results A total of 261 evaluations were collected from 2006 to 2007; exploratory factor analysis yielded 5 factors with Eigenvalues >1.0 ; each contained ≥ 4 items, with factor loadings >0.4 corresponding with the following competencies: (1) medical knowledge and practice-based learning and improvement, (2) patient care and systems-based practice, (3) interpersonal skills, (4) communication skills, and (5) professionalism. Cronbach α for the final 53-item instrument was 0.989. There was also significant responsiveness of the tool to the year of training.

Conclusion A substantively and statistically validated rubric evaluation of PEM fellows is a reliable tool for formative and summative evaluation.

Editor's Note: The online version of this article contains the rubric-based assessment instrument used in this study and tables of mean faculty scores and self-evaluations scores.

Background

Educational assessment serves 2 purposes: (1) providing support for student learning processes (formative assessment), and (2) determining the status of learning and

Both authors are at Baylor College of Medicine and Texas Children's Hospital. **Deborah C. Hsu, MD, MEd**, is Associate Professor, Fellowship Program Director, and Director of Education in the Department of Pediatrics Section of Emergency Medicine; **Charles G. Macias, MD, MPH**, is Associate Professor, Chief of Academic Programs, and Research Director, Department of Pediatrics, Section of Emergency Medicine.

An earlier draft of this article was presented as a Medical education platform at the Annual Conference of the Pediatric Academic Societies in Honolulu, HI, and as a poster presentation at the Baylor College of Medicine Academy of Distinguished Educators Showcase, Houston, TX.

The authors would like to thank the Baylor College of Medicine pediatric emergency medicine fellows and faculty without whom this project would not have been possible; Will A. Weber, PhD (deceased), and Catherine Horn, PhD, University of Houston, College of Education, for their guidance in conducting educational research and instrument development; and Erin Endom, MD, for her help in manuscript review.

Corresponding author: Deborah Hsu, MD, MEd, Pediatric Emergency Medicine, Texas Children's Hospital, 6621 Fannin St Ste A 210 MC1-1481, Houston, TX 77030, 832.824.5487, dchsu@texaschildrens.org

Received May 10, 2010; revision received June 12, 2010; accepted June 24, 2010.

DOI: 10.4300/JGME-D-10-00083.1

performance (summative assessment).¹ In addition, an educational assessment should be fair, be based on learner ability, ensure all learners receive “the same or equal opportunity to perform.”^{2(p3)} In graduate medical education, global ratings of resident physicians by faculty are the most widely used method of assessment, and often use Likert-type rating scales to measure competence,^{1,3} yet research regarding such rating forms show wide variability in validity and reliability.^{2,4-10} Likert-type rating-scale assessments that consist of numeric ratings, even when accompanied by qualitative labels, such as *competent* or *not competent*, often yield scores that are subjectively derived with limited value in formative evaluation¹ because they lack detailed requirements of performance expectations and behavioral descriptions for each domain.¹¹

The Accreditation Council for Graduate Medical Education (ACGME) requires that training programs evaluate trainee acquisition of the core competencies using dependable measures; therefore, evaluation processes must be qualitatively and quantitatively validated.¹² Criteria for developing evaluation items include (1) consensus among evaluators that the items reflect the intent or definition of the given competency; (2) frequent occurrence of items or actions; and (3) transparency of items or actions.¹³ Additionally, the application of principles of psychometric theory may provide quantitative evidence of validity and reliability. Well-defined scoring/rating criteria and training of observers and raters have been associated with higher

reliability of such tools.¹ Scales incorporating distinct behavioral descriptions that provide specific information may contribute to the learning process.

