

## PERSPECTIVES

## Improving In-Training Evaluation Programs

Jeffrey Turnbull, MD, Jean Gray, MD, John MacFadyen, MD

**I**n-training evaluation (ITE) is the process of observing and systematically documenting the ongoing performance of a learner in real clinical settings during a specific period of training. Clinical performance (what the learner does) is in part a consequence of competency (what the learner knows) but is also influenced by other variables.<sup>1-3</sup> In-training evaluation may include the assessment of competencies, such as knowledge or the acquisition and development of skills. It also includes the assessment of performance behaviors resulting from these competencies.<sup>4-8</sup> Behaviors that should develop and mature during clinical training, such as professionalism, are most appropriately evaluated during clinical interactions.<sup>9</sup>

Objective competency-based assessment methods should be considered as part of any system of in-training evaluation and include written examinations, objective structured clinical examinations (OSCEs) using standardized patients, formal or structured oral examinations, and chart-stimulated recall assessment.<sup>10</sup> Performance-based methods of in-training assessment include the introduction of standardized patients into daily clinical activity, videotaped patient encounters, chart review, log books, and rating forms used to document observed behaviors. Most techniques used in evaluating residents are quantitative in nature, but qualitative evaluations (such as field notes) are also of value.<sup>10</sup>

Although objective evaluation methods must be a part of the overall ITE, we focus on issues of *performance assessment* with special emphasis on the residency years. The ongoing assessment of residents (ITE) as they participate in patient care is the best opportunity to evaluate many essential practice behaviors and to provide meaningful formative and summative information. Present forms of assessment often fail to provide sufficient information to facilitate resident learning and assist training programs to make promotion decisions.

Current approaches used in resident performance assessment rely heavily on input from attending staff to a program director who integrates all of this material and creates a final document.<sup>11,12</sup> Increasingly, however, the use of other observers, particularly nurses,<sup>1,13,14</sup> has been suggested to be of value, particularly in the assessment of communication skills, ethical behaviors, reliability, and integrity.<sup>15-17</sup> Peer and resident assessments remain primarily research tools at the present but will likely be useful in future evaluation.<sup>3,18</sup> Patients have also been shown to provide valuable information pertaining to resident performance.<sup>16</sup> Self-evaluation has not been extensively studied, but the development of such skills has been proposed as a form of guidance and support to residents.<sup>19</sup>

### THE ROLES OF IN-TRAINING EVALUATION

The purpose of ITE is to provide an accurate measurement of the learner's performance abilities. Program directors and ultimately the public must know that competent physicians are entering practice. Effective programs of ITE must also provide feedback to the trainee and to the program to enhance further learning and modifications in training.<sup>20-22</sup> Systems of ITE must be based on clear learning objectives as these objectives will motivate students and direct learning. Programs evaluate what they perceive to be important. Although learners may have unique learning approaches, the learning style they ultimately utilize is a function of multiple variables. Prominent among these is the evaluation process.<sup>23</sup>

Systems of ITE must be perceived as having validity and reliability—that is, as evaluating essential competencies necessary for future practice in a reproducible fashion utilizing methods that are feasible, defensible, and accountable.<sup>24,25</sup> This information can be used for purposes such as licensure, the acquisition and maintenance of hospital privileges, annual promotion, and certification or recertification.

### PROBLEMS WITH EXISTING SYSTEMS

As a method of evaluation, ITE has been widely criticized. Traditional ITEs lack validity in that they assess a restricted range of competencies and do not appear to consider many of the essential skills necessary for future practice. Tools, such as resident rating scales, have been found to have limited interrater and intrarater reliability

---

Received from the Department of Medicine, University of Ottawa, Ont., Canada (JT, JM), and the Departments of Medicine and Pharmacology, Dalhousie University, Halifax, N.S., Canada (JG).

