DOCUMENT RESUME

ED 132 211 TM 005 948

AUTHOR Willoughby, T. Lee
TITLE Evaluating the Information Base of Students in Health Science Programs.
PUB DATE [76]
NOTE 19p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (New Orleans, Louisiana, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bachelors Degrees; *Computer Oriented Programs; Item Analysis; *Item Banks; *Medical Education; Student Testing; Test Reliability; Test Validity

ABSTRACT

Concommitant with changes in education, health care roles, and evaluation in recent years, a new six year combined B.S. and M.D. program was inititated at the University of Missouri at Kansas City. Evaluation was integrated into the academic plan in order to assess student progress toward the overall goal of becoming a safe physician, as well as achievement in specific areas of the curriculum. Although several evaluation procedures are in place at this institution, this paper describes only the Item Library. The Item Library is a computerized information retrieval system containing, at this time, 10,000 test questions and support information. Each entry of the Item Library consists of the test question, Medical Subject Heading (MeSH) code, date submitted, source, history of use and update, item analysis data, and a specific bibliographic reference. The psychometric data indicate that the Quarterly Profile Exams generated from the Item Library are reliable and valid measures. Extending the computer assisted testing procedures of the Item Library into other allied health fields is in progress. (Author/RC)

Evaluating the Information Base of Students in Health Science Programs[1]

T. Lee Willoughby
University of Missouri - Kansas City

In recent years several developments have occurred regarding the education and role of members of the health care team. Developments in the educational process included combining programs in a manner which cuts across the traditional departments and disciplines. The process of integrating programs yielded curricula which include community involvement more so than in the past. As the concept of the health care team has evolved, new and expanded roles have developed. For example, the role of the physician assistant has developed within the past decade. Continuing education and career mobility have been emphasized recently much more so than in the past. The December, 1975 Journal of Medical Education is commended to your reading for more information on the above points.

Concomitantly, related developments in education occurred. There has been a movement to specify curriculum goals and objectives in far greater detail than in the past. Testing of student mastery of information has been geared to these curriculum objectives. Furthermore, the computer has been introduced to curriculum, evaluation, and instructional aspects of the educational process. The program described in this pre-

[1]Presented at the annual meeting of Southeastern Psychological Association; New Orleans, La., 1976 as part of the Symposium "Selecting and Evaluating Students for their Future Role on the Health Care Team."

sentation reflects many of the trends in health and education described thus far.

The University of Missouri - Kansas City School of Medicine admits students directly from high school to a six-year combined undergraduate-medical education program. One unique feature of the curriculum organization is the assignment of students in small groups to physician-instructors called "docents". Each docent has responsibility for the medical content of the curriculum. In the first two years of study, liberal arts and medicine represent 75% and 25% respectively of the curriculum. During years three through six these proportions are reversed. Thus, students receive their bachelors and medical degrees simultaneously at the conclusion of the program.

Evaluation was integrated into the academic plan at this new School of Medicine to assess student progress toward the overall goal of becoming a safe physician, as well as achievement in specific areas of the curriculum. The evaluation process is frequently used to make prescriptive statements about individual student's curriculum plan. This paper focuses on one aspect of the overall evaluation process, the Item Library. In the next several minutes, I will describe the Item Library, discuss its functions, propose its use in health care programs, describe psychometric properties of exams generated from the Item Library, and discuss the significance of our use of computer technology.

The Item Library is a computerized information retrieval system containing 9000 test questions and support information;

the number continues to increase each month. Each item of the Item Library consists of the test question, Medical Subject Heading (MeSH) code, date submitted, source, history of use, item analysis data, and a specific bibliographic reference. ~~Slide 1 illustrates the content of each item.~~ Some of these components warrant additional comment.

For our purposes, a "good" test question should include a statement of fact, the question, and a set of five response alternatives. Faculty, students, and visiting docents contribute test questions relevant to their field of specialty. The quantity and quality of contributions has steadily increased with time. The School of Medicine at the University of Wisconsin and the Rockford School of Medicine have been generous in allowing us access to their question banks. The 7000 test questions currently in the Item Library represents less than the number of the questions on hand, because each question is extensively reviewed, referenced and coded for MeSH before entered in the computer.

The MeSH code is a tree structure indexing method developed by the National Library of Medicine. Each test question is assigned at least one and no more than twelve MeSH codes, according to its content. Item analysis data is the proportion of examinees responding to each answer choice at years 1-6 and faculty/staff for the most recent use of the questions and combined use of the questions. An index of item reliability ~~will be added soon~~ is also stored.

