

In-Training Performance Assessment in Family Practice

James Molineux, MA, Brian K. Hennen, MD, MA, and Ian R. McWhinney, MD
London, Ontario

The In-Training Performance Assessment (ITPA) is an evaluation instrument derived from 174 basic objectives in family medicine. The instrument was applied to two consecutive classes of fourth-year medical undergraduates during the family medicine clerkship. Comparisons were made of two scoring systems: one measuring mastery of the objectives using the criterion of performance expected of a fully qualified family physician, the other using traditional categories of "poor," "satisfactory," "good," and "outstanding." The mastery evaluation model made it more difficult to achieve the objectives but had no effect on the discriminatory ability of the objectives when compared with the traditional evaluation method. The mastery model showed "management" to have the greatest differential between medical students and qualified family physicians. The evaluating supervisor was most influenced by the student's assessment in "problem-solving" using the traditional method, and by "management" using the mastery model. Management skills accounted for 89 percent of the variance of the overall competence assessments.

In September 1970, the Department of Family Medicine at the University of Western Ontario received a grant from the Ontario Ministry of Health to apply different kinds of evaluation techniques to the assessment of family physician performance and to determine the efficiency of these techniques at various levels of family physician development. There was an urgent need to evaluate family physician performance in order to: (1) provide continuing evaluation of undergraduate and graduate students of family medicine as they progress through formal curricula, to aid in learning and in career direction and selection; (2) evaluate performance of doctors in various postgraduate programs; and (3) compare different aspects of primary care in terms of efficiency and effectiveness.

This paper presents three years' experience with one method of assessment. The method used two basic principles of evaluation which combine practical considerations with the

classically required criteria of good evaluation. The first principle is that assessment should be based on the demonstration of having achieved or not achieved the defined objectives. This principle is now generally accepted, with justification, in writings by Bloom,¹ Miller² and Mager,³ among others. Such assessment allows for comparative peer assessment and evaluation against one's own performance and can provide reports of achievement which satisfy traditional evaluative reporting methods. The second principle is that the standard of performance expected should be that of qualified family physicians.

The following points support these principles:

1. The ultimate aim is to graduate competent family doctors (not to rank students or residents in terms of their peers).
2. Grades have no substantive meaning in terms of medical competence. In the world of practicing physicians we do not refer to 74 percent pediatricians or 87 percent family doctors, and medical school grades have not been accurate predictors of the quality of physicians' prac-

tice (Clute,⁴ Peterson⁵).

3. In our own department, a preliminary study comparing assessments based on objective achievement versus traditional excellent-good-satisfactory categories showed virtually no correlation.
4. Continuing medical education (CME) literature⁶ suggests that those physicians who study in response to internalized objectives based on self-recognized needs are more capable than those who simply attend refresher days offered by their medical organizations.
5. Learning towards objectives fits best with an undergraduate philosophy of self-learning and self-evaluation, a policy that the University of Western Ontario advocates.
6. Grading on the curve has problems including, for example, the student who "looks good" if he personally achieves a great deal or if his colleagues achieve poorly.
7. Assessment based against a common end-point permits accurate monitoring of changes in the quality of individual and group performance.
8. The majority of instructors are part-time physicians, usually chosen on the basis of competence, practice organization, and interest in teaching; most of them do not have the educational sophistication which is required to apply complex evaluation criteria. This is also true for many full-time teachers of family medicine. Most instructors in family medicine have a good idea, however, of the level of performance they expect of themselves or their colleagues in the context of the practice situation.

Method

The method involves the use of a rating scale, the In-Training Performance Assessment (ITPA), which is a modification of the Byrne-Freeman modification of the Supervisory Rating Scale developed by Blum et al in 1965.

A departmental objectives committee took as its base the publication *Educational Objectives for Certification of the Canadian College of Family*

From the Department of Family Medicine, The University of Western Ontario, London, Ontario. Requests for reprints should be addressed to Dr. Brian K. Hennen, Dalhousie Family Medicine Centre, Dalhousie University, Halifax, Nova Scotia.

This factor is concerned with the student's skill in establishing appropriate professional relationships with fellow students, paramedical personnel, and attending staff.

