How Accurate Are Faculty Evaluations of Clinical Competence?

JEROME E. HERBERS, JR., MD, MAJ, MC, USA, GORDON L. NOEL, MD, GLINDA S. COOPER, MS, JOAN HARVEY, MD, CDR, MC, USN, LOUIS N. PANGARO, MD, LTC, MC, USA, MICHAEL J. WEAVER, MD, COL, MC, USA

Objective: *To determine the degree and sources of variability in faculty evaluations of residents for the American Board of Internal Medicine (ABIM) Clinical Evaluation Exercise (CEX).*

Design: *Videotaped simulated CEX containing programmed resident strengths and weaknesses shown to faculty evaluators, with responses elicited using the open-ended form recommended by the ABIM followed by detailed questionnaires.*

Setting: *University hospital.*

Participants: *Thirty-two full-time faculty internists.*

Intervention: *After the open-ended form was completed and collected, faculty members rated the resident's performance on a five-point scale and rated the importance of various aspects of the history and physical examination for the patient shown.*

Measurements and Main Results: *Very few of the resident's strengths and weaknesses were mentioned on the open-ended form, although responses to specific questions revealed that faculty members actually had observed many errors and some strengths that they had failed to document. Faculty members also displayed wide variance in the global assessment of the resident: 50% rated him marginal, 25% failed him, and 25% rated him satisfactory. Only for performance areas not directly related to the patient's problems could substantial variability be explained by disagreement on standards.*

Conclusions: *Faculty internists vary markedly in their observations of a resident and document little. To be useful for resident feedback and evaluation, exercises such as the CEX may need to use more specific and detailed forms to document strengths and weaknesses, and faculty evaluators probably need to be trained as observers.*

Key words: *clinical competence; internship and residency; internal medicine; certification.* J GEN INTERN MED 1989; 4:202–208.

INTERNAL MEDICINE RESIDENCY PROGRAM directors and their evaluation committees bear primary responsibility for ensuring that trainees become clinically competent. Although attainment of uniform standards for all of the more than 440 residencies in the United States and Canada is extremely difficult, the American Board of Internal Medicine (ABIM) has attempted to foster uniformity by providing definitions of essential skills and expectations of performance for certified internists.[1] Among the strategies suggested by the ABIM for the assessment of clinical competence is the clinical evaluation exercise (CEX). It is included in the evaluation process in over 70 % of Internal Medicine residencies.[2] The ABIM recommends that experienced physicians administer the CEX to all residents during the first six months of training; many programs administer the CEX annually throughout the residency. Faculty findings are used to determine the level of competence of each resident as well as to provide feedback; CEX results are one component of the assessment of clinical competence which program directors submit to the ABIM.

Unfortunately, the CEX, like the oral examinations abandoned by the ABIM in 1972, has several shortcomings. First, evaluators may vary considerably in their abilities to discern strengths and weaknesses in residents, and they may apply different standards when judging a resident's performance.[5, 6] Second, evaluators may be positively or negatively influenced in their assessments of residents because of expectations or biases.[3] Finally, a resident's performance may vary considerably from patient to patient and from encounter to encounter,[4] so that a single appraisal of a resident's performance is unlikely to represent his overall clinical competence accurately.

In spite of these shortcomings, direct observation of a resident by a faculty member continues to offer important opportunities for feedback and instruction. Its importance is based in part on the traditional value placed by physicians and society on direct observation and guidance of physicians-in-training. The opportunities the CEX affords for assessment of interpersonal skills as well as overall clinical judgment are not fully shared by other measures of clinical competence, which focus on knowledge base, interpretative and problem-solving abilities, specific history-taking, or procedural skills. But if direct observation of a resident by a faculty member during a history and physical examination is to provide truly useful guidance, then the accuracy of that observation and its thorough documentation must be assured.

In this study, we sought to elucidate the sources of faculty variability as a necessary step toward achieving

Received from the Fellowship Program in General Internal Medicine, Departments of Medicine, Walter Reed Army Medical Center, Washington, D.C., and the Uniformed Services University of the Health Sciences, Bethesda, Maryland.

