Journal of Contemporary Medical Education

available at www.scopemed.org

Original Research

Accuracy of faculty evaluations of residents' medical knowledge: comparison of subjective summative evaluations to in-training examination scores

Matthew Stenerson[1], Amy Weaver[2], Jason Homme[1]

[1]Department of Pediatric and Adolescent Medicine and [2]Health Sciences Research
Mayo Clinic, Rochester, MN, USA

Received: March 13, 2013

Accepted: May 08, 2013

Published Online: July 15, 2013

DOI: 10.5455/jcme.20130508094057

Corresponding Author:
Jason Homme, Mayo Clinic, 200 First Street, SW, Rochester, MN 55905
homme.jason@mayo.edu

Keywords: *Evaluation/assessment of clinical performance, Faculty development, Graduate Medical education*

ABSTRACT

Although faculty evaluations are an integral component in the assessment of residents' medical knowledge, there is conflicting evidence about the accuracy of such evaluations. In contrast, the in-training examination (ITE) is an objective measure of residents' medical knowledge that has proven to predict future performance. This study investigates the accuracy of faculty evaluations by comparing scores of medical knowledge on faculty completed evaluations to in-training examination scores. Subjects included categorical pediatrics and pediatric neurology residents at Mayo Clinic from 2006-2010. For each year of training the standard ITE score was compared to the corresponding composite faculty evaluation score for medical knowledge and patient care competencies using the Spearman's rank correlation coefficient. Forty-two residents were analyzed for the first year of training and 35 for the second. For both years, there was only a mild correlation between ITE score and either the medical knowledge (r=0.26 and 0.19, respectively) or patient care competency (r=0.27 and 0.29, respectively). Faculty assessments of residents' medical knowledge did not correlate well with a standard objective measure. Training programs should consider targeted faculty development to improve evaluation skills and supplementing existing evaluation methods with additional assessment tools.

INTRODUCTION

Faculty evaluations have long been an integral component of resident competency assessment in graduate medical education training programs. In 1999, the Accreditation Council for Graduate Medical Education (ACGME) defined six core competencies of physician training: patient care; medical knowledge; practice-based learning and improvement; interpersonal and communication skills; professionalism; and systems-based practice [1]. These competencies serve as a framework for educating and evaluating residents and fellows. Although maximizing residents' medical knowledge is essential to producing clinically savvy physicians, there is conflicting evidence regarding the ability of faculty evaluators to accurately assess this competency [2-13].

Each specialty board offers an in-training examination (ITE) administered yearly. This examination is composed of 200 speciality specific multi-choice questions and is administered at a designated time of year in a secure, proctored manner. Performance on the ITE has demonstrated predictive value for the likelihood of passing the board certification examination on the first attempt [13-22]. There is also evidence that performance on standardized measures of medical knowledge is predictive of future performance on objective measures of quality clinical care [23]. Thus, the in-training examination (ITE) offers an objective measure of medical knowledge against which the accuracy of faculty evaluations can be assessed.

This study seeks to assess the accuracy of faculty evaluations by comparing residents' evaluation-derived medical knowledge scores to ITE scores. Performance scores related to the patient care core competency (which measures faculty assessment of direct patient care activities by trainees) were also reviewed as a comparison measure. This competency was chosen, as a comparison measure, because patient care activities are commonly those most closely observed by faculty, the quality of these activities relate significantly to medical knowledge, and this core competency is one of the best understood by faculty, thus our contention that the assessments may track together. The other core competencies were not included in this study because of the perception that they did not have as direct a connection with the medical knowledge domain. Our null hypothesis was that the faculty scores would not correlate with the ITE scores.

METHODS

Subjects for this retrospective record review study included all categorical pediatric and pediatric neurology residents at Mayo Clinic between 2006-2010 who completed ITEs after each year of training and were evaluated by faculty on all core ACGME competencies after each clinical rotation.

The Pediatric and Adolescent Medicine Residency Program at Mayo Clinic trains both categorical pediatric residents and pediatric neurology residents each year. The categorical residents train for three years and sit for three ITEs, while the pediatric neurology residents sit for two ITEs during the pediatric-specific part of their training. ITEs are administered in July for every year of training.