The Ineffective Student:

1. ___ is uncommunicative, cold, or may attempt to minimize professional contact with staff and peers;
2. ___ may be defensive, tactless, or inconsiderate toward staff and peers;
3. ___ is reluctant to assist others;
4. ___ has difficulty in giving or taking advice gracefully;
5. ___ deals with paramedical and public health personnel as subordinates rather than as professional colleagues;
6. ___ is unable to distribute appropriate aspects of care to others;
7. ___ has difficulty expressing himself;
8. ___ has difficulty expressing himself;
9. ___ does not acknowledge the contribution of others;
10. ___ does not evoke the confidence and cooperation of those with whom he works;
11. ___ does not work well in a team situation.

The Effective Student:

- seeks professional contact with staff and peers;
- is straightforward in his approach and contact with staff and peers;
- is available to assist others as required;
- offers and receives advice in a tactful and discreet manner;
- treats paramedical and public health personnel as professional members of a medical team;
- is able to obtain and organize the assistance of others;
- relates well to others;
- communicates easily;
- gives credit to others for their contribution;
- is respected by staff and peers;
- has a cooperative attitude and participates well in a team situation.

Figure 1. Relationship with Colleagues

Physicians.⁹ Descriptive behavioral paragraphs were divided into performance lists that could be checked by the assessor, and the 174 objectives were grouped into 11 categories: relationship with patients and family (RP), history and interviewing skills (HS), physical examinations (PE), problem solving and clinical judgment (PR), implementation of management plan (MP), use of the laboratory (LB) and medical records (RS), clinical responsibility (CR), relationship with colleagues (RC), orientation toward family practice (OP), and overall competence as a family physician (OC).

The behavioral descriptions were written in performance terms as shown in the example for the factor, "relationship with colleagues" (RC), illustrated in Figure 1.

The objectives committee also took each objective and empirically decided whether it was likely to be achievable by a third-year medical student, fourth-year medical student, or a resident in one of the four phases of the training program. We called this the *staging* of the objectives. We tried to decide the student's readiness to accomplish a particular objective. Our experience has shown that our empirical judgments were not always correct in this regard.

By using the achievement of objectives as the basis of measuring student learning and by expecting a level of performance equivalent to that of a qualified family physician, we felt that this was, in effect, a *mastery evaluation model*.

All the performance items are described in terms which indicate, for a specific task, what is clearly ineffective behavior and what is clearly effective behavior for a qualified family physician. The staging of the objectives attempted to assess the student against mastery for items of behavior which we thought, in terms of his readiness to learn and his current curriculum, he could reasonably achieve. For example, in a brief (16-hour) preceptorship experience for third-year medical students, only 24 of the 174 overall objectives were considered appropriate. A third-year student is not expected to set out a complex management plan for a patient, but he might be expected to relate comfortably to most patients in an interview.

At the end of each category, which comprised from 10 to 25 objectives in performance terms, a summary rating scale for that category was scored in one of two ways. The traditional way included the usual categories of "poor," "satisfactory," etc, whereas

the mastery model scored the student according to his progress in meeting objectives against the level of performance expected of a qualified family doctor (Figure 2).

This evaluation instrument was applied to two groups of 86 and 103 students. These groups represented two years of the fourth-year clinical clerkship in family medicine.

Results

Using various statistical techniques we made the following observations:

1. Analysis of item difficulty showed that the same objectives were more difficult to achieve when assessed by the mastery evaluation model (Table 1).
2. Using discrimination upper and lower indices, the discrimination power of the objectives was not altered when assessed by one or another evaluation model; that is to say, none of the individual performance items stood out as the predictor of the student's overall performance on the entire evaluation (Table 2).
3. Management is the performance area with the lowest average assessment under the mastery model and hence represents the greatest differential between medical students and practicing