Address correspondence and reprint requests to Dr. Turnbull: University of Ottawa, 451 Smyth Rd., Ottawa, Ont. K1H 8M5, Canada.

and thus do not discriminate between residents or performances (i.e., they lack reproducibility).<sup>7,20,25,26</sup> Such weaknesses arise from two major factors. First, faculty must serve as both teacher and evaluator but receive little training to assume this dual role. The lack of training may result in observations that are not fully accountable or legally defensible.<sup>27</sup> Rater errors include those of *distribution* such as the leniency/severity error (doves/hawks) and the central tendency error (failure to use the entire rating scale),<sup>20</sup> and those of *correlation* (commonly called the halo effect) in which assessments reflect the tendency of a rater to be influenced by some information or impression about a resident and to allow that impression to spill over into the judgment of other components of the resident's performance.<sup>20</sup> Faculty training can help to reduce these errors.

Second, even when practice behaviors are observed on a daily basis, they frequently are not documented. Consequently, evaluations are not done in a timely fashion, and the opportunity for the systematic objective evaluation of performance behaviors is lost, and with it, the opportunity for meaningful summative assessment and feedback to the learner. Observations that are necessary for meaningful evaluations do require considerable faculty time. Evaluations that take place long after the event, and are based on faculty recall of rotational performance, are not sufficiently detailed to be helpful for decision making.

### ADVANTAGES OF RECORDING PERFORMANCE ASSESSMENT

Daily or frequent resident performance can and should be observed, evaluated, and documented.<sup>20</sup> This form of evaluation captures information not available by other measurement techniques. For the resident, ongoing evaluation and feedback on an individual patient basis, pertaining to essential external performance criteria, encourages the internalization of performance standards and is a prerequisite for learning.<sup>21,28</sup> For the faculty, individual patient-based evaluation permits multiple observations of actual behaviors in the clinical setting.

Important educational initiatives over the last decade based on the evolving needs and expectations of society have led to the identification of new roles and responsibilities for physicians. Roles such as the communicator, scholar, advocate, collaborator, manager, and professional and their associated competencies are most appropriately assessed in the context of ITE.<sup>3,23,29,30</sup> Frequent in-training assessment permits the opportunity for the assessment of history taking and physical examination skills, as well as communication skills as trainees relate to colleagues and patients in a variety of practice milieus.<sup>31</sup> Newer definitions of competency include essential roles such as diagnostic reasoning skills, as well as patient and practice management skills including health promotion, advocacy, collaboration, counseling, and psy-

chomotor skills, which can be best assessed in usual clinical settings. Critical appraisal, continued learning, and teaching skills are also evaluated most appropriately in this setting. Finally, professional and interpersonal behaviors can be best evaluated through in-training assessment as residents relate to patients, allied health professionals, and colleagues.<sup>1,13-17</sup>

### AGENDA FOR DEVELOPING IMPROVED PROGRAMS

General principles underlying the formulation of any comprehensive program for ITE are presented below. These have arisen from broader evaluation concepts and will serve as a guide to the development of future programs for ITE. The following case description is utilized to illustrate these principles.

*John is a second-year medical resident who has recently received in the mail his rotational evaluation after having completed his cardiology rotation 2 months ago. His supervisor, in filling out the ITE form 6 weeks after the completion of the rotation, felt that John was an average resident who was knowledgeable and who got along well with the staff and patients alike. With this in mind, he gave John a very good mark and turned to the item checklist to justify this overall score.*

#### 1. Reliability

Reliability pertains to the reproducibility or consistency of results produced by an evaluation method. Enhanced reliability results from the reduction of variance introduced by random error and the improved detection of true differences in performance. Observations must be consistent or reproducible and distinguish between different levels of ability. To achieve this, faculty should be encouraged to observe performances that occur on a regular daily basis and document these in a standardized fashion. Global evaluations done several weeks after the completion of a rotation will not provide John or his supervisors with reliable or useful feedback. Performance on a single test item, be it a multiple choice question or OSCE station, only loosely predicts performance in another domain. Multiple assessments of a single domain using multiple items or cases are necessary to improve the reliability of assessments and ensure that results are generalizable (e.g., not limited by content specificity).<sup>25,26</sup>

Reliable in-training assessment must document meaningful differences (between and within individual performances) in order to provide formative feedback to the trainee and summative information to the program. Faculty development initiatives focusing on items on ITE reports may lead to modestly improved interrater and intrarater reliability<sup>32</sup> and discrimination. The reported demonstration of high correlations of the ITE form and a global rating scale, however, suggests that increasing or modifying the items on the forms is not likely to increase reliability. Thus, the best way to improve reproducibility is to increase the number of ratings.

