DOCUMENT RESUME

ED 226 010 TM 820 878

AUTHOR Markert, Ronald J.; Kolmen, Samuel N.
TITLE WSUSOM Curriculum Evaluation: Analysis Using NBME Examinations. Program Evaluation Studies, Report No. 8.
INSTITUTION Wright State Univ., Xenia, OH. School of Medicine.
PUB DATE Sep 82
NOTE 9p.; This study was partially funded by the Miami Valley Area Health Education Center.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Course Content; *Curriculum Evaluation; *Evaluation Methods; *Graduate Medical Education; Higher Education; Medical School Faculty; National Norms; Occupational Tests; Scores
IDENTIFIERS *National Board of Medical Examiners

ABSTRACT

This report presents a method for analyzing curriculum using the examinations of the National Board of Medical Examiners (NBME). The method provides a system by which the NBME examinations can be used each year to compare the content of an external licensing test with the content of the curriculum at the Wright State University School of Medicine (WSUSOM). The NBME test items were classified by faculty content experts on four dimensions: school-wide topics, discipline topics, cognitive levels, and item types. Then, the percent of WSUSOM students and percent of national sample answering the item correctly were recorded. The analysis allows comparison of the performance of WSUSOM student and the national sample on each school-wide or discipline topic. The Faculty Curriculum Committee is the most interested user of data related to school-wide topics, while the individual department is the most appropriate audience for the discipline topic information. Curriculum decision-makers should take special caution in examining the results for a single year's administration, since test content varies from year to year. (Author/BW)

Program Evaluation Studies
Wright State University
School of Medicine

Report No. 8 September 1982

WSUSOM Curriculum Evaluation: Analysis Using NBME Examinations

Ronald J. Markert, Ph.D.

Samuel N. Kolmen, Ph.D.

Purpose

The examinations of the National Board of Medical Examinations (NBME) are used extensively by U.S. medical schools to evaluate institutional effectiveness. The 1980 Annual Report of the National Board reported that about 75 percent of U.S. medical schools use the NBME exams for institutional assessment. Despite their program evaluation importance to medical schools, very little has been reported in the literature to guide local efforts in making the best use of the NBME examinations.

This report presents a method for analyzing curriculum using the examinations of the NBME. The method provides a system by which the NBME examinations can be used each year to compare the content of an external

This study was partially funded by the Miami Valley Area Health Education Center

licensing test with the content of the WSUSOM curriculum. The analyses provide data to medical school curriculum decision-makers which can be used to examine topics which cut across individual disciplines. Multi-disciplinary information of this kind permit a cooperative effort within the medical school in evaluating teaching effectiveness. In addition, departments are provided with data on the content areas identified as specific to their disciplines.

Review of the Literature

Curriculum evaluation using the National Boards has most frequently used the NBME Part I (Sanazaro, 1967; Kennedy et al., 1970; Wile, 1978; Garrard et al., 1978; McGaghie et al., 1980; McGaghie et al., 1981; Baum et al., 1981). Studies using Part I generally have focused on (1) the match between NBME content and local content (Kennedy et al., 1970; Wile, 1978; Garrard et al., 1978; McGaghie et al., 1980; McGaghie et al., 1981) and (2) the relationship between local curriculum emphasis allotted to NBME content and student achievement (Sanazaro, 1967; Wile, 1970; Garrard et al., 1978; McGaghie et al., 1980; McGaghie et al., 1981).

While evaluation studies using Part I have examined all or nearly all of the component disciplines, Part II investigations have focused on one individual discipline: surgery (Linn et al., 1979; Benenson et al., 1981), obstetrics and gynecology (Spellacy and Dockery, 1980), and medicine (Calhoun et al., 1980). The validity of local evaluation measures was the focus of three of these studies (Linn et al., 1979; Spellacy and Dockery, 1980; Benenson et al., 1981). The Calhoun et al. study (1980) examined the relationship between student achievement and type of clerkship activity (patient care, education, joint patient care and education).

Method

As a response to a mandate from the Wright State University School of Medicine Faculty Curriculum Committee, the authors formulated a plan to analyze the test items on the Part I and II examinations of the NBME. Part I was analyzed following its administration in June 1981 and Part II following its administration in September 1981. This first round of analysis in 1981 will be followed each succeeding year by similar investigations.