Residents spend four to five weeks on each clinical rotation and have direct, interactive contact with faculty for approximately ten to 15 hours per week. Much of this contact is spent either at the bedside on patient-centered rounds or in the outpatient clinic setting. During a typical rotation, a faculty member may spend one to two weeks with a particular resident. At the end of the rotation, each faculty member is given an opportunity to complete an online evaluation form on each resident with whom sufficient time was spent (sufficiency left to faculty discretion). The end of rotation evaluation forms are 15 questions long and are intended to be comprehensive in scope (global rating scale). Faculty is encouraged to complete the entire form but may skip questions if they determine insufficient exposure to provide a meaningful assessment. Each question is scored on a scale of one (needs improvement) to five (top ten percent of residents) and questions are each anchored to one of the six ACGME core competencies (Figure 1). A composite score for each competency is computed by the online evaluation program (Integrated Scheduling and Evaluation System, a product of the Mayo Clinic's Education Technology Center). This composite score is in fact the mean score for the competency domain and is calculated by dividing the sum of all individual evaluation questions addressing the particular competency by the total number of evaluations. For this study the faculty scores for all clinical rotations for a resident in a particular year of training were averaged to derive a final composite faculty evaluation score for the two competencies of interest, medical knowledge and patient care.

Standard scores for each resident's ITE taken at the beginning of the second and third (if applicable) years of training were recorded. (The ITE taken at the beginning of the first year of residency training was not factored into this analysis, as it reflects only knowledge attained during medical school.)

The Spearman's rank correlation coefficient was calculated between ITE score and medical knowledge and patient care composite evaluation scores for each year of training. A two-sided p-value testing whether the correlation was significantly different from zero was also calculated.

RESULTS

Forty-two out of 44 residents who completed the first year of training during the study period met inclusion criteria and were included in the analysis, compared to 35 out of 43 residents for the second year. The mean standard ITE score after the first year of training was 372 (range 90-630), compared to 422 (range 160-630) after the second year. Composite medical knowledge scores for the first and second year of training were 3.90 (range 3.32-4.42) and 4.03 (range 3.60-4.38), respectively. Composite patient care scores for the first and second year of training were 3.92 (range 3.17-4.43) and 4.09 (range 3.52-4.41), respectively (Figure 2). The mean number of faculty evaluations that were utilized to determine the medical knowledge composite scores were 44.1 (range 15-98) and 35.8 (range 12-80) for the first and second year of training respectively. For the patient care composite scores the mean number of faculty evaluations were 105.5 (range 32-190) and 73.3 (range 16-133) for the first and second year of training respectively (Figure 3).

Results for the correlation between each year's ITE score and the medical knowledge and patient care competencies are shown in Table 1. As noted in the table, there was a positive trend but none of the correlations analyzed achieved statistical significance.

1. (Patient Care) Accuracy and completeness of gathered information: Patient history, Records, Physical exam						
	1	2	3	4	5	
2. (Interpersonal & Communication Skills) Effective and concise case presentations						
	1	2	3	4	5	
3. (Medical Knowledge) Interprets patient data effectively : Patient history, Physical exam, Laboratory data						
	1	2	3	4	5	
4. (Systems-Based Practice) Selection of appropriate diagnostic tests: Incorporates consideration of cost awareness and risk-benefit analysis in patient care						
	1	2	3	4	5	
5. (Patient Care) Formulation of an effective plan and management strategy for his/her patients: Makes therapeutic decisions and formulates and carries out an effective plan and management strategy for his/her patients						
	1	2	3	4	5	
6. (Practice-Based Learning) Commitment to his/her own education: Use of information technology to access, appraise, and apply knowledge to patient care; Self-directed learning; Conference attendance; Asked appropriate questions						
	1	2	3	4	5	
7. (Practice-Based Learning) Organizes time to balance both teaching and care giving: Start and finish rounds on time, Minimized delays and interruptions						
	1	2	3	4	5	
8. (Medical Knowledge) Incorporated learning topics as outlined by the curriculum						
	1	2	3	4	5	
9. (Professionalism) Provided excellent teaching on a regular basis: Case-based, Didactic, Integrated teaching into rounds						
	1	2	3	4	5	
10. (Practice-Based Learning) Asked learners to discuss differential diagnosis on most patients.						
	1	2	3	4	5	
11. (Practice-Based Learning) Asked team members to discuss alternative management options for patients nursing, peers, faculty, and other services						
	1	2	3	4	5	
12. (Interpersonal & Communication Skills) Effective communications with all members of the health care team:						
	1	2	3	4	5	
13. (Professionalism) Level of integrity: Honesty, citizenship, trustworthiness, and reliability						
	1	2	3	4	5	
14. (Interpersonal & Communication Skills) Effective communication with families and patients: Incorporation of patient preferences in clinical decision making						
	1	2	3	4	5	
15. (Systems-Based Practice) Using the health care system to optimize the care of the patient: Advocates for quality patient care and uses the health care system to optimize the care of the patient						
	1	2	3	4	5	
Comments:						
Rotation Grade:						

Figure 1. End of rotation competency based summative evaluation form.