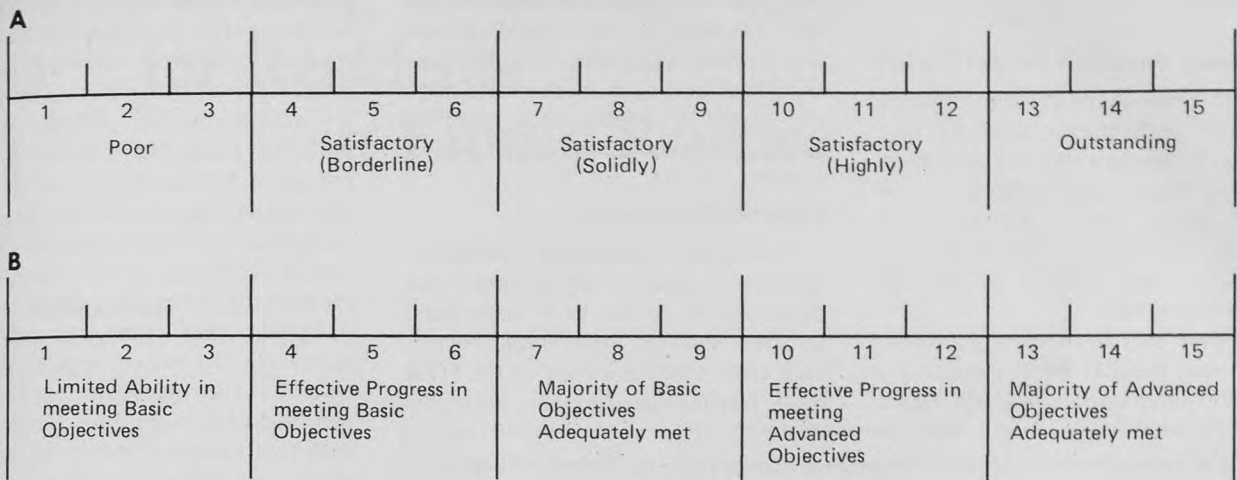


Figure 2. Methods of assigning subscale performance ratings. A, traditional model; B, mastery model.

family physicians (Table 3).

4. Using multiple regression analysis we determined that what most influenced the supervisor in scoring the overall competence of his students on the standard scale was their assessment on:

- a. problem-solving,
- b. relationship with colleagues, and
- c. relationship with patients.

What most influenced the supervisor in scoring students on the mastery scale was their assessment on:

- a. management,
- b. relationship with patients, and
- c. relationship with colleagues (Table 6).

5. Principle factors analysis of the two methods indicated that both scales are based on highly integrated measures of clinical skills, with the management subscale being the most representative in both cases. A second factor relating to social skills was also identified, with the subscale for relationship with colleagues being the most representative for both scales (Table 7).

6. Generally speaking, a good medical student overall is a good medical student in particular areas of performance.

7. Using the mastery model did not disrupt or impair any assessments previously made using

Table 1. Distribution of Difficulty Indices Under the Two Scaling Conditions

Condition	Range of Values			
	<20	20-39	40-59	≥60
Traditional	10	51	30	18
Mastery	48	46	11	4

$\chi^2 = 42.89 \quad p < .01$

Table 2. Distribution of Discrimination Indices Under the Two Scaling Conditions

Condition	Range of Values			
	<.10	.10-.29	.30-.49	≥.50
Traditional	20	58	22	9
Mastery	22	44	37	6

$\chi^2 = 6.19 \text{ N.S.}$

Table 3. Means and Standard Deviations of the Performance Ratings for the Two Scaling Conditions

Variable	Interpretation	Traditional		Mastery	
		\bar{x}	s	\bar{x}	s
OC	Overall competence as a family physician	10.2	2.23	7.4	1.94
RP	Relationship with patients and family	10.5	2.06	7.7	2.01
HS	History and interviewing skills	9.9	2.20	7.6	1.98
PE	Physical examinations	10.2	1.82	7.7	2.01
PR	Problem solving and clinical judgment	9.7	2.34	7.3	2.05
MP	Management plan	9.7	2.05	6.9	1.87
LB	Use of the laboratory	9.8	2.02	7.3	2.01
RS	Medical records	11.0	2.15	8.4	2.66
CR	Clinical responsibility	10.8	2.16	7.9	2.12
RC	Relationship with colleagues	10.7	2.49	8.9	2.23
OP	Orientation toward family practice	10.7	2.49	9.1	2.35

traditional methods (Tables 4 and 5).

Preliminary Results of Further Studies

Further use of the ITPA to evaluate residents' performance suggests that the use of the mastery criteria is more objective than traditional scales of "fair," "good," etc, in that the score given on the mastery model is proportionate to the number of specific behaviors checked.

We have also found (using another instrument, the office-visit assessment) that inter-rater correlations are higher

with the mastery criteria applied. In this case the evaluators were all full-time teachers, so this would support the previous suggestion that full-time as well as part-time teachers are more able to use the criteria of the qualified family physician's expected performance.

Reporting Performance

How is this information transmitted back to the student? In the case of the clerks, who up to 1975 have been involved over a three-week period, each clerk receives a copy of the ITPA when starting his clerkship. When the

instructor fills it out he does so in the presence of the student, once during and once after the clerkship. Thus, the student gets direct feedback immediately. To residents we are able to give a printed profile which is now available from the computer. It is presented to the resident in the form of a graph, one for each of the major categories of performance. He is scored according to the phase of the program he is in, can look at his own progress over time, is able to appreciate to what degree he has mastered the objectives expected of him, and can compare his performance with that expected of him on the basis of previous residents' experience.