Although objectivity is a fundamental component of reliability, subjective evaluations raise special concerns for clinical supervisors. Objective assessment requires that observations must be unbiased by factors other than resident performance. To accomplish this, the documentation of performance must be standardized between observers or raters and based on observed behaviors (e.g., tardiness) and not on inferences (e.g., disrespect).<sup>24,25</sup> John's cardiology supervisor tends to give all residents a very good evaluation, and this score is influenced by a multitude of factors independent of John's clinical performance (the halo effect described earlier). Important behaviors reflecting clinical performance must be the unit of evaluation to permit discrimination.

Increasing the number of observations by a single faculty supervisor, or using observations by staff, colleagues, and allied health professionals will improve the reliability of in-training assessment as well as its validity.<sup>25,30,33</sup> A profile of John's performance could be constructed with input from allied health professionals, peers, and patients.

## 2. Validity

Validity pertains to whether a test measures what it is intended to measure, that is, the appropriateness of generalizations made about the resident based on the observations. In-training assessment methods should focus on the observation of actual performance (what the resident does) as opposed to competence (what the resident knows or is capable of doing). Although not always the case, the closer an evaluation mimics reality, the more valid or realistic it becomes. Although the training environment is an artificial situation, it does reflect the practice environment, and all assessment efforts should be directed toward the application of what the learner knows or is capable of doing in actual practice circumstances.<sup>9,24,34</sup>

In addition to being a measure of John's socialization skills, his ITE should evaluate all relevant domains of clinical practice in cardiology as outlined in the training objectives of the internal medicine program and predict future practice performance.

## 3. Flexibility

The assessment of resident performance must be flexible so that the complete spectrum of clinical practice competency can be evaluated in multiple circumstances. Whatever system is developed, it should be applicable to the inpatient, ambulatory, and community environments. John's evaluation provides little useful information on his skills and capabilities in a variety of practice environments, such as being on call in the coronary care unit or seeing patients in the cardiology clinic, and should be altered to reflect these.

## 4. Comprehensiveness

All relevant objectives and the corresponding resident performance must be assessed and documented through ITE. All the essential competencies necessary for future practice that can only be evaluated by the observation of ongoing performance must be evaluated through ITE. If communication, collaboration, professionalism, and advocacy are important, programs for ITE must define the corresponding behaviors that can be observed and document these systematically. The form used in John's program probably did not encompass these various competencies.

## 5. Feasibility

Recognizing the demands placed on human and physical resources within an academic environment, any assessment strategy requiring additional time and money is unlikely to be acceptable. Programs for ITE must measure behaviors without additional effort. New and expensive evaluation technologies are neither feasible nor necessary for most circumstances. Existing strategies for in-training assessment are not cost-effective in terms of time and resources.<sup>22</sup> To be successful, John's ITE must be effective, must not incur additional cost, and must be part of the daily practice of medicine.

## 6. Timeliness

To be effective, direct observation of in-training performance must occur as close to the behavior as possible (not 2 months later, as in John's case) to ensure that the observations are accurate and student learning is maximized. Individual clinical interactions must be the focus of evaluation, not the rotation itself. If documentation is delayed, evaluation is less effective as a learning tool, more subject to bias, and less defensible.<sup>22</sup>

## 7. Accountability

In-training assessment must be accountable to the program, its objectives, and ultimately the community within which the resident will work. Evaluation must be defensible and capable of providing a justifiable analysis or explanation of results.<sup>27</sup> John's evaluation failed to provide such explanations. Programs must ensure that performances are being evaluated in a fair and transparent fashion in keeping with the principles of natural justice.