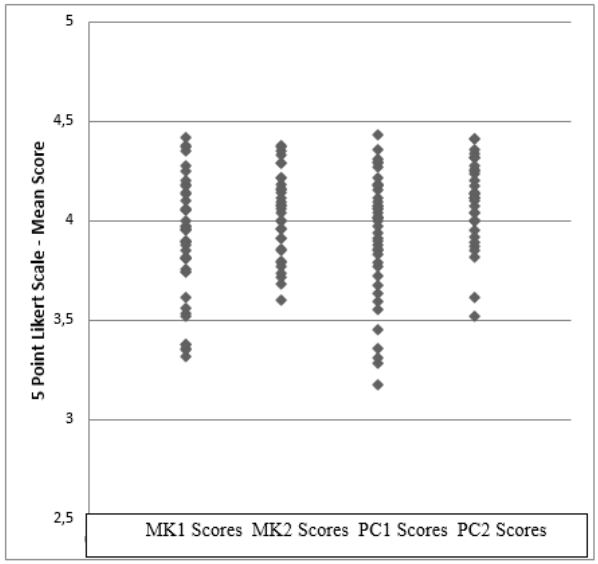

Figure 2. Distribution of composite scores for medical knowledge and patient care.

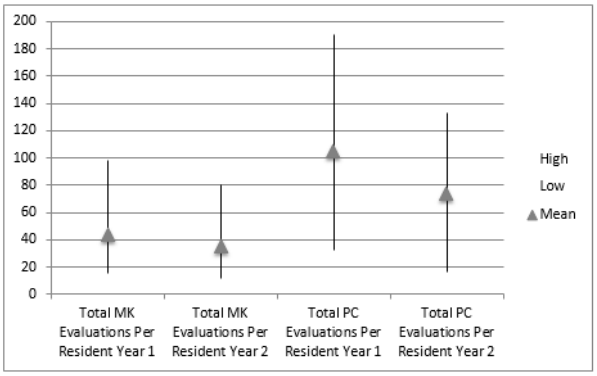

Figure 3: Range of number of evaluations to produce composite score.

Table 1. Correlation between ite score and medical knowledge and patient care competencies.

Correlation Between Core Competency Score and ITE	
Medical Knowledge	
PGY-1	0.26 (p = 0.10)
PGY-2	0.19 (p = 0.28)
Patient Care	
PGY-1	0.27 (p = 0.08)
PGY-2	0.29 (p = 0.10)

ITE = In-training examination, PGY = Post-graduate year

DISCUSSION

Our findings indicate that as a group, faculty members' subjective assessment of medical knowledge demonstrate a poor but positive correlation with a standard objective measure, the ITE score, and does not reach statistical significance. The same is true for assessment of the patient care competency.

With the exception of the study published by Kolars et al. in 2003, each of the previous studies comparing faculty evaluations to ITE scores were performed prior to the ACGME's 1999 statement defining the six core competencies of physician training [2-13]. The majority of the studies indicated a poor correlation between the two measures [2-8,13], with only Davis et al. and Kastner et al. reporting a strong correlation – although the Kastner study reported a strong correlation only on evaluations by general inpatient faculty, not for outpatient or ICU-based faculty [9,10]. (Of note, Risucci et al. and Buckwalter et al. also found a positive correlation between evaluation and ITE score, but evaluations in these studies were based on overall performance, not specifically medical knowledge [11,12]). Our study, being the first in the ACGME-core-competency era to compare ITE scores to faculty evaluations across an entire year of diverse rotations, suggests that despite efforts to improve the evaluation process, faculty's subjective evaluations still fail to correspond to objective markers of a residents' global medical knowledge.

Aside from its small sample size, the principal limitation of our study is that not all faculty members filled out evaluations on residents whom they supervised (non-participation), thus leaving their input out of the analysis (unfortunately we do not have a way to quantify precise amount of non-participation in this study population). However, over the span of one to two years of training, in which many clinical rotations are repeated (e.g. three months of neonatal intensive care in the first year), it is believed that the non-participation would likely affect all residents similarly and thus not result in any major statistical anomalies.