The Item Library is used to generate our Quarterly Profile Exam and to provide self-assessment learning experiences via remote terminals. The Quarterly Profile Examination (QPE) is a set of 400 questions intended to assess factual knowledge in the area of medicine. Each quarter the QPE is generated according to specifications of content area (MeSH codes) and the proportion of old/new questions. The MeSH codes appropriate for five major content areas, Internal Medicine, Pediatrics, Ob/Gyn, Surgery, and Basic Science are used to select the test items. Usually, 80% of the test questions have not been used in a previous QPE. Since test questions generally have more than one MeSH code, we are able to establish additional categories for scoring. A given test question may, therefore, contribute to the score of more than one category. The list of categories and their appropriate MeSH code(s) is shown in slide 2.

The QPE is administered to all students, years one through six, as well as interested faculty, staff, and allied health professionals. Students are encouraged to view the exam as an assessment/learning experience. Since students leave the exam with their exam booklet, references to most questions, and receive scores and answer key in a few days, the likelihood of viewing the exam as a learning experience is increased. They are aware, however, that the Council on Evaluation weighs heavily QPE performance in their promotional considerations. Two flagging criteria are applied to each quarters' scores to facilitate Council review: (1) bottom 15% of peer group and; (2) score below mean of class one year level under peer group.

At this point I should add that test questions are flagged on each exam, as well as student's scores. Any item flagged by one of these two criteria is reviewed: (1) year 6 per cent correct minus year 1 per cent correct less than 20%; and, (2) year 6 per cent correct less than 60%. In addition to the flagging by item analysis data, challenges to the examination are encouraged by the Council on Evaluation.

The self-assessment opportunity via remote terminals is by nature a learning experience. Ten cathode ray tube computer terminals are available for the students' use to access the Item Library. In effect, students have access to the pool of questions from which their future QPE's will come. To generate a quiz at the terminal, one specifies number of questions and the MeSH code(s) appropriate to the content. The content area may be as general or specific as the MeSH tree structure allows. For example, a student might seek questions for the general area of Hemic and Lympathic Disease (C9) or questions on a topic as specific as Thrombocytosis (C9.33.43.109). Up to twelve MeSH may be combined with Boolean logic to generate a search. Then the question and set of response alternatives are presented on the terminal screen. After the student's response, the correct answer is indicated. The student may also ask for item analysis data on each question.

The potential application of the Item Library, conceptually and concretely, to other programs of medicine and other fields of the health sciences appears promising. In fact, we began about two years ago building into the Item Library the capa-

bility of generating examinations for the non-physician primary care practioner. The Profile Examination for the Physician Extender (PEPE) will be the counterpart of the QPE. The next portion of this presentation will describe the major features of the PEPE project.

Rather than gear our generation of PEPE items to a single physician extender curriculum, we took an approach which would apply to programs in general. Dr. Ned Smull and Ms. Jane Kerber developed a curriculum outline which represents the likely didactic material of physician extender programs. Including major categories and their subdivisions, there are about 1000 elements in the curriculum outline. MeSH codes were assigned each element of the curriculum outline. The next step was to study the curricula of physician extender programs and record the number of class hours devoted to the teaching of topics of the curriculum outline. The curriculum outline, MeSH codes, and hours of teaching by different programs forms what we call the curriculum matrix. The accompanying slide illustrates the matrix.

The curriculum matrix is the core upon which curriculum documentation, item generation, exam generation, score reporting, and curriculum prescription depend. The curriculum matrix is "computerized" so that it interfaces with the Item Library. Currently we use this system to conduct periodic inventories of the Item Library by content. This information guides our item generation efforts by identifying content areas with few items. The system has also been used to generate pro-

totype PEPE's. The MeSH codes assigned to elements of the matrix were used to retrieve from the Item Library exams representative of the curriculum content.

How would anyone use this system to generate exams for their specific curriculum? The answer is fairly straightforward. Match your curriculum objectives with the PEPE curriculum matrix. Decide total length of exam and proportional representation by content area. Use MeSH codes and proportion data to search the Item Library. Obviously, the actual process is more difficult than the description leads one to believe; furthermore, we feel that a couple more years of effort are necessary in order to develop the system's full capabilities. Thus, the presentation should be taken as a progress report on a project for which all of us concerned are excited, not a signal of completion. However, we have data to report on the QPE which will help you analyze the quality of the system.

Analysis of the psychometric properties of the QPE is an ongoing project. My next remarks summarize the analyses of reliability and validity.