Finally, with increasing experience and further analysis of the data we already have available, we believe we will be able to pare down these assessment scales to focus more accurately on those objectives capable of being met at a given level of training, within a prescribed time, and with a minimum number of dimensions needed to describe clinical performance. Our data are beginning to tell us which of the objectives third-year students are capable of meeting, and it would appear that faculty opinion which empirically staged the objectives, does not necessarily correlate with what the students can, in fact, do. We have already used some of our analyses of clerkships to develop the objectives and assessment techniques for the third-year preceptorship preceding the clerkship.

References

1. Bloom BS: Taxonomy of Educational Objectives. New York, McKay, 1956
2. Miller GE (ed): Teaching and Learning in Medical School. Cambridge, Mass, Harvard University Press, 1968
3. Mager RF: Preparing Instructional Objectives. Belmont, Calif, Reardon, 1962
4. Clute KF: The General Practitioner. Toronto, University of Toronto Press, 1963
5. Peterson OL, Andrews LP, Spain RS, et al: Analytic study of North Carolina general practice, 1953-1954. J Med Educ 31(12, pt 2):1-165, 1956
6. Brown CR, Fleisher DS: The bi-cycle concept - relating continuing education directly to patient care. Continuing medical education in community hospitals, a manual for program development. N Engl J Med 284 (suppl), May 20, 1971, pp 88-97
7. Byrne PS, Long BEL: Learning to Care. London, Churchill Livingstone, 1973
8. Blum JF, Fitzpatrick R: Critical Performance Requirements for Orthopedic Surgery. University of Illinois, College of Medicine, Office of Research in Medical Education, 1956
9. Committee on Educational Objectives: Educational Objectives for Certification in Family Medicine, Pt 1. Toronto, College of Family Physicians of Canada, 1971

Table 4. Intercorrelations Between Performance Ratings Using the Traditional Scaling Procedure

Variable	RP	HS	PE	PR	MP	LB	RS	CR	RC	OP
OC	.75	.76	.59	.88	.84	.73	.59	.74	.78	.71
RP		.63	.53	.65	.62	.61	.46	.48	.54	.59
HS			.62	.78	.70	.65	.52	.58	.60	.56
PE				.64	.56	.46	.45	.41	.36	.34
PR					.82	.75	.47	.63	.64	.64
MP						.70	.56	.69	.74	.71
LB							.39	.51	.57	.66
RS								.51	.54	.47
CR									.79	.67
RC										.82

Table 5. Intercorrelations Between Performance Ratings Using the Mastery Scaling Procedure

Variable	RP	HS	PE	PR	MP	LB	RS	CR	RC	OP
OC	.81	.81	.72	.81	.87	.68	.63	.79	.60	.66
RP		.80	.69	.69	.69	.54	.58	.74	.61	.67
HS			.61	.73	.72	.53	.61	.70	.55	.51
PE				.72	.68	.50	.54	.68	.44	.57
PR					.77	.62	.53	.62	.40	.52
MP						.70	.54	.69	.47	.53
LB							.43	.44	.32	.38
RS								.65	.61	.56
CR									.61	.65
RC										.78

Table 6. Multiple Regression with Stepwise Addition of Variables Under the Two Scaling Conditions

Traditional Scaling Model				Mastery Scaling Model			
Step #	Variable	R	R ²	Step #	Variable	R	R ²
1	PS	.88	.78	1	MP	.87	.75
2	CR	.92	.85	2	RP	.91	.83
3	RP	.94	.88	3	CR	.92	.85

Table 7. Principal Factors Analysis of the Two Scaling Conditions

Variable	Rotated Factor Matrices			
	Traditional Scaling Model		Mastery Scaling Model	
	Factor 1	Factor 2	Factor 1	Factor 2
RP	.729	.145	.867	-.102
HS	.818	.237	.841	-.011
PE	.626	.416	.797	-.062
PR	.886	.214	.857	-.241
MP	.889	.015	.872	-.185
LB	.762	.101	.676	-.220
RS	.615	-.019	.698	.226
CR	.766	-.234	.822	.177
RC	.850	-.471	.642	.643
OP	.804	-.307	.716	.444