The scoring rubric is a method of assessment that has been extensively studied and is gaining recognition in professional education.^{10,14–25} It uses specified evaluation criteria and proficiency levels to gauge student achievement; each point on a fixed scale is described by a list of performance characteristics.^{10,26–28} Rubrics can aid teachers in the measurement of “products, progress, and the process of learning,”^{26(p2)} as well as provide clear performance targets.^{27,28} Advantages associated with well-written rubrics include their relative ease of use by instructors and learners, their ability to provide informative feedback to students, their consistency in scoring, their ability to facilitate communication between evaluators and learners, their support for learner self-assessment and skill development, and their familiarity to physician evaluators (Apgar Score and the Glasgow Coma Scale).^{10,15,16} The challenges with rubrics are related to the development process in which criteria for evaluation are identified, levels of performance are described, and definitive examples of performance at various levels are written in measurable terms.²⁷

Appropriately designed scoring rubrics with objective criteria and strong psychometric properties have benefits as an assessment method for evaluating resident physicians’ acquisition of competencies and providing them with both formative and summative evaluations. We sought to construct and validate a rubric assessment instrument for use by pediatric emergency medicine (PEM) faculty to evaluate the competence of PEM fellows and to allow fellows to self-assess.

Methods

Study Design, Setting, and Population

This study was conducted prospectively at a PEM fellowship program, which enrolls 15 full-time fellows annually and is staffed by 23 faculty members certified in the subspecialty of PEM. This study received an exemption from our Institutional Review Board. The practice setting is an urban, freestanding, children’s hospital emergency department with an annual census of >85 000 patients. Quarterly global evaluations of fellows by faculty comprise 1 part of a multimethod and multievaluator assessment system used by the program.

Study Protocol

Instrument Development—Item Writing and Refinement

A pool of items for the study instrument was initially developed and written in rubric form by D.C.H., our study’s principal investigator, who used and modified items from unpublished evaluation forms in use at institutions willing to share their forms with D.C.H. and from previously published resident-physician evaluation forms obtained in a

review of the literature (C. Hallstrom, MD, written communication).^{29–31} Behaviorally focused items, rated on a 5-point scale, were written in rubric format using modified Dreyfus-model levels of proficiency as headers (1, Novice; 2, Intermediate; 3, Competent; 4, Proficient; 5, Expert; 0, Cannot Assess) and were designed to demonstrate fellow attainment of the ACGME 6 general competencies.²⁹ (See supplemental APPENDIX 1 for the assessment instrument). The draft instrument was sent to local subject-matter experts in graduate medical education and PEM for review and revision; further review was conducted through a focus group of 12 key stakeholders in the fellowship program’s assessment process, including current PEM fellows, junior faculty who were recent graduates of the fellowship program, and senior faculty with significant experience in education and clinical practice. All group members were given a description of ACGME core competencies and a global description of fellow competency levels. The goals of the reviews were to (1) determine which items to retain in the instrument or to delete secondary to inadequate data for providing meaningful evaluation; (2) provide clarification of the intended meaning of each item with appropriate wording; (3) add specific behaviorally focused evaluation criteria; and (4) classify items into appropriate competencies.¹³

Data Collection Faculty and fellows were trained to use the rubric evaluation form via group and individual instruction. Training sessions entailed a division-specific faculty-development workshop on feedback and evaluation, division staff meetings, a meeting of the fellowship program’s faculty mentors, a meeting between the fellowship program directors and fellows, and, on an individual basis, to fellows and faculty who missed the group training sessions. In forum, we discussed the philosophic underpinnings of rubric evaluations, introduced segments of the tool, and provided examples of the domains covered by the evaluation. Ten to 15 minutes in each session were sufficient to discuss how the evaluation form was to be used by faculty and fellow evaluators. Different forums were used to ensure education of all faculty and fellows on the use of this new form. We believe this process helped (1) to ensure that faculty evaluated fellows based on the behavioral descriptions provided on the instrument, and (2) to make fellows aware of behaviors associated with the various competency levels for which they were providing self-assessments and being evaluated by faculty. After training, fellows provided self-evaluations and PEM faculty evaluated the fellows quarterly for 4 consecutive academic quarters using the instrument. Each faculty member was randomly assigned 4 fellows to evaluate each quarter. Each evaluation cycle was open for 3 weeks to ensure adequate time for completion and submission of evaluations. Evaluations were administered electronically via an institutionally restricted, secure, and confidential website-based form. After data analysis for each academic quarter,