The discrepant assessment of medical knowledge seen with faculty evaluations creates two possible types of error, each with concerning results for the trainee. The first is underestimation of skill – where the trainee may dismiss the evaluation as inaccurate because they know their objective measure of medical knowledge (ITE) reflects a higher level of achievement. The second is an overestimation of skill – with missed opportunities for improvement for the trainee. In both scenarios, inaccurate faculty assessments may result in a detriment to residents' future efforts to improve their medical knowledge. Without accurate input into their deficiencies from supervising physicians, trainees are

forced to navigate the complex field of medicine like a sailor with a faulty compass.

Why faculty members' assessments of residents' medical knowledge lack correlation with the objective measures of knowledge is unclear, although several factors may contribute. For example, in a study of family medicine preceptors, Taylor and Lipsky found that the seven faculty who rated lowest in their ability to predict residents' cognitive levels had been teaching for significantly fewer years than the seven best faculty predictors, suggesting a strong role for faculty experience [3]. In addition, in their systematic review of resident evaluations since 1999, Lurie et al. found that none of the published evaluation tools could effectively distinguish one core competency from another, suggesting that faculty may score residents based on a gestalt sense of overall clinical performance, rather than by each specific competency [1]. Although further research is needed to better distinguish the factors that make certain faculty more or less adept at evaluating residents' medical knowledge, it is likely that all evaluators would benefit from ongoing faculty development targeted at improving skills in evaluating residents' competencies. Targeted faculty development has been demonstrated to improve the specificity of feedback, the ability to assess competency-based and level-specific learning objectives, and can produce lasting changes in faculty evaluation behaviors [24-26]. In turn, with more accurate evaluations to guide further study, trainees would benefit as well.

It also deserves note that the traditional rotation-based faculty evaluations are not the only means of assessing resident competency. Indeed, widely used as a competency assessment tool in medical school, the observed structured clinical exam (OSCE) has also shown promise in residency programs as a means of assessing a number of the core competencies [27]. Although standardizing such examinations may be challenging in clinically demanding residency programs, the additional opportunity for direct faculty observation may be quite valuable to identify areas for improvement. For medical knowledge in particular, end-of-rotation examinations may also provide a timely assessment upon which further study can be based. Although also challenging to accomplish in the day-to-day practice of clinical medicine, standardizing such examinations would provide an objective assessment of residents' knowledge for each rotation where there currently exists only subjectivity.

In summary, our study indicates that despite efforts by the ACGME to improve the resident evaluation process, faculty as a group are not accurate evaluators of residents' medical knowledge. While further research is needed to better assess the reasons for the discrepancy between evaluations and ITE scores, it is likely that structured faculty development programs would benefit all evaluators and thus, the trainee. In addition, augmenting clinical rotation-based evaluations with other subjective (e.g. OSCE) and objective (e.g. end-of-rotation examinations) assessment tools may provide residents with additional input into their deficiencies to help guide further study. Further research is needed to evaluate the costs and benefits of such a multi-faceted approach in the context of clinically demanding training programs.

PRACTICE POINTS

- Faculty evaluations are important to assessing resident competence.
- Faculty evaluations are commonly subjective in nature.
- Objective measures of medical knowledge such as specialty specific in-training-evaluations exist and predict future performance.
- Correlation between subjective faculty evaluations and objective measures is often lacking as demonstrated by this study.
- Targeted faculty development may improve accuracy of evaluation methods.

Declaration of Interest

The authors report no conflicts of interest including no financial relationships relevant to this article to disclose.

Clinical Trial Registration Number

N/A.

Grants/Fellowships

None.

Notes on Contributors

Matthew Stenerson, M.D., is currently a third year pediatric endocrinology fellow at Stanford University in Palo Alto, California. The research included in this manuscript occurred during his pediatric residency at Mayo Clinic in Rochester, Minnesota. Prior to residency Doctor Stenerson graduated from the (University of California San Francisco) UCSF School of Medicine.

Amy Weaver, M.S., is a biostatistician in the Division of Biomedical Statistics and Informations in the Department of Health Sciences Research at Mayo Clinic in Rochester, Minnesota. She has provided statistical and study design support for numerous

research studies including large population-based projects.

Jason Homme, M.D., is an Assistant Professor of Pediatrics at the Mayo Clinic College of Medicine and the current Residency Program Director for pediatrics in the Mayo School of Graduate Medical Education in Rochester, Minnesota.