Presented at the Tenth Annual Meeting of the Society of General Internal Medicine, April 30, 1987, San Diego, California.

Supported in part by the Department of Clinical Investigation, Walter Reed Army Medical Center.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

Address correspondence and reprint requests to Dr. Herbers: Geneal Medicine Service, Department of Medicine, Walter Reed Army Medical Center, Washington, DC 20307-5001.

TABLE 1
Programmed Strengths and Weaknesses in Resident's Performance

	History	Physical Examination
Strengths	Good use of open-ended questions Good alcohol history Good rheumatic fever history Good tobacco use counseling	Good organization and control Good exam for peripheral edema Good exam of peripheral pulses
Weaknesses	Ignored psychosocial cues/clues Drug compliance not assessed No family history Incomplete marital history No sexual history No sleep or dietary history	Cardiac exam inadequate to assess possible valvular disease Poor thyroid exam Incomplete exam of lymph nodes

reliability in resident evaluations. We eliminated resident and patient variability by showing a group of faculty members the same resident – patient encounter in a videotaped simulation. We sought to determine the extent of faculty variability and to elucidate how much of that variability arises from each of several sources: 1) differences in how well individual faculty members discern strengths and weaknesses in a given resident; 2) failure of some faculty members to document important strengths and weaknesses which are in fact observed; and 3) differences in standards of performance among evaluators.

METHODS

Subjects

Forty-five staff internists at Walter Reed Army Medical Center who had been selected by the internal medicine residency program director to conduct clinical evaluation exercises in the spring of 1986 were asked to participate in the study. Thirty-two of these physicians (29 male, 3 female) volunteered. All were full-time salaried staff members whose duties included teaching, research, patient care, and administration, and all gave written informed consent. The study protocol was approved by the Walter Reed Investigational Review Board.

Videotaped Simulated Clinical Evaluation Exercise

A 60-minute simulated clinical evaluation exercise (CEX) was produced in a professional television studio. Included in the simulation were a complete history and physical examination, a brief presentation by the resident to a staff physician, and a final discussion of problems and plans between the resident and the patient. The "patient" was an actor trained to portray a middle-aged man with congestive heart failure and angina following a myocardial infarction; he was also depressed and noncompliant with his medical regimen. In addition, his history included a work-related shoulder injury, a childhood illness suggestive of rheumatic fever, and a history of a cardiac murmur heard in adulthood. The physician who portrayed the resident was in fact a board-certified internist, appearing to be in his late twenties, who was unknown to the study subjects. He performed a complete and generally competent history and physical examination, except for several specific errors of omission and of commission that were programmed into the simulation (Table 1). With the exception of these programmed errors, his performance was designed to be typical of an average finishing resident; in some areas his performance was unusually good. The camera was positioned so that subjects viewing the videotape would have the same perspective a staff physician would have during an actual CEX, except that multiple camera angles and close-up shots were also used to demonstrate nonverbal communication and pertinent parts of the physical examination.

Questionnaires

The 1985 ABIM clinical evaluation exercise (CEX) form (Questionnaire 1) was used.[7] The form included six sections: History, Physical Examination, Judgment and Synthesis, Medical Care, Humanistic Attributes, and Overall Clinical Competence. Each section contained a brief summary describing the skills to be evaluated in the exercise and provided a space for comments. Each section also included a four-category rating scale with the following response options: unsatisfactory, marginal, satisfactory, and superior. The form concluded with a request for an estimate of overall clinical competence using the same scale.

Two additional questionnaires (Questionnaires 2 and 3) were developed to determine whether areas of resident performance not described by faculty members on Questionnaire 1 were simply not observed, were observed but not felt to be important enough to comment on, or were not commented on despite being observed and considered to be important; these questionnaires also included questions to assess the level of agreement among evaluators on what should have been included in the history and physical examination and how the resident's time should have been allocated.