With the increased emphasis on competencies that can be evaluated only in the setting of ITE, programs must accept responsibility for maintaining their societal contract. Programs must be seen to be maintaining the public trust.<sup>27</sup>

## 8. Relevance

In-training assessment of residents must be seen as important by both residents and faculty. Evaluations, both positive and negative on specific items or global ratings, must be acted upon, and performance assessment must influence promotion decisions and learning. Negative evaluations must lead to remediation, delayed promotion, or possibly expulsion, while positive evaluations must favorably affect advancement. A 2-month delay in providing feedback without any specific recommendations does not permit remedial or corrective action by the resident or by the program. If not viewed as important by John or his cardiology supervisor, ITE will not be effective as a summative tool, and its other functions, which include the facilitation of learning, will be lost.<sup>22</sup>

### HOW TO EVALUATE

*In-training evaluation must be carried out at the time of the observed behavior in a standardized fashion that does not significantly increase the workload placed on the faculty.* Performance evaluation must be based on observed behaviors on a per-patient basis, rather than on more global impressions of performance across a rotation or period of time.<sup>9,20</sup> Objective evaluations of resident performance must be part of the care of an individual patient. For example, opportunities to document behaviors could include rating scales incorporating duplicate sheets of the history and physical examination, the consultation note, or the reverse side of a billing card or daily patient lists. These are a part of the routine practice of medicine and can be used to document important aspects of resident performance in an ongoing standardized fashion. Observed technical skills can be documented through the use of on-line computer or dictating systems. Computer-based evaluation software and optical scanners can facilitate the collection of performance data and their subsequent analysis and presentation. Behaviors take place and are observed daily. If they are not documented, the information is lost.

### AVAILABLE METHODS OF EVALUATION

*Multiple assessment methods must be utilized to evaluate all essential competencies.* The most commonly used method for in-training assessment is the rotational resident rating scale.<sup>20</sup> Items should be behaviorally based with a clear description of potential responses. Clinical supervisors generally rate two overall constructs: professional and interpersonal behaviors, and clinical competency.<sup>18</sup> To improve feasibility and compliance with these forms, the total number of items should be commensurate with adequate reliability and validity. Other objective measures of performance should complement the resident rating scale. The structured assessment of case presentations and written medical notes could be used to evaluate objectively many aspects of competence.<sup>35</sup>

Formal oral examinations, OSCEs, or written examinations could complement these methods if resources permit, assuming that they too are subject to scrutiny in terms of validity and reliability assessment.<sup>10</sup>

### WHO SHOULD EVALUATE

*Effective programs of in-training assessment should move beyond one clinical supervisor evaluating resident performance to multiple observers from different aspects of the health care system.* In addition to one clinical supervisor, other clinical faculty, allied health professionals, patients, colleagues, and the residents themselves can provide meaningful objective performance information.<sup>1,13-17</sup> This can be incorporated into the day-to-day activities of patient care by asking questions of patients on discharge pertaining to resident performance, and by including the standardized evaluation of residents by allied health professionals in multidisciplinary team rounds. The focus of resident evaluation should not be on the basis of a rotation but on the care of an individual patient, which can be observed from many different perspectives.<sup>20</sup>

The bias introduced by evaluators who are reluctant to provide direct individual negative feedback can be minimized by evaluating practice behaviors, not personal traits, utilizing multiple observers and an evaluation committee. Input should be sought widely, and decisions should be made by the evaluation committee.

Faculty development and the feedback of evaluations from colleagues for comparison may also improve reliability.<sup>25,31,33</sup> These faculty development initiatives should focus on the agreement of objectives and assessment methods and standards. They should also include ways to maximize learning through meaningful feedback.

### WHEN TO EVALUATE

*For evaluation to be meaningful, it should significantly influence a resident's progress.* Similar performance criteria throughout all residency rotations are not necessary nor advisable. For some rotations, perhaps only specific skills should be assessed, such as procedural techniques, as well as a few general objectives, such as professional behaviors and the evaluation of progression with training. For certain rotations, especially those that occur close to the time of important promotion decisions (at the beginning of training, when moving from junior to senior resident, and at the completion of training), perhaps a more in-depth in-training assessment system with observed performance assessment, and possibly additional objective competency measures could be developed. This would allow resources and attention to be focused at critical points during training. By tailoring in-training assessment systems, specific competencies can be measured in accordance with program objectives with appropriate support and faculty development.