For Part I one of the authors (SNK) classified the items for the seven examinations (anatomy, behaviorial science, biochemistry, microbiology, pathology, pharmacology and physiology). For Part II a WSUSOM faculty content expert from each of the six disciplines (medicine, obstetrics/gynecology, pediatrics, preventive medicine/public health, psychiatry and surgery) classified the items. WSUSOM faculty content experts were not used for Part I because this first investigation was considered a pilot study of a new methodology. However, all succeeding analyses will employ WSUSOM faculty content experts for both Part I and Part II.

For the 1981 round of analysis Table 1 shows the four dimensions on which each item was classified and the variation in classification systems used in Part I and Part II. A more standardized classification system, where appropriate, will be developed for the 1982 investigations. For Part II in 1981 (and for all succeeding analyses of Part I and Part II), content experts were permitted to use more than one category in classifying items by school-wide topic and discipline topic. Multiple classifications permit a better understanding of an exam's content.

Table 1: Classification of Items on NBME Part I and Part II

	Part I	Part II
School-wide topics	23	33
Discipline topics	varies	varies
Cognitive levels	2[a]	3[b]
Item types	5[c]	4[d]

[a] factual recognition/knowledge and application of knowledge/ problem-solving
[b] factual recognition/knowledge, diagnosis, treatment/management
[c] one best response, one best response (except), matching list, matching compare/contrast, multiple true/false
[d] one best response, matching list, matching compare/contrast, multiple true/false

Finally, the percent of WSUSOM students and percent of the national sample answering the item correctly were recorded. This permitted each item to be categorized on the basis of WSUSOM versus national performance with intervals of 5 percentage points used to form cells. Thus, when WSUSOM and the national sample differed by 5 percentage points or less on an item, the item was classified as same (i.e., no difference between WSUSOM and the national sample). A difference of greater than 5 percentage points resulted in an item being placed either in the WSUSOM or nation column.

A few examples from the 1981 Part I and Part II analysis should suffice to demonstrate how curriculum decision-makers can examine content which cuts across disciplines. School-wide Topic A had 119 items on Part I and 27 items on Part II. On about one-half (55) of the part I items, WSUSOM students scored the same as the national sample. However, on the

remaining 64 Part I items, WSUSOM students were lower on 47 (73%) of the items. Performance on School-wide Topic A did not improve on Part II. While two-thirds of the 27 items (18) were in the same column, WSUSOM was outscored on 8 of the 9 remaining items. School-wide Topic A is an example of low WSUSOM performance on both National Boards Part I and Part II. There were 35 School-wide Topic B items on Part I and 19 School-wide Topic B items on Part II. WSUSOM scored the same as the national sample on 16 Part I items, better on 11 items, and worse on 8 items. Similarly, WSUSOM scored the same on 10 Part II items, better on 7 items and worse on 2 items. School-wide Topic B illustrates content on which WSUSOM scored better than the national sample on both Part I and Part II.

Other school-wide topics can be found on which WSUSOM performance was better (or worse) than the national sample on both Part I and Part II. In addition, there are school-wide topics on which WSUSOM Part I performance was high while Part II performance was low, and vice versa. School-wide Topic C had WSUSOM outperforming the national sample on a 2 to 1 basis on those items where there was a difference on Part I. However, on Part II the national sample outperformed WSUSOM by a small margin. In contrast, the national sample did better by more than a 3 to 1 ratio on School-wide Topic D for Part I, but WSUSOM reversed the results on Part II with a 2 to 1 ratio of better performance on items where there was a difference.

In addition to comparison from Part I to Part II, faculty can examine the disciplines which were most influential in producing a positive (or negative) result. For example, was the poor performance on School-wide Topic A due to a few of the eleven disciplines included or were the low results uniform across disciplines. Similarly, was high achievement on

School-wide Topic B due to a few or most disciplines. Finally, the second type of content classification-within departments-can be beneficial to faculty. For the 1981 analysis each department classified items according to its view of how its discipline is organized. Examination of the results from this analysis can be helpful to departmental faculty in evaluating discipline topics as they are taught in the curriculum.