Questionnaires 2 and 3 consisted of structured questions with responses limited to a specific five-point scale developed for each section. An additional response "can't remember" was also available.

Questionnaire 2 was divided into two parts. In the first, 45 questions were asked about various aspects of the observed history (e.g., "How well did the resident explore the patient's compliance with the prescribed drug regimen?") and physical examination (e.g., "How well did the resident perform cardiac auscultation?"). The second part of Questionnaire 2 related to the resident's allocation of time to 15 areas in the history and physical examination (e.g., "How would you characterize the time or emphasis the resident expended on the neurological exam?").

Questionnaire 3 provided five brief summaries of important data from the patient's history; after each summary, subjects were asked to rate the importance in this patient of specific items of history and physical examination. For example, one series of questions was preceded by the following text: "This patient is being treated for heart failure, hypertension, and angina. Recently, his symptoms appear to have worsened. In your opinion, how important is each of the following in this patient?" Among the items that followed were, "a detailed review of how he takes his medicine" and "detailed information on his diet." Another group of questions followed the text, "The patient has a history of a murmur heard in childhood and now has symptoms of heart disease. In your opinion, how important is each of the following in examining this patient?" Among the seven items which followed were, "inquiring about recent dental work" and "listening for the murmur of mitral stenosis."

Subjects were also asked questions about their professional backgrounds, current activities, attitudes about their work, and experience as CEX evaluators.

Administration of Simulated CEX

The simulated CEX was shown in a quiet room with distractions kept to a minimum. Subjects first completed the personal characteristics questionnaire and reviewed the ABIM CEX form (Questionnaire 1); they were asked not to discuss the videotape among themselves. They were instructed to evaluate the resident shown on the videotape in the same way as they would evaluate a resident in their own program and to take notes as desired. They were initially shown the history portion of the simulated CEX and asked to complete the first part of Questionnaire 1 ("History"). They were then shown the physical examination portion and requested to complete the corresponding part of the questionnaire. Thereafter the remainder of the videotape (presentation by the resident to the faculty member and closing discussion with the patient) was shown, and subjects were allowed to complete the rest of Questionnaire 1 and to revise their earlier comments and ratings.

After collection of Questionnaire 1, the other questionnaires were sequentially distributed and collected. Unlimited time was allowed for completion of each questionnaire. The total time expenditure for the session was two hours or less.

Statistical Methods

Agreement among subjects on each item of resident performance on Questionnaires 2 and 3 was evaluated as the average of the agreement beyond chance between each pair of subjects using the kappa statistic,[8] with agreement weighted according to closeness of subjects' responses (weights of 1.0, 0.75, 0.50, 0.25, and 0 were used for agreement within one, two, three, four, and five categories, respectively). These kappas were used only for comparison of items within this study, since criteria suggested for the interpretation of unweighted kappas may not apply.

RESULTS

Subject Characteristics

Among the 32 subjects, eight of the internal medicine subspecialties were represented. Only two subjects had not had subspecialty training. Subjects had attended 28 medical schools, including two foreign medical schools, and had completed internal medicine residency training in 16 different programs; all were certified by the ABIM. Twenty-seven of the subjects had completed internal medicine residencies within the preceding 11 years. Twenty-five subjects indicated that most of their professional time was allocated to clinical practice and teaching; seven described their time as being primarily devoted to research and/or administration. Nine subjects described providing comprehensive internal medicine care to at least 60% of their patients, whereas half (16 of 32) reported providing such care for less than 30% of their patients. Six subjects reported never having served as a CEX evaluator and 11 had served only once; only seven had been a CEX evaluator more than three times previously. Six subjects described themselves as never having been evaluated doing a complete history and physical examination and 13 reported having been evaluated only once or twice.

