An In-Training Assessment Examination in Family Medicine: Report of a Pilot Project

John E. Donnelly, MD, Bonnie Yankaskas, MPH, Craig Gjerde, PhD, John C. Wright, MD, and Douglas P. Longnecker, MD
Farmington, Connecticut and Dayton, Ohio

A family medicine in-training assessment examination was developed and piloted in 20 programs across the country. Core Content Review questions were used for the examination. Reporting of scores used both the traditional, normative-referenced approach, and a criterion-referenced approach. The latter permitted family medicine faculty to set passing standards for the examination.

The pilot project was well received and the examination will be offered to all family medicine residency programs this year.

Resident assessment examinations have been in use for over a decade. They were first designed to assist residents in preparation for board examinations and gradually have developed into examinations designed to aid individual residents and residency programs in assessment of strengths and weaknesses. The specialties of orthopedics, neurosurgery, ophthalmology, obstetrics and gynecology, and pediatrics have all developed resident assessment examinations.[1-4] These are annual, norm-referenced examinations.

Family medicine, with its rapid development and growth, has not developed a national resident assessment examination. In 1975, the Core Content Review Board formed a subcommittee charged with the responsibility of developing an assessment examination that could be offered to family medicine programs nationally. The examination was to be designed in such a fashion that programs could assess how well they were meeting their instructional goals and offer feedback to individual residents on their areas of strength and weakness.

The paper describes the nationwide pilot project that accepted the following goals: to develop a criterion-referenced examination for family medicine residents; to test the acceptance of an in-training assessment examination by residency programs; and, to test the acceptance of the criterion-referenced approach to setting passing standards.

Methodology

Study Group

Thirty family medicine residencies were selected nationwide and invited to participate in a pilot examination scheduled for July 1976. Twenty-eight residencies, with a total of 465 residents, agreed to participate. There were no quotas set as to numbers of residents or year of training. In some programs only a portion of the residents took the examination.

Two hundred eighty-five residents from 20 residencies completed the examination: 110 were beginning their first year, 93 were beginning their second year, and 82 were beginning their third year. The participants were from many geographically distinct areas and represented both university-based and community hospital-based programs.

Selection of Questions

Questions for the examination were taken from the bank of questions of the Core Content Review, a self-assessment multiple-choice examination developed by the Connecticut and Ohio Academies of Family Physicians for family practitioners. Core Content Review (CCR) questions are written by experts in various disciplines, and screened for teaching value and relevance by an executive committee of family physicians of CCR.

All CCR questions from the 1974-1975 and 1975-1976 examinations were subcategorized by content using the International Classification of Health Problems in Primary Care (ICHPPC) guidelines. From the 1,000 available questions, 400 were selected. This resulted in the desired balance of questions from the five major subject areas: medicine, pediatrics, surgery, psychiatry, and obstetrics and gynecology.

Questions were also categorized according to the cognitive level required to answer the question: recall, understanding, or problem-solving. Thirty percent of the questions were recall, 35 percent understanding, and 35 percent problem-solving questions. The breakdown of the subcategories is shown in Table 1.

Setting a Minimum Pass Level (MPL)

For this examination, criterion-referenced passing standards were used.[5-7] The criterion-referenced examination offers an alternative to the traditional norm-referenced approach. The difference is that norm-referenced evaluation compares an individual against average performance, while criterion-referenced evaluation measures the individual's performance against previously determined standards. The examination was designed to assess mastery of a core body of content and there was no guarantee that normative scores based on a self-selected group of residents would be meaningful.

Ten faculty members of family medicine residency programs were

From the Departments of Family Medicine and Research in Health Education, University of Connecticut, Farmington, Connecticut, and the Department of Family Practice, Wright State University, Dayton, Ohio. Requests for reprints should be addressed to Dr. John E. Donnelly, Department of Family Medicine, University of Connecticut, Farmington, CT 06032.

Table 1. Number of Questions in Each Major Subject by ICHPPC Classification and Cognitive Level

	Major Subject				
	MED	PED	PSYCH	SURG	OB/GYN
ICHPPC Classification					
Accidents, poison, violence	3	10			
Cardiology	15	5			
Congenital and Perinatal	—	10			
Dermatology	5	5			
Endocrinology	10	10			
Gastroenterology	10	5			
Hematology	5	2			
Infectious Disease	15	25			
Mental Disorder	—	5	100		
Musculoskeletal System and Connective Tissue	5	5			
Oncology	7	3			
Renal Disease	10	5			
Respiratory Disease	10	10			
Other	5	—			
Total	100	100	100	50	50
Cognitive Level					
Recall	20	24	39	15	23
Understanding	21	43	45	19	13
Problem Solving	59	33	16	16	14
Total	100	100	100	50	50

recruited to help set the MPL for the examination. Every answer to every question was evaluated and scored by at least two different raters. In establishing the criterion-referenced passing standards, raters assigned to each multiple-choice possibility a weight equal to the degree to which choosing that particular item would correctly reflected appropriate mastery of the material. The item MPL weights were averaged across raters and summed. These sums were then transformed to percent figures with separate MPLs set for the total examination and each of the five subtests. The MPLs were 66.2 percent for the total examination, 66.0 percent for medicine, 65.8 percent for pediatrics, 65.3 percent for psychiatry, 68.9 percent for obstetrics and gynecology, and 66.4 percent for surgery. For the pilot project, the MPL criterion was based on the desired knowledge of a resident who had completed two years of training (a beginning third year resident).