REFERENCES

1. 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 Mar;84(3):301-09.
2. Kolars JC, McDonald FS, Subhiyah RG, Edson RS. Knowledge base evaluation of medicine residents on the gastroenterology service: Implications for competency assessments by faculty. Clin Gastroenterol Hepatol. 2003 Jan;1(1):64-68.
3. Taylor C, Lipsky MS. A study of the ability of physician faculty members to predict resident performance. Fam Med 1990;22:296-98.
4. Hawkins RE, Sumption KF, Gaglione MM, Holmboe ES. The in-training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty predictions of resident performance. Am J Med. 1999 Feb;106(2):206-10.
5. Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluation, OSCE, and ABSITE as measures of surgical intern performance. Am J Surg. 1995;169:414-17.
6. Brailovsky CA, Grand'Maison P, Lescop J. Residency directors' predictions of candidates' performance on a licensing examination. Acad Med. 1997;70:410-14.
7. Lazar HL, DeLand EC, Tompkins RK. Clinical performance versus in-training examinations as measures of surgical competence. Surgery 1980;87:357-62.
8. Nattress LW. Orthopaedic In-Training Examination. J Med Ed. 1969;44:878-89.
9. Davis JK, Inamdar S, Stone RK. Interrater agreement and predictive validity of faculty ratings of pediatric residents. J Med Educ. 1986 Nov;61(11):901-05.
10. Kastner L, Gore E, Novack AH. Pediatric residents' attitudes and cognitive knowledge, and faculty ratings. J Pediatr. 1984 Jun;104(6):814-18.
11. Risucci DA, Tortolani AJ, Ward RF. Ratings of surgical residents by self, supervisors and peers. Surg Gynecol Obstet. 1989;169:519-26.
12. Buckwalter JA, Schumacher R, Albright JP, Cooper RR. The validity of orthopaedic in-training examination scores. J Bone Joint Surg Am. 1981 Jul;63(6):1001-06.
13. Borlase BC, Bartle EJ, Moore EE. Does the in-service training examination correlate with clinical performance in surgery? Curr Surg. 1985 Jul-Aug;42(4):290-92.
14. Garvin PJ, Kaminski DL. Significance of the in-training examination in a surgical residency program. Surgery. 1984 Jul;96(1):109-13.
15. Shetler PL. Observations on the American Board of Surgery In-Training examination, board results, and conference attendance. Am J Surg. 1982 Sep;144(3):292-94.
16. Grosman RS, Fincher RM, Layne RD, Seelig CB, Berkowitz LR, Levine MA. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. J Gen Intern Med. 1992;7(1):63-67.
17. Cantwell JD. The Mendoza Line and in-training examination scores. Ann Intern Med. 1993;119(6):541.
18. Waxman H, Braunstein G, Dantzker D, Goldberg S, Lefrak S, Lichstein E, Ratzan K, Schiffman F. Performance on the internal medicine second-year residency in-training examination predicts the outcome of the ABIM certifying examination. J Gen Intern Med. 1994;9(12):692-94.
19. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. Am J Med. Aug 2007;120(8):735-40.
20. Rollins LK, Martindale JR, Edmond M, Manser T, Scheld WM. Predicting pass rates on the American Board of Internal Medicine certifying examination. J Gen Intern Med. 1998;13(6):414-16.
21. Leigh TM, Johnson TP, Pisacano NJ. Predictive validity of the American Board of Family Practice In-Training Examination. Acad Med. 1990;65:454-57.
22. Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. J Pediatr. 2008 Sep;153(3):425-28.
23. Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailovsky C. Association between licensure examination scores and practice in primary care. JAMA. 2002;288:3019-26.
24. Regan-Smith M, Hirschmann K, Iobst W. Direct observation of faculty with feedback: an effective means of improving patient-centered and learner-centered teaching skills. Teaching and Learning in Medicine. 19(3):278-86.
25. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence. A randomized trial. Ann Intern Med. 2004;140:874-81.
26. Narayan AP, Whicker SA, McGann KA. An innovative process for faculty development in residency training. Teaching and Learning in Medicine. 2012;24(3),248-56.
27. Jain SS, DeLisa JA, Eyles MA, et al: Further experience in development of an objective structured clinical examination for physical medicine and rehabilitation residents. Am J Phys Med Rehabil 1998;77:306-10.