Discussion

The method presented in this report provides information to curriculum decision-makers such as the Faculty Curriculum Committee, department chairmen, and course directors through evaluation of NBME examinations. The Faculty Curriculum Committee is the most interested user of data related to school-wide topics while the individual department is the most appropriate audience for the discipline topic information. Curriculum decision-makers are advised to take special caution in examining the results for a single year's administration. The content of NBME exams varies from administration to administration, and only through the compilation of results over a number of years can validity be attached to the findings. The system presented in this report will be employed each year to analyze NBME Part I and Part II examinations with the intent of providing curriculum decision-makers with valuable data on which to make judgments.

References

Baum, J.H., Wimmer, R.S., Smalley, R.V., Moldawar, N.P. and Frankel, M.C. Is there an oncology domain in Part I of the NBME tests? J. Med. Educ., 56: 525-526, 1981.

Benenson, T.F., Stimmel, B. and Aufses, A., Jr. Concordance of surgical clerkship performance and National Board of Medical Examinations Part II subtest scores. A validation model. Surg., 89: 692-696, 1981.

Calhoun, J.G., Hull, A.L. and Davis, W.K. The relationship between medical student clerkship activities and performance on NBME Part II. Annu. Conf. Res. Med. Educ., 19: 103-108, 1980.

Garrard, J, McCollister, R.J. and Harris, I. Review by medical teachers of a certification examination: Rationale, method and application. Med. Educ., 12: 421-426, 1978.

Kennedy, W.B., Kelley, P.R. and Hubbard, J.P. The Relevance of National Board Part I Examinations to Medical School Curricula. Philadelphia: National Board of Medical Examiners, 1970.

Linn, B.S., Schimmel, R., Wirch, J., Pratt, T. and Zeppa, R. Where National Board Examinations Pass and Fail in Evaluating Knowledge of Surgical Clerkship. J. Surg. Res., 26: 97-100, 1979.

McGaghie, W.C., Burford, H.J. and Harward, D.H. Content representativeness and student performance on National Board Part I special subject examinations. Annu. Conf. Res. Med. Educ., 19: 15-20, 1980.

McGaghie, W.C., Burford, H.J. and Harward, D.H. External and internal tests of preclinical learning: Content representativeness, item educational emphasis, and student achievement. Annu. Conf. Res. Med. Educ., 20: 115-120, 1981.

Sanazaro, P.J. Educational Self-Study by Schools of Medicine. Evanston, IL: Association of American Medical Colleges, 1967.

Spellacy, W.N. and Dockery, J.L. A comparison of medical student performance on the obstetrics and gynecology National Board Part II examination and a comparable examination given during the clerkship. J. Reprod. Med., 24: 76-78, 1980.

Wile, M.Z. External examinations for internal evaluation: The National Board Part I test as a case. J. Med. Educ., 53: 92-97, 1978.

PROGRAM EVALUATION STUDIES*

Report No. 1	Program Evaluation Studies: An Overview	Nov. 1980
Report No. 2	The Prediction of National Board Performance, Specialty, and Location of Residency for the Wright State University School of Medicine Class of 1980	April 1981
Report No. 3	Medical Education at Wright State University: An Evaluation During the First Year of Residency by the Class of 1980 and Their Residency Supervisors	June 1981
Report No. 4	Stability and Change of Medical Specialty Choice: Wright State University School of Medicine Class of 1981	Dec. 1981
Report No. 5	The Relationship Between First-Year Medical School Grades and Academic, Personality, and Attitudes Measures: Wright State University School of Medicine Class of 1984	March 1982
Report No. 6	The Prediction of Licensibility, Specialty, and Location of Residency for the Wright State University School of Medicine Class of 1981	August 1982
Report No. 7	Stability and Change of Medical Specialty Choice: Wright State University School of Medicine Classes of 1981 and 1982	August 1982
Report No. 8	WSUSOM Curriculum Evaluation: Analysis Using NBME Examinations	Sept. 1982

*Available from:

Ronald J. Markert, Ph.D.
Dept. Postgraduate Medicine and
Continuing Education
Greene Memorial Hospital
Xenia, Ohio 45385