TABLE 1 ITEMS THAT UNDERWENT CHANGES IN ACGME CORE COMPETENCY CLASSIFICATION AFTER EXPLORATORY FACTOR ANALYSIS

Item	Initial Competency: Competency Item Placed Into During Review by SMEs and Focus Group (EFA Factor Loading of Item Within This Competency)	Final Competency: Competency Item Placed Into by Exploratory Factor Analysis (EFA Factor Loading of Item Within This Competency)
Sensitivity to others	Professionalism (.351)	Interpersonal skills (.696)
Compassion; ability to provide support	Professionalism (.301)	Interpersonal skills (.743)
Personal appearance	Professionalism (.078)	Interpersonal skills (.490)
Ability to apply knowledge of testing modalities in patient management plans	Medical knowledge (.510)	Patient care (.587)
Ability to interpret diagnostic tests	Medical knowledge (.306)	Patient care (.682)
Ability to recognize personal limitations	Practice based learning and improvement (.421)	Professionalism (.542)
Acceptance of feedback	Practice based learning and improvement (.452)	Professionalism (.667)
General efforts to improve self	Practice based learning and improvement (.437)	Professionalism (.621)

Abbreviations: EFA, exploratory factor analysis; SME, subject matter expert.

fellows met individually with their faculty mentors to discuss their evaluations with a template of the evaluation form on hand.

Data Analysis Data were coded, entered into a data file, and analyzed using SPSS for Windows software, Release 15.0 (SPSS Inc, Chicago, IL). Factor analysis is a commonly used measure of construct validity.³² Exploratory factor analysis was conducted on data from faculty evaluations of fellows, using principal-components extraction and Varimax rotation. Any factor with an Eigenvalue of >1.0 was defined as a category measurable by the instrument.^{33–35} Based on the factor analysis and other substantive criteria, items were recategorized and included in the final instrument.

Internal consistency of the entire instrument, each factor, and each ACGME core competency was measured using Cronbach α coefficient statistic. Mean faculty scores obtained within individual competencies by fellows at different years of fellowship training were compared using Kruskal-Wallis 1-way analysis of variance followed by a Mann-Whitney U test if results were statistically significant. Mean faculty scores and fellow self-assessment scores obtained within each competency by fellows at different years of training were compared using independent-sample t test.

Results

A total of 261 faculty evaluations and 51 self-evaluations for 18 fellows were collected during the study period between October 2006 and October 2007. Exploratory factor analysis of faculty evaluation data yielded 5 factors with Eigenvalues ≥ 1.0 , with each factor containing at least

4 items with factor loadings greater than 0.4. Kaiser-Meyer-Olkin measure of sampling adequacy was .965 and Bartlett Test of Sphericity was significant with $P < .001$; these analyses demonstrated that exploratory factor analysis was appropriate for this data.³³ With factor analysis, some items initially classified under certain core competencies by subject-matter expert and focus group review changed classifications (TABLE 1). The 5 identified factors corresponded with the following competencies and together accounted for 71% of the total variance: (1) medical knowledge and practice-based learning and improvement, (2) patient care and systems-based practice, (3) interpersonal skills, (4) communication skills, and (5) professionalism. In the final instrument, all items mentioned in TABLE 1 were reclassified into the indicated core competencies from the exploratory factor analysis.

Faculty evaluating fellows selected the *cannot assess* response 0.4% to 28% of the time. Based on the Reisdorff criteria,¹³ items were excluded from the final instrument if they had a high frequency of *cannot assess* responses or were redundant. In the final instrument, 36 items (68%) had $\leq 4\%$ *cannot assess* response frequencies, 11 items (21%) had 6% to 10% *cannot assess* response frequencies, 4 items (8%) had 11% to 15% *cannot assess* response frequencies, and 2 items (3%) had 20% to 25% *cannot assess* response frequencies. Of the 4 items with 11% to 15% *cannot assess* response frequencies, 1 pertained to procedural technical skills, another to fellow assessment of patient discomfort, a third to chart documentation practices, and the final 1 to navigation of the health care system. The 2 items with the 20% to 25% rate of *cannot assess* responses were retained in the final instrument because they pertained to evaluating fellows on their abilities to recognize patients requiring