Test Administration

The examination was available for administration between June 20 and July 15, 1976. Program directors received detailed instructions on administration. Six hours were allowed for testing and no references were permitted. Residents recorded their answers on computer scorable sheets and all testing materials were collected after the examination.

Following the computer scoring of results each residency director received the results in package form. The package contained individual results of each resident and composite results of all the residents in that particular program by year of training. In addition, summary results of all residents who took the examination were included, also by year of training.

Surveys

The director of each participating program was sent a questionnaire before the examination to ascertain what types of resident assessments were in current use and to establish some data about the director's expectations.

Six weeks after the results were mailed, a telephone interview was conducted with 15 of the 20 program directors. Its purpose was to get feedback on difficulties in administration of the examination, problems in evaluating its results, use of its information, and desirability for an annual standardized examination like the pilot project.

Results

In the pretest questionnaire, 15 out of 17 directors who replied felt there was a need for an annual assessment examination. Six programs had used the standard Core Content Review for their residents as a self-assessment examination, two programs used patient management problems, and one used a pretest examination. The balance had not used any written test in their programs.

All program directors interviewed by telephone had received and reviewed the results of the examination. They had all used the examination for program assessment and for assessments of individual residents in a supportive and positive manner. No punitive action resulted from the pilot examination. Fourteen of the 15 program directors repeated their desire for a yearly examination of this type.

The MPL approach was acceptable to most programs and it was frequently suggested that multiple pass levels be used (eg, one for each year of training) rather than one MPL for all residents.

Examination results are reported in Table 2. It should be repeated that the MPL was set for a beginning third year resident. The examination did serve as a pretest for residents beginning their first year of training. An analysis of test results comparing years one, two, and three on the examination showed

Table 2. Examination Results (MPLs reported in percent)

	Total	MED	PED	PSYCH	OB/GYN	SURG
All Residents N=285						
MPL	66.2	66.0	65.8	65.3	68.9	66.4
MEAN	63.4	60.0	62.3	67.8	63.1	63.8
SD*	6.5	7.9	8.2	7.2	9.8	7.7
# Res above MPL	104	75	97	200	85	112
# Res below MPL	181	210	188	85	200	173
Third Year Residents N=82						
MPL	66.2	66.0	65.8	65.3	68.9	66.4
MEAN	65.6	63.2	65.0	68.9	64.6	65.9
SD	6.0	7.4	6.7	6.6	9.7	6.6
# Res above MPL	47	33	39	61	30	41
# Res below MPL	35	49	43	21	52	41
Second Year Residents N=93						
MPL	66.2	66.0	65.8	65.3	68.9	66.4
MEAN	65.3	62.2	65.3	68.2	65.6	65.1
SD	5.5	6.9	7.7	6.1	8.7	7.0
# Res above MPL	42	34	47	64	36	41
# Res below MPL	51	59	46	29	57	52
First Year Residents N=100						
MPL	66.2	66.0	65.8	65.3	68.9	66.4
MEAN	60.1	55.7	57.6	66.7	59.8	61.2
SD	6.2	7.1	7.5	8.2	9.9	8.1
# Res above MPL	15	8	11	75	19	30
# Res below MPL	95	102	99	35	91	80

*SD = Standard Deviation

year two and year three had significantly higher scores than year one in all the subtests except psychiatry (analysis of variance, $P<.001$). On the total examination, 14 percent of the first year residents, 45 percent of the second year, and 57 percent of the third year residents attained the preset passing level (MPL).

Conclusion

As a pilot project, the results were encouraging. Interest in an in-training assessment examination was documented. The method employed (MPL) together with the scoring results indicated that a criterion-referenced examination could be used as an in-training assessment vehicle. Finally, the residency directors used the scores to assess strengths and weaknesses of individual residents and total programs.

In the future, as universally acceptable teaching objectives are developed, questions can be formulated to test those objectives. The use of a criterion-referenced examination would permit educators in family medicine to establish minimally acceptable passing standards for all programs and simultaneously assess strengths and weaknesses of individual residents and programs.

Acknowledgement

The support of the Core Content Review of the Connecticut and Ohio Academies of Family Physicians and the participating residency directors is gratefully acknowledged.

References

1. Natress LW: Orthopaedic in-training examination. J Med Educ 44:878, 1969
2. Hubbard JP, Furlow LT, Matson DD: An in-training examination for residents as a guide to learning. N Engl J Med 276:448, 1967
3. Rubin ML: The ophthalmology resident in-training examination. Am J Ophthalmol 67:70, 1969
4. Meskauskas J, Newton M, Russell K: The 1973 in-training examination in obstetrics and gynecology. Obstet Gynecol 44:463, 1974
5. Taylor DD, Reid JC: Criterion-referenced evaluation of student performance. J Med Educ 47:970, 1972
6. Nedelsky L: Absolute grading standards for objective tests. Educ Psychol Meas 14:3, 1954
7. Bobula J: Minimum Pass Level Study. Report to Faculty, Center for Educational Development, College of Medicine, University of Illinois. Urbana, Illinois, University of Illinois, 1974, pp 25-31