## WHAT COMPETENCIES TO ASSESS

There are many essential program objectives that can only be measured through in-training assessment. These must be determined and their evaluation made explicit. It is essential that rotational objectives include the necessary competencies for the future practice of all physicians, such as professional behaviors, communication skills, ethical decision-making skills, collaboration skills, patient management skills, and problem-solving skills. Although easily defined in more general terms, the characterization of more specific behavioral objectives will be essential but difficult. Faculty development programs will allow supervisors to agree on what is important and how it is to be evaluated.

## AN EXAMPLE

Recognizing the difficulties with existing systems of ITE, a pilot project was undertaken during the medical clerkship at the University of Ottawa based on the principles of timely and objective documentation of behaviors utilizing multiple observers in multiple circumstances.

Student performance information was gathered from attending physicians, allied health professionals, and patients utilizing a brief observational checklist (4 to 7 items) relevant to the observer and the behavior evaluated. Physician supervisors completed two assessments, the first at the time of the initial encounter with the patient (reflecting the student's clinical assessment abilities), and the second at the time of the discharge of that same patient (reflecting the student's management abilities). Members of the multidisciplinary team kept notes at the nursing station, and these were discussed and formally documented at the time of the multidisciplinary team rounds. Patients at discharge were asked to comment on the competencies of their attending medical student.

Collection of these data was made feasible by incorporating data collection into the usual day-to-day activities on the general medicine ward. The history and physical examination form was modified so that the final page (where the case is summarized by the student and comments are provided by faculty) has a noncarbon copy with an appended 5-item rating scale. At the discharge of the patient, staff are required to complete a billing card, the reverse of which is an 8-item rating scale. To make data collection and collation easier, these forms as well as those used by the multidisciplinary team and the patient are all computer scannable, and drop-off boxes were strategically placed on the ward.

Students receive immediate feedback pertaining to their performance and at the time of the completion of the rotation. A final summary sheet is computer-generated outlining the number of evaluations from the different sources provided, the student's mean mark pertaining to the items evaluated, the range, and the performance of the student's peers.

Utilizing this method, individual behaviors are assessed as they occur by multiple observers in a standardized fashion. On average, each student received 17 evaluations over the course of a 1-month clerkship rotation in internal medicine, and the corresponding reliability and validity studies are under way.

Although this is an example of utilizing the stated principles for the assessment of clinical clerks on a ward-based general medicine rotation, they could be applied to other circumstances. Surgical performance could be evaluated immediately after surgery using a telephone dictation system. An ambulatory care rotation could be evaluated by a brief scannable questionnaire and a rating scale of several items at the bottom of the patient's billing sheet. Consultation skills could be assessed through a computer scannable noncarbon copy of the consultation sheet that would have a brief rating scale. These assessment tools could be used to track a resident's performance over the course of a 4-year residency.

The fundamental principle of this approach is that the unit of evaluation is no longer a rotation completed at some distant date but an observed performance rated in an objective fashion at the time. Fewer items are completed in a greater number of circumstances to facilitate compliance.

## AGENDA FOR FUTURE RESEARCH

The development and implementation of improved strategies for in-training assessment will not require new evaluation methodologies, but rather the implementation of existing methods in a more consistent and objective fashion. To demonstrate efficacy, new programs of in-training assessment must be both reliable and feasible. A generalizability study utilizing *objective evaluations by multiple observers at the time of the behavior*, such as in our pilot project, will be an important adjunct to the literature. Standard setting is also a problem for in-training assessment as trainees of different levels and backgrounds perform similar tasks with differing expectations. Faculty must also consider not only an individual's performance, but also whether improvement is occurring in keeping with the expectations of the program. Future research must also consider the relative contribution of different evaluators and how many evaluations are required to provide a reliable estimate of resident performance and whether ITE correlates with future practice performance as a measure of predictive validity.