TABLE 2 RELIABILITIES OF EACH INDIVIDUAL FACTOR AND CORE COMPETENCY CATEGORY		
Factors Identified by EFA	No. of Items in Factor/Competency	Cronbach α
Factor 1: medical knowledge and practice-based learning and improvement	6	.94
Factor 2: patient care and systems-based practice	21	.98
Factor 3: interpersonal skills	11	.944
Factor 4: communication skills	4	.858
Factor 5: professionalism	11	.944
ACGME core competency		
Medical knowledge	3	.876
Practice-based learning and improvement	3	.92
Patient care	16	.973
Systems-based practice	5	.915
Interpersonal and communication skills	15	.956
Professionalism	11	.944

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; EFA, exploratory factor analysis.

resuscitations and to direct patient resuscitations, infrequent but important events. These items were retained in our final instrument because they pertained to opportunity-dependent processes fellows are unlikely to encounter every day.

The final instrument contains 53 items: 3 in medical knowledge (6%), 3 in practice-based learning and improvement (6%), 16 in patient care (30%), 5 in systems-based practice (9%), 15 in interpersonal and communication skills (28%), and 11 in professionalism (21%) (see online supplemental APPENDIX 1). Because this instrument was developed to measure the competency levels of PEM fellows within the ACGME core competencies, the reliability analyses in TABLE 2 include items contained within categories specified by the ACGME core competencies, as well as within the factors identified by exploratory factor analysis. Mean score comparisons were conducted based on the core competency categories.

Cronbach α for the entire final instrument was .989, indicating a high degree of internal consistency; a high degree of internal consistency within each category was also obtained (TABLE 2).

The mean scores within all competencies obtained by fellows from faculty increased with time spent in the fellowship; senior-level fellows received higher scores than junior fellows (online supplemental APPENDIX 2). The differences in mean scores obtained by fellows in each year

of training were statistically significant. Junior fellows were rated competent to proficient in all core competencies; senior fellows received ratings more in the proficient to expert range.

Self-evaluation scores increased with time spent in fellowship (online supplemental APPENDIX 3); senior fellows self-assessments were higher than the self-assessments of junior fellows were. The differences in mean scores between fellows in different years of fellowship were statistically significant only in the competencies of practice-based learning and improvement, patient care, interpersonal and communication skills, and professionalism. Average scores self-assessed by fellows were closer to competent for first-year fellows and closer to proficient for fellows in their final year.

Independent-sample *t* tests yielded statistically significant differences between faculty's evaluation and fellow self-evaluations within all competencies for fellows in their final of year of fellowship, in all competencies, with the exception of medical knowledge for second year, and in 4 competencies (practice-based learning and improvement, patient care, interpersonal and communication skills, and professionalism) for first-year fellows (TABLE 3). Fellows' self-assessments yielded lower scores than the corresponding faculty evaluations. Across all years, fellows were very unlikely to give themselves ratings in the expert range in any competency. Faculty evaluators were more

TABLE 3 MEAN SCORE COMPARISONS: COMPARISONS OF MEAN SCORES WITHIN EACH CORE COMPETENCY SELF-ASSESSED BY THE FELLOWS VERSUS SCORES GIVEN TO FELLOWS BY FACULTY

Core Competency	Mean Core Competency Scores								
	First-Year Fellow			Second-Year Fellow			Third-Year Fellow		
	Self	Faculty	P Value	Self	Faculty	P Value	Self	Faculty	P Value
Medical knowledge	3.40	3.49	.680	3.51	3.86	.166	3.82	4.42	.032
Practice-based learning and improvement	3.21	3.60	.001	3.25	3.78	.002	3.64	4.46	.002
Patient care	3.40	3.68	≤.001	3.57	3.95	≤.001	3.90	4.44	≤.001
Systems-based practice	3.33	3.53	.277	3.49	3.80	.023	3.75	4.32	.005
Interpersonal and communication skills	3.31	3.63	≤.001	3.48	3.74	≤.001	3.77	4.32	≤.001
Professionalism	3.47	3.49	≤.001	3.68	3.92	.027	3.87	4.42	≤.001

likely to give the highest-performing fellows expert ratings, whereas those same fellows rarely rated themselves higher than proficient on any item in any category.