Responses on ABIM Form (Questionnaire 1)

Written comments on the open-ended form suggested by the ABIM for the CEX (Questionnaire 1) ranged in length from a few words to several paragraphs. Most comments were brief and general rather than specific. The numbers of subjects writing specific comments differed greatly among performance areas

TABLE 2

Evaluation of Resident's Performance by 32 Subjects: Questionnaire 1 (Open-ended) and Questionnaire 2 (Structured)

Performance	Questionnaire 1* No. (%)	Questionnaire 2 Unacceptable → outstanding† 1	2	3	4	5	Kappa‡
Poor compliance history	5 (16)	8	14	9	0	0	0.50
Poor attention to concerns	19 (59)	6	10	13	3	0	0.39
Poor psychiatric history	30 (94)	26	6	0	0	0	0.81
Good open-ended questions	1 (3)	2	6	18	5	1	0.46
Poor thyroid exam	4 (13)	6	4	18	1	1	0.40
Poor lymph node exam	8 (25)	12	13	6	0	0	0.40
Poor cardiac exam	17 (53)	9	9	14	0	0	0.44
Poor abdominal exam	6 (19)	1	4	24	2	0	0.72

*Number of evaluators writing comments on open-ended ABIM form.

†Rating scale for faculty assessment of resident's performance: 1, unacceptable, glaringly in error; 2, barely adequate, room for much improvement; 3, average, typical of a good resident; 4, done very well, better than most residents; 5, outstanding, nearly perfect. Table does not include "can't remember" responses, so totals may not equal 32 for every item.

‡Agreement beyond chance for all pairs of observers, weighted according to degree of agreement.

(Table 2). Thirty of the 32 subjects correctly described the resident's attention to psychosocial issues as inadequate. However, only seven (22%) noted that the resident obtained no information on sexual functioning, despite hints of marital dissatisfaction. Further, only five subjects made any note that the resident failed to ask the patient about compliance with his medical regimen, even though the medical regimen was complex, and the patient seemed unsure of why each medicine had been prescribed, was unsure of his dosing schedule, and expressed concern over whether some of his symptoms might be caused by his medications. More subjects (19 of 32) noted that the resident ignored several of the patient's concerns and questions, such as his concern over possible medication side-effects and questions about the basis for his worsening symptoms of congestive heart failure. Little comment was made about areas in which the resident performed well, such as the resident's consistent attempts to optimize communication by asking frequent open-ended questions, using facilitative gestures and phrases, and attempting to maintain eye contact despite the patient's detached, depressed affect.

In the physical examination of this patient with current symptoms of congestive heart failure, a childhood history strongly suggestive of rheumatic fever, and a history of a cardiac murmur, the resident failed to examine the heart with the patient sitting up or in the left lateral decubitus position. Only 17 of 32 subjects wrote any comment describing the cardiac exam as poor or incomplete. Although the resident palpated the neck several inches above the thyroid gland, an error clearly shown in a close-up view, only four subjects commented on this deficiency. Similarly, errors in the physical examination of the lymph nodes and abdomen were mentioned infrequently on the ABIM form (Table 2).

Responses on Questionnaire 2

After subjects had completed and turned in Questionnaire 1, they answered 45 questions about specific areas of resident performance. Although subjects generally responded correctly more frequently than on Questionnaire 1, considerable variability persisted. Responses to eight key questions are shown in Table 2. As on the open-ended form, the greatest agreement was on how well the resident asked about psychiatric symptoms (kappa = 0.81), with all 32 subjects now describing this area as unacceptable or barely adequate. Marked improvement in recorded recognition occurred for the resident's failure to inquire about drug compliance; although only five subjects wrote comments about this deficiency on the open-ended form, when asked a direct question 22 indicated they recognized that this area was poorly handled. However, nine subjects still described the resident's performance in this area as adequate. Substantial variability was evident in responses to numerous other questions about the history. Although six subjects felt the resident made *above-average* use of open-ended questions, only one subject had recorded this finding on Questionnaire 1; furthermore, eight others disagreed, stating that this aspect of the history-taking was *below average* (kappa = 0.46). Six subjects indicated that the resident was above average in his eye contact with the patient whereas ten felt he was deficient in this area (kappa = 0.41).