The reliability coefficients for the total score on all QPE's since May, '73 have been in the range of .95-.98 when computed by the Kuder-Richardson formula 20. As expected, the reliability of the various category scores is less than that of the total score. The reliability of categories on the November '73 and March '74 exams is presented in slide 4, as representative of usual results. Each quarter the Council on Evaluation

considers the scores on Internal Medicine, Pediatrics, Ob/Gyn, Surgery, and Basic Science for the purpose of reviewing student progress. Of these five major categories, the lowest observed reliability was .859. Recognizing the relationship between test length and reliability, the scores of minor exam categories are not considered for promotional purposes.

The preceding data indicates that the QPE has desirable properties of reliability. Now, let us consider its validity. The nature and function of the QPE is that of an achievement test; therefore, an appropriate type of validity to consider is content validity. The following points lead to a logical assumption of adequate content validity:

1. Curriculum objectives were specified and coded for content.
2. Questions of factual information were written and coded for content with the same MeSH code.
3. Each exam represents a sample of our pool of questions.

While it is safe to assume content validity of the QPE in its present form, two future refinements should improve content validity of the exam: (1) we will attempt to have questions for every curriculum objective pertaining to factual information; and, (2) exam questions will be selected for various content areas by a stratified random process.

Mindful of the difficulty in assessing content validity with statistical methods, I think data from our exam adds support to the assumption of construct validity. First, the mean performance of groups on each exam is consistently of a pattern one

would expect from an educational standpoint. In other words, mean performance has consistently been directly related to year level in school. Slide 5 illustrates this pattern for scores on the August, 1973, exam. The second example is somewhat more involved. Half of the year two students take their Ob/Gyn course Fall Semester; the other year 2 students take Pediatrics. As a group the year two students taking the Ob/Gyn course have scores significantly higher on the Ob/Gyn category than those students not enrolled in the course; on the other hand, the Peds students have scored significantly higher in Pediatrics. The two groups did not differ significantly on any other category of the examination. Although it is important to consider content validity of an achievement test, correlations with external criteria also yield significant information. A study of data from all students (N=171) conducted Spring of 1973 indicated the QPE to correlate .43, .36, and .30 with clinical evaluations, Arts and Sciences GPA, and current status in medical school, respectively. Multiple regression analysis of 20 students' scores on National Board of Medicine exam, Part 1, and their QPE total scores on the two previous exams yielded a multiple R of .827.

Many of the previous remarks dealt with psychometric characteristics of our program for computerized assessment of factual information. Sound psychometric properties are indeed desirable for any evaluation procedure. However, the practical usefulness of the assessment device must be significant enough to warrant its use, no matter how great its psychometric properties. Also,

when using a computerized system, a cost/benefit consideration becomes crucial. At this point I would like to comment on the manner in which this testing program has been integrated into the educational process. Exam scores are treated as indications of strengths and weaknesses. Generally speaking a student's low scores are not seen as low grades, but as indications of areas for future study. It is not uncommon for a student's program to be adjusted in consideration of QPE performance.

Another feature of the Item Library points to the integration of testing into the educational process at UMKC. Curriculum objectives, library reference materials, audio-visual packages and the test questions are all coded for content by the MeSH code. Use of the MeSH code facilitates the integration of curriculum, support material, and evaluation aspects of the program. This aspect of our procedure makes it possible for the Item Library to serve evaluation and learning experience functions.

(1) INFECTIOUS MONONUCLEOSIS IS OFTEN A DIFFICULT DIAGNOSIS TO MAKE WITH ANY CERTAINTY. IN ABOUT 10% - 15% OF CASES A RASH MAY OCCUR WHICH SOMETIMES COMPLICATES THE DIAGNOSIS FURTHER. THE RASH USUALLY RESEMBLES THAT OF:

1. RUBELLA
2. CHICKEN POX
3. FIFTH DISEASE
4. TINEA CORPORIS
5. ROSEOLA INFANTUM

SUBJECT: INTERNAL MEDICINE
INFECTIOUS MONONUCLEOSIS

MESH: G2.91.21.
C1.100.36.

DATE SUBMITTED: 01/72
SOURCE: 3034: WISCONSIN DATA BANK ITEM NUMBER 3034

EXAMS: 3/01/74(32)

CURRENT ANALYSIS

RESPONSES	YEAR 1	YEAR 2	YEAR 3	YEAR 4	YEAR 5	YEAR 6	FAC/STAFF
NONE	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1*	44.3	57.1	56.1	64.3	72.7	25.0	67.5
2	24.3	21.4	9.8	10.7	0.0	0.0	7.5
3	2.9	2.9	4.9	7.1	13.6	0.0	7.5
4	20.0	1.4	9.8	3.6	4.5	0.0	0.0
5	8.6	17.1	19.5	14.3	9.1	75.0	17.5