Discussion

Our study provides evidence supporting the reliability and validity of a scoring rubric for PEM fellows to use to self-assess and for PEM faculty to evaluate fellows objectively, on all 6 ACGME core competencies. We determined content and construct validity of our instrument through extensive review by subject-matter experts and focus groups of key stakeholders, and secondly by exploratory factor analysis. Exploratory factor analysis yielded 5 factors corresponding closely to the ACGME core competencies. Items fitting descriptions for the competencies of medical knowledge and practice-based learning and improvement were in 1 factor; this result intuitively makes sense because resident physicians gain a large portion of knowledge through the process of learning to practice medicine. Likewise, items within the competencies of patient care and systems-based practice probably were included in the same factor because trainee physicians practice and learn patient care within a given system. An interesting result of the exploratory factor analysis was that items for interpersonal skills and communication skills were in separate factors. Whereas many relate to others using these skills in conjunction, not all individuals with good communication skills have good interpersonal skills, and vice versa. For our final instrument, we chose to keep all items classified within the 6 ACGME competencies. However, to improve instrument flow, we changed the order of how items appear in the instrument, such that items for practice-based learning and improvement followed medical knowledge, and items for systems-based practice followed patient care.

Another interesting finding in our study was the reclassification of certain items by exploratory factor analysis. Items in TABLE 1 initially categorized under professionalism, medical knowledge, and practice-based learning and improvement resulted in the reclassification of these items into factors corresponding to other competencies, demonstrating the overlapping nature of the ACGME competencies.

As a measure of reliability, our instrument showed a high degree of internal consistency with Cronbach α for the entire instrument and for each separate factor at or above the ACGME recommended 0.85.³⁶ Finally, as a measure of construct validity, our instrument showed ability to discriminate between the competence levels of fellows in advancing years of training. Fellows achieved improvements in faculty ratings in all core competencies as they advanced through their fellowship; this finding indicates that with the use of our scoring rubric, our more skilled and experienced fellows obtain higher ratings from faculty than our less-experienced fellows.

From the standpoint of feasibility, we had 93% participation from faculty and 94% participation from fellows in submitting evaluations during the study period. The only fellows who missed providing self-assessments during the study period were on vacation or leave of absence. Our evaluations currently are distributed via a noninstitution-dependent website-based system, and our fellows' completion rate is consistently 100%. Faculty participation is also slightly improved, although we have yet to attain 100% participation consistently. There were 61 items in the study instrument; it took our faculty about 20 minutes to complete 1 evaluation. Now that the instrument has been decreased to 53 items and the format is familiar to faculty, the average amount of time spent by faculty to complete 1 evaluation is 10–15 minutes.

We found that random assignment of fellows for faculty to evaluate did not consistently result in the collection of adequate numbers of evaluations for an individual fellow. Faculty were apt to suspend an evaluation if they were randomly assigned to evaluate a fellow with whom they had not worked; faculty suspension of evaluations sometimes resulted in the collection of only 1 to 2 faculty evaluations for a fellow in a quarter. We changed our program's evaluation assignment process to have each fellow identify at least 6 faculty members they have worked with in a given academic quarter; and at least 4 of those faculty members are randomly selected to provide evaluations.

Before using the study rubric, our previously used, 5-point, numeric, Likert-type rating scale failed to distinguish between experienced and less-experienced fellows. Generally, all fellows received ratings of 4 or 5 across all items, regardless of level of training, and if they received scores of 3 or less, fellows were unhappy with their scores; behavior-specific feedback was also lacking. After we began using our study rubric, fellows no longer became upset with 3 ratings because behaviors ascribed to 3 ratings were perceived by fellows as an acceptable level; if they received scores of ≤ 2 , fellows had specific behavioral descriptions available to make self-improvements. One of the unforeseen benefits of using the rubric instrument in our setting was that it improved the constructiveness of the comments faculty provided in the freehand comments section of the instrument. With the use of our study rubric, faculty began providing more constructive feedback for our fellows, with behavior-specific suggestions provided more often than when using the prior Likert-type rating form.