For the physical examination, there was essentially no change in the frequency of appropriate responses regarding the cardiac exam; nearly half continued to rate an incomplete exam as acceptable. Sixteen subjects felt the neurologic exam was average or better, but 16 also felt it was poor (kappa = 0.42). The least agreement on Questionnaire 2 regarded how well the

TABLE 3

Responses of 32 Subjects to Selected Questions about Standards of Performance for the Simulated CEX (Questionnaire 3)

	Rating of Item's Importance*					
	None ⟶ critical					
Performance Area	1	2	3	4	5	Kappa†
Detailed review: how patient takes his medications	0	0	2	10	20	0.63
Information about marital satisfaction	0	0	3	23	6	0.70
Percussion of the diaphragm	5	7	15	4	1	0.34
Listening for egophony	5	5	14	5	3	0.23
Percussion of the left heart border	3	9	10	9	1	0.29
Listening for mitral stenosis	0	0	0	9	23	0.75

*Rating scale for importance: 1, not important, should be left out; 2, of little importance, could be included or left out without harm; 3, important, but not likely to be critical; 4, very important, must have information for good care; 5, of critical importance, not to know might be harmful.

†Agreement beyond chance for all pairs of observers, weighted according to degree of agreement.

thyroid exam was done (kappa = 0.40); two subjects felt that this completely inadequate exam was very good or outstanding, 18 felt it was adequate, 10 felt it was unacceptable or barely adequate, and two couldn't remember.

Questionnaire 2 also contained 15 questions on how the resident allocated his time among various parts of the history and physical exam. The resident spent several minutes inquiring about the patient's understanding of the hazards of cigarette smoking and offering simple advice on quitting smoking. Disagreement was extreme: two subjects felt the amount of time expended by the resident in this area was excessive and unacceptable, two felt it was excessive but not a major problem, seven felt it was too little but not a major problem, and six felt it was insufficient and unacceptable (kappa = 0.30). Whereas 15 subjects described time spent on the cardiac exam as unacceptable, 12 felt it was just right, in spite of the fact that major portions were not done (kappa = 0.40). Marked disagreement was also noted in the evaluation of the resident's performance of a detailed screening neurologic exam: six subjects rated the time allocation as excessive, nine rated it as insufficient, and 17 rated it as just right (kappa = 0.31).

Responses on Questionnaire 3

In the third questionnaire, subjects were given brief descriptions to remind them of certain parts of the patient's history and then asked 35 detailed questions about what they felt should have been done to evaluate this patient. Subjects showed substantial agreement in some areas but little agreement in others (Table 3). For example, almost all subjects felt it was very important to ascertain information about the patient's marital satisfaction, to carefully review compliance with medications, and to specifically listen for the murmurs of mitral stenosis and aortic insufficiency, in spite of the fact that most had not commented on the programmed resident errors in these areas on Questionnaire 1. In other areas, however, subjects disagreed substantially, with some believing certain behaviors to be very important and others indicating those same behaviors to be of little importance. For example, subjects disagreed markedly on the importance of questions about possible endocarditis and about which components of the pulmonary and neurologic examinations were essential.

Assessments of Overall Clinical Competence

On Questionnaire 1, five of the 31 subjects responding rated the resident's overall clinical competence "unacceptable," 18 rated it "marginal," and eight rated it "satisfactory." The final question on Questionnaire 2 asked for a second assessment of overall clinical competence. Seven subjects now rated the resident's performance "unacceptable," 17 rated it "marginal," and eight rated it "satisfactory." At the end of the final questionnaire, subjects were once again asked to assess overall clinical competence. The distributions of assessments remained essentially unchanged when compared with Questionnaires 1 and 2.