## SUMMARY

For faculties of medicine and training colleges to develop reliable, valid, and accountable systems of in-training assessment (Table 1), there must be a major change in the way we view this process. New systems of residency performance assessment that accurately document the characteristics necessary for the future practice of medi-

**Table 1. Criteria for Effective In-Training Evaluation****Reliability**

1. Are evaluations based on individual patient interactions?
2. Are there sufficient observations recorded for each resident?
3. Are there important differences detected between resident performances (differences between residents, or at different times with one resident at different stages of learning)?
4. Are differences in keeping with other available evaluation methods?
5. Are resident performances documented in a standard reproducible fashion?
6. Are observations based on performance behaviors rather than character judgments?
7. Is there a faculty development program in place to encourage consistency?

**Validity**

8. Does the ITE process measure behaviors that reflect relevant and important program- and rotation-specific performance objectives?
9. Are these completed by the appropriate people (clinical supervisors, peers, allied health professionals, patients)?

**Feasibility**

10. Do faculty believe that, although time-consuming, the ITE process is nevertheless worthwhile?
11. What measure is there of the compliance with such a program?
12. Are the principles of ITE and the program itself capable of being applied in multiple settings, and do they assess multiple levels of learning and learning objectives?

**Comprehensiveness**

13. Is the ITE process assessing all relevant objectives that can be evaluated through this system? (Is it complete to the extent that important objectives are measured through ITE where appropriate?)

**Timeliness**

14. Are evaluations completed at the time of the interaction to minimize bias and to maximize student learning?

**Accountability**

15. Are the principles of natural justice adhered to (i.e., timely evaluation based on clearly stated objectives with opportunity for remediation)?

**Relevance**

16. Do meaningful actions take place based on the ITE, from the perspective of the program (summative), and from the perspective of the resident (summative and formative)?

cine must make evaluations an ongoing part of day-to-day practice. The solution to the problem of ineffective in-training assessment is not the further revision of existing forms, but the restructuring of how we evaluate our trainees and reorient our faculty development and reward systems to recognize this essential part of our practice. An improved system for in-training assessment will not only provide an accountable system for the identification of those individuals who are in difficulty, but also encourage and facilitate learning for all residents in accordance with the objectives of the training program.

*As part of the In-Training project, this work was supported by the Educating Future Physicians of Ontario (EFPO) project and the Royal College of Physicians and Surgeons of Canada.*

**REFERENCES**

1. Butterfield PS, Mazzaferri EL, Sachs LS. Nurses as evaluators of the humanistic behavior of internal medicine residents. *J Med Educ.* 1987;62:842-9.
2. Dauphinée WD. Assessing clinical performance: where do we stand and what might we expect? *JAMA.* 1995;274(9):741-3. Editorial.
3. Nash DB, Markson LE, Howell S, Hildreth EA. Evaluating the competence of physicians in practice: from peer review to performance assessment. *Acad Med.* 1993;68(2):S19-23.
4. Hull AL, Hodder S, Berger B, et al. Validity of three clinical performance assessments of internal medicine clerks. *Acad Med.* 1995;70(6):517-22.
5. Jansen JJM, Tan LHC, van der Vleuten CPM, van Luijk SJ, Reithans JJ, Grol RPTM. Assessment of competence in technical clinical skills of general practitioners. *Med Educ.* 1995;29:247-53.
6. Kane MT. The assessment of professional competence. *Evaluating Health Professionals.* 1992;15:163-82.
7. Littlefield JH, DaRosa DA, Anderson KD, Bell RM, Nicholas GG, Wolfson PJ. Assessing performance in clerkships: accuracy of surgery clerkship performance raters. *Acad Med.* 1991;66(9):S16-8.
8. Miller GR. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65:S63-7.
9. Phelan S. Evaluation of the non-cognitive professional traits of medical students. *Acad Med.* 1993;68(10):799-803.
10. Gray JD. Primer on resident evaluation. *Ann R Coll Phys Surg.* 1996;29(2):91-94.
11. Parenti CM, Harris I. Faculty evaluation of student performance: a step towards improving the process. *Med Teacher.* 1992;14(2/3):185-8.
12. Resnick R, Taylor B, Maudsley R, et al. In-training evaluation—it's more than just a form! *Ann R Coll Phys Surg Can.* 1991;24:415-20.
13. Butterfield PS, Mazzaferri EL. New rating form for use by nurses in assessing residents humanistic behavior. *J Gen Intern Med.* 1991;6:155-61.
14. Kaplan CB, Centor RM. The use of nurses to evaluate house officers' humanistic behavior. *J Gen Intern Med.* 1990;5:410-4.
15. Henkin Y, Friedman M, Bouskila D, Kushnir D, Glick S. The use of patients as student evaluators. *Med Teacher.* 1990;12(3/4):279-89.