As TABLE 3 shows, fellows routinely provide self-assessment ratings that are lower in all competencies when compared with their faculty ratings. This phenomenon has been previously described in the literature. "Top performers have been found to underestimate their percentile rank relative to the people with whom they compare themselves... they tend to underestimate how their performance compares with that of others."^{37(p85)}

Our study has several limitations, including testing the instrument at 1 institution with 1 set of faculty and fellows, reducing generalizability. The form is also specific to the evaluation of PEM fellows and may not be valid for use in evaluating residents and fellows in other specialties. However, many of the items in our instrument are generic enough to be used or modified by other programs.

Conclusions

Development of assessment instruments for use in medical education is a time-consuming process that requires numerous steps to establish validity and reliability.

We have created a substantively and statistically validated rubric evaluation of PEM fellows that is a reliable tool for formative and summative evaluation. Although this instrument fits under the category of a *global rating* form.

As defined by the ACGME, we developed it to be used as just 1 part of a more comprehensive assessment system in our PEM program that incorporates multiple methods and evaluators. Use of the rubric assessment increased our ability to deliver more constant and reliable feedback that fellows are willing to hear and incorporate into making constructive changes, and we have observed a decrease in the number of fellows who are unhappy, dissatisfied, or disagree with their faculty assessments.

References

- Swing S. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med*. 2002;9(11):1278–1288.
- Lynch DC, Swing SR; Accreditation Council for Graduate Medical Education. Key considerations for selecting assessment instruments and implementing assessment systems. Available at: <http://www.acgme.org/outcome/assess/keyconsids.pdf>. Accessed April 9, 2008.
- Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Meloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med*. 2004;79(6):549–556.
- Davis J, Inamdar S, Stone R. Inter-rater agreement and predictive validity of faculty ratings of pediatric fellows. *J Med Educ*. 1986;61:901–905.
- Haber R, Avins A. Do ratings on the American Board of Internal Medicine Resident Evaluation Form detect differences in clinical competence? *J Gen Intern Med*. 1994;9:140–145.
- Thompson W, Lipkin M Jr, Gilbert D, Guzzo R, Roberson L. Evaluating evaluations: assessment of the American Board of Internal Medicine Resident Evaluation Form. *J Gen Intern Med*. 1990;5:214–217.
- Epstein RM, Hunder EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226–235.
- Pulito AR, Donnelly MB, Plymale M, Mentzer RM. What do faculty observe of medical students' clinical performance? *Teach Learn Med*. 2006;18(2):100–104.
- Siegel BS, Greenber LW. Effective evaluation of residency education: how do we know it when we see it? *Pediatrics*. 2000;105(4):964–965.
- Boateng BA, Bass LD, Blaszk RT, Farrar HC. The development of a competency-based assessment rubric to measure resident milestones. *J Grad Med Educ*. 2009;1(1):45–48.
- Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: a systematic review. *Acad Med*. 2009;84(3):301–309.
- Reisdorff E, Carlson D, Reeves M, Walker G, Hayes O, Reynolds B. Quantitative validation of a general competency composite assessment evaluation. *Acad Emerg Med*. 2004;11(8):881–884.
- Reisdorff E, Hayes O, Walker G, Carlson D. Evaluating systems-based practice in emergency medicine. *Acad Emerg Med*. 2002;9:1350–1354.
- Moskal B. Scoring rubrics: what, when, and how? *Pract Assess Res Eval*. 2000;7(3). Available at: <http://PAREonline.net/getvn.asp?v=7&n=3>. Accessed October 30, 2006.
- Andrade, H. Using rubrics to promote thinking and learning. *Educ Leadersh*. 2000;57(5):13–18.
- Musial J, Rubinfeld I, Parker A, et al. Developing a scoring rubric for resident research presentations: a pilot study. *J Surg Res*. 2007;142:304–307.
- Moni RW, Beswick E, Moni KB. Using student feedback to construct an assessment rubric for a concept map in physiology. *Adv Physiol Educ*. 2005;29:197–203.
- Moni RW, Moni KB. Student perceptions and use of an assessment rubric for a group concept map in physiology. *Adv Physiol Educ*. 2008;32:47–54.
- Brown MC, Conway J, Sorensen TD. Development and implementation of a scoring rubric for aseptic technique. *Am J Pharm Educ*. 2006;70(6):1–6.
- Nicholson P, Gillis S, Dunning AM. The use of scoring rubrics to determine clinical performance in the operating suite. *Nurse Educ Today*. 2009;29:73–82.
- Blood-Sieffried JE, Short NM, Rapp CG, et al. A rubric for improving the quality of online courses. *Int J Nurs Educ Scholarsh*. 2008;5(1):article 34.
- Lunney M, Sammarco A. Scoring rubric for grading students' participation in online discussions. *Comput Inform Nurs*. 2009; 27(1):26–31.
- Licari FW, Knight GW, Guenzel PJ. Designing evaluation forms to facilitate student learning. *J Dent Educ*. 2008;72(1):48–58.
- Isaacson JJ, Stacy AS. Rubrics for clinical evaluation: objectifying the subjective experience. *Nurs Educ Pract*. 2009;9:134–140.