DISCUSSION

The observation by experienced clinicians of medical students and residents interviewing and examining patients is a fundamental component of clinical teaching. However, the usefulness of these evaluations, whether done informally on rounds or in clinics or formally as in the ABIM Clinical Competence Exercise (CEX), depends on the accuracy of the evaluator. Although recent studies have raised concerns about day-to-day variability in a resident's performance and the problems of adjusting for differences in the complexity of patients,[4, 9] this study provides evidence that variability among faculty evaluators is likely to be at least as important. The major source of variability appeared to be the differences among faculty members in how much they record when provided an open-ended instrument, such as that recommended by the ABIM. Espe-

cially striking was the overall paucity of specific comments that would be useful for either evaluation or feedback. Regarding several important errors in the physical examination, most subjects made no comment. For areas in which the resident performed very well, even less was written; for example, only one subject mentioned that the resident demonstrated good interviewing skills.

The extent to which faculty evaluators discern important findings in a CEX, even if they fail to record them on an open-ended form, was evaluated by providing subjects with a structured questionnaire asking for ratings of specific areas of the history and physical examination. In many cases responses using this directed format revealed that subjects did indeed recognize important findings in the simulated CEX that they had failed to document on the ABIM form. For example, whereas only 16% noted on the open-ended questionnaire that the resident failed to determine the patient's compliance with his medical regimen, 69% indicated on Questionnaire 2 that this area was unacceptable or barely adequate. However, substantial variability among evaluators persisted in responses on Questionnaire 2, and in some areas there was no evidence that some subjects ever recognized important errors. For example, nearly half of the subjects did not recognize a completely inadequate cardiac examination.

To determine whether differences in subjects' standards accounted for failure to correctly respond on both open-ended and directed questionnaires (Questionnaires 1 and 2), we administered a third questionnaire that sought subjects' views on what *should* have been done by the resident in interviewing and examining this patient. For those areas of the history and physical examination that related directly to the patient's major problems, subjects were almost always in agreement. In particular, all 32 subjects felt a careful cardiac examination for mitral stenosis and aortic insufficiency to be very important, and almost all indicated that questions about marital satisfaction and medication compliance were essential. However, there was very little agreement regarding aspects of the physical examination that did not relate directly to the patient's complaints, such as several parts of the pulmonary and neurologic examinations (Table 3).

The evaluators who were the subjects of this study were all full-time faculty internists in a university hospital and almost all were subspecialists. Whether our results are indicative of the performances of evaluators in other settings is of possible concern. Subsequent to the work described here, we have used these methods with experienced evaluators at other universities and in workshops at national meetings. Among academic general internists and subspecialists alike, results have been strikingly similar: individual evaluators vary substantially in the acuity of their observations, the degree to which they document their findings, and their standards for resident performance.

Evaluator variability limits the effectiveness of any method of performance assessment, not just faculty observation of a resident's clinical skills.[3, 8, 10] The reliability of physicians as evaluators of trainees, however, has received little attention.[11] Orkin and Greenhow presented 34 faculty members with 27 simulated evaluations of anesthesiology residents, each containing ratings based on six criteria.[12] There was no consensus on how the different criteria should be weighed in making an overall assessment of clinical competence. Wigton asked internal medicine residents, volunteer faculty members, and full-time faculty members to rank performance dimensions according to their importance in the evaluation of first-year house officers.[13] Full-time and volunteer faculty differed significantly on the importance of the quality of the written record and on the importance of skills in the critical evaluation of data. Woolliscroft and coworkers showed 20 faculty internists two short videotapes and asked them whether specific physical examination maneuvers were performed; four or more disagreed on whether a maneuver was performed for five of 18 items in the cardiovascular exam.[5] Although Butzin and coworkers showed impressive levels of agreement among examiners for the American Board of Pediatrics oral examination,[14] they considered only global assessments of competence. In contrast to these studies, we examined faculty skills as observers of an *entire* CEX and discovered poor faculty reliability in discerning and recording many important details of the resident's performance.

In this study some variability was undoubtedly due to differences among subjects in the degree to which they formed a fixed global assessment of the resident that diminished their ability to accurately rate specific performance areas—the so-called halo effect.[3, 15] A possible negative halo effect might explain the tendency of subjects to fail to comment on good interviewing skills if they had already judged the resident negatively because psychosocial issues were not attended to with appropriate specific questions. Further, since the resident was unknown to the subjects, some may have been less reluctant to give negative ratings. In most actual CEX settings, faculty members are familiar with the residents being evaluated, and halo error is felt to skew positively assessments of clinical competence.[6] The problem of halo error illustrates the need in any method of direct observation to delineate the degree and sources of observer variability.