16. Tamblyn R, Benaroya S, Snell L, McLeod P, Schnarch B, Abrahamowicz M. The feasibility and value of using patient satisfaction ratings to evaluate internal medicine residents. *J Gen Intern Med.* 1994;9:146-52.
17. Weaver MJ, Ow C, Walker DJ, Degenhardt EF. A questionnaire for patients' evaluations of their physicians' humanistic behaviors. *J Gen Intern Med.* 1993;8:135-9.
18. Ramsay PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer rating to evaluate physician performance. *JAMA.* 1993;269(13):1655-60.
19. Gordon MJ. Cutting the Gordian knot: a two part approach to the evaluation and professional development of residents. *Acad Med.* 1997;72:876-80.
20. Gray JD. Global rating scales in residency education. *Acad Med.* 1996;71(1 suppl):S55-63.
21. Hill DA, Guinea AI, McCarthy WH. Formative assessment: a student perspective. *Med Educ.* 1994;28:394-9.
22. Hunt DD. Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. *Acad Med.* 1992;67(4):254-9.
23. Newble DI. Assessing clinical competence at the undergraduate level. *Med Educ.* 1992;26:504-11.
24. Norman GR, van der Vleuten CPM, de Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ.* 1991;25:119-26.
25. van der Vleuten CPM, Norman GR, de Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ.* 1991;25:110-8.
26. Maxim BR, Dielman TE. Dimensionality, internal consistency and inter rater reliability of clinical performance ratings. *Med Educ.* 1987;27:130-7.
27. Phillips SE. Legal issues in performance assessment. *West's Educ Law Q.* 1993;2(2):329-58.
28. Stillman PL. Positive effects of a clinical performance assessment program. *Acad Med.* 1991;66(8):481-3.
29. Rolfe IE, Andren JM, Pearson S, Hensley MS, Gordon JJ. Clinical competence of interns. *Med Educ.* 1995;29:225-30.
30. Neufeld V, Maudsley R, Pickering R, et al. Demand-side medical education: educating future physicians for Ontario. *Can Med Assoc J.* 1993;148:1471-7.
31. McCrae HM, Vu NV, Graham B, Word-Sims M, Colliver JA, Robbs RS. Comparing checklists and databases with physicians ratings as measures of students' history and physical-examination skills. *Acad Med.* 1995;70(4):313-7.
32. Haber RJ, Avins AL. Do ratings on the American Board of Internal Medicine resident evaluation form detect differences in clinical competence? *J Gen Intern Med.* 1994;9:140-5.
33. Smith RM. The triple-jump examination of an assessment tool in the problem-based medical curriculum at the University of Hawaii. *Acad Med.* 1993;68(5):366-72.
34. McGuire C. Perspectives in assessment. *Acad Med.* 1993;68(2)S3-8.
35. McLeod PJ. Assessing the value of student case write-ups and write-up evaluations. *Acad Med.* 1989;64:273-4.



## ANNOUNCEMENT

**JGIM Website — Visit us online today!**

Please visit the JGIM World-Wide Website:

**<http://www.blacksci.co.uk/products/journals/xjgim.htm>**