- 25 O'Brien CE, Franks AM, Stowe CD. Multiple rubric-based assessments of student case presentations. *Am J Pharm Educ*. 2008;72(3):article 58.
- 26 Montgomery K. Classroom rubrics: systematizing what teachers do naturally. *Clearing House*. 2000;73:324–328.
- 27 Hanna M, Smith J. Using rubrics for documentation of clinical work supervision. *Couns Educ Superv*. 1998;37(4). Available at: <http://web.ebscohost.com.ezproxy.lib.uh.edu/ehost/detail?sid=1c18de34-9383-4b05-8735-c23a5b80383d%40sessionmgr7>. Accessed October 30, 2006.
- 28 Allen D, Tanner K. Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sci Educ*. 2006;5:197–2003.
- 29 Dreyfus HL, Dreyfus SE. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York, NY: Free Press; 1986.
- 30 Hatcher RL, Lassiter KD. *Report on Practicum Competencies*. Washington, DC, and Miami FL: The Association of Directors of Psychology Training Clinics (ADPTC) Practicum Competencies Workgroup. 2004;1:49–53. http://www.appic.org/downloads/Practicum_Competencie8C8F8.doc. Accessed: July 12, 2006.
- 31 Atherton JS. Doceo: Competence, Proficiency and beyond. 2003. <http://www.doceo.co.uk/background/expertise.htm>. Accessed July 12, 2006.
- 32 Accreditation Council of Graduate Medical Education Board. 2007. Common program requirements: general competencies. Available at: <http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>. Accessed March 5, 2008.
- 33 de Wet C, Spence W, Mash R, Johnson P, Bowie P. The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care*. 2010;10.1136/qshc.2008.031062. Accessed April 22, 2010.
- 34 Bierer SB, Hull AL. Examination of a clinical teaching effectiveness instrument used for summative faculty assessment. *Eval Health Profs*. 2007;30:339–361.
- 35 Haidet P, Kelly A, Chou C; Communication, Curriculum, Culture Study Group. Characterizing the patient-centeredness of hidden curricula in medical schools: development and validation of a new measure. *Acad Med*. 2005;80(1):44–50.
- 36 Lynch DC, Swing SR. Key considerations for selecting assessment instruments and implementing assessment systems. Accreditation Council for Graduate Medical Education. <http://www.acgme.org/outcome/assess/keyconsids.pdf>. Accessed April 9, 2008.
- 37 Dunning D, Johnson K, Ehrlinger J, Kruger J. Why people fail to recognize their own incompetence. *Curr Dir Psychol Sci*. 2003;12(3):83–87.