It is likely that the accurate assessment of a resident's clinical competence requires multiple approaches.[16] Stillman et al. have demonstrated the usefulness of standardized patients in the evaluation of history and physical examination skills.[4] However, reliable estimates of communication and diagnostic skills were projected to require at least ten resident–patient encounters; therefore, for these components of clinical competence standardized patients may be impractical. Because of its feasibility and capacity to encompass

several aspects of performance, including problem-solving, the CEX remains a potentially valuable component of clinical competence assessment, especially if it is repeated several times early in residency training. However, our results strongly suggest that untrained evaluators are too variable and document too little for such exercises to achieve their full potential either as a formative evaluation or to establish a resident's clinical competence. These findings do, however, suggest several steps toward improving faculty performance as evaluators.

One approach entails alterations in the evaluation forms provided. Our results point out the limitations of open-ended rating forms as they are currently used. Although specific questionnaires tailored to the problems of each patient are not practical, questions could easily be added to the current CEX form to require assessment by evaluators of interviewing and physical examination skills needed for all patients. With the ABIM we have initiated a controlled multisite trial to test the effects of a brief training intervention and a more directive evaluation form on the quality of faculty assessments. If proven effective, such changes could make the CEX and other observations of students' and residents' interviewing and physical examination skills more valuable in the assessment of clinical competence.

The authors thank Dr. Ronald Sen of the Bethesda Naval Hospital for portraying the resident in the simulated CEX.

REFERENCES

1. American Board of Internal Medicine. Evaluation of clinical competence. Portland, OR: American Board of Internal Medicine, 1986.
2. Blank LL, Grosso LJ, Benson JA. A survey of clinical skills evaluation practices in internal medicine residency programs. J Med Educ. 1984; 59:401-6.
3. Cooper WH. Ubiquitous halo. Psychol Bull. 1981; 90:218-44.
4. Stillman PL, Swanson DB, Smee S, et al. Assessing clinical skills of residents with standardized patients. Ann Intern Med. 1986; 105:762-71.
5. Woolliscroft JO, Stross JK, Silva J. Clinical competence certification: A critical appraisal. J Med Educ. 1984; 59:799-805.
6. Kroboth FJ, Kapoor W, Brown FH, Karpf M, Levey GS. A comparative trial of the clinical evaluation exercise. Arch Intern Med. 1985; 145:1121-3.
7. American Board of Internal Medicine. Evaluation of clinical competence. Portland, OR: American Board of Internal Medicine, 1985.
8. Feinstein, A. Clinical epidemiology: the architecture of clinical research. Philadelphia: W. B. Saunders, 1985; 184-6, 638-9.
9. Moulopoulos SD, Stamatelopoulos S, Nanas S, Economides K. Medical education and experience affecting intra-observer variability. Med Educ. 1986; 20:133-5.
10. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown, 1985:22-32.
11. Wakefield J. Direct observation. In: Neufeld VR, Norman GR, eds. Assessing clinical competence. New York: Springer, 1985:51-70.
12. Orkin FK, Greenhow DE. A study of decision making: how faculty define competence. Anesthesiology. 1978; 48:267-71.
13. Wigton RS. Factors important in the evaluation of clinical performance of internal medicine residents. J Med Educ. 1980; 55:206-8.
14. Butzin DW, Finberg L, Brownlee RC, Guerin RO. A study of the reliability of the grading process used in the American Board of Pediatrics oral examination. J Med Educ. 1982; 57:944-6.
15. Borman WC. Consistency of rating accuracy and rating errors in the judgment of human performance. Organ Behav Hum Perform. 1977; 20:238-52.
16. Neufeld VR. Implications for education. In: Neufeld VR, Norman GR, eds. Assessing clinical competence. New York: Springer, 1985; 301-5.