TOTAL ANALYSIS

RESPONSES	YEAR 1	YEAR 2	YEAR 3	YEAR 4
NONE	0.0	0.0	0.0	0.0
1*	44.3	57.1	56.1	64.3
2	24.3	21.4	9.8	10.7
3	2.9	2.9	4.9	7.1
4	20.0	1.4	9.8	3.6
5	8.6	17.1	19.5	14.3

* INDICATES THE CORRECT RESPONSE

MEDICAL SUBJECT HEADING

CATEGORY	MeSH CODES
Internal Medicine	G2.403.776.409
Pediatrics	G2.403.776.671
Ob/Gyn	G2.403.776.542, G2.403.776.342
Surgery	G2.403.776.909
Basic Science	
Physiology	G1.782
Biochemistry	G1.201
Anatomy	G1.100
Pharmacology	G1.703
Microbiology	G1.273.540
Etc.....	

TEXT	QUEST AVIL	QUST OSRD	DUKE HRS	DUKE PCT	SHEPPARD HRS	SHEPPARD PCT	HRS	PCT
MESH CODES								
ANATOMY & PHYSIOLOGY G-G1,100 OR G-G1,782 OR G-G7, OR G-G8, OR G-G9, OR G-G10, OR G-G11	1268		97.0	100.0	345.0	100.0		
INTEGUMENTARY SYSTEM G-A1,835 AND (G-G1,000 OR G-G1,782) OR G-G10,795	9							
SKIN A1,835 AND (G-G1,100 OR G-G1,782)	2		2.0	2.0				
ACCESSORY ORGANS (A1,835,288 OR A1,835,472 OR A1,835,657 OR A1,835,744 ORA1,835,858) AND (G-G1,100 OR G-G1,782)	1		1.0	1.0				
SKIN ABSORPTION G10,795,633	0							
BODY TEMPERATURE & REGULATION G-G7,315 OR G7,621,265	5							
MUSCULOSKELETAL SYSTEM & BODY REGIONS G-G11,427 OR (G-A1, OR G-A2,) AND (G-G1,100 OR G-G1,782)	116							
SKELETON G-A2,835 AND (G-G1,100 OR G-G1,782)	37							
UPPER EXTREMITIES (G-A1,378,209 OR G-A2,835,232,87) AND G-G1,100	8		6.0	6.1	7.0	2.0		
LOWER EXTREMITIES (G-A1,378,592, OR G-A2,835,232,484) AND G-G1,100	5		5.0	5.1	12.0	3.4		
SKULL G-A2,835,232,781 AND G-G1,100	7		3.0	3.0	20.0	5.7		
SPINE G-A2,835,232,834 AND G-G1,100	12		2.0	2.0	7.0	2.0		
THORAX (G-A2,835,232,904 OR G-A1,913) AND G-G1,100	6		3.0	3.0				
TRUNK (G-A1,47 OR G-A1,176) AND G-G1,100	11		7.0	7.2				
PELVIS G-A2,835,232,611 AND G-G1,100	2		2.0	2.0				
HEAD & NECK (G-A1,456 OR A1,598 OR G-A2,835,232,307) AND G-G1,100	6		7.0	7.2				
JOINTS G-A2,835,583 AND (G-G1,100 OR G-G1,782)	4							
BONE PHYSIOLOGY G-G7,551,121 OR G-G11,427,111 OR G11,427,156 OR G-G11,427,201	14				3.0	.8		
MUSCLES G-A2,637 AND (G-G1,100 OR G-G1,782)	20		1.0	1.0	7.5	2.1		
MOVEMENT & MUSCLE CONTRACTION G-G11,427,614 OR G11,427,661	22		1.0	1.0	7.5	2.1		

Reliability of Categories on the
November, 1973 and March, 1974 Exams

Category	November, 1973 #Items	November, 1973 Reliability[1]	March, 1974 #Items	March, 1974 Reliability[1]
Internal Medicine	178	.952	175	.949
Pediatrics	113	.907	72	.859
Ob/Gyn	91	.891	55	.889
Surgery	122	.935	92	.910
Basic Science	-	-	175	.943
Physiology	43	.829	45	.837
Biochemistry	18	.771	34	.811
Anatomy	17	.511	58	.863
Pharmacology	50	.862	54	.857
Microbiology	27	.806	35	.740
Infectious Diseases	53	.866	34	.835
Neoplasms	25	.669	27	.794
Musculoskeletal	17	.570	19	.519
Digestive	33	.804	14	.579
Urogenital	57	.838	38	.858
Pulmonary	23	.765	19	.601
Endocrine	25	.750	17	.737
Cardiovascular	30	.782	26	.645
Hemic	18	.620	15	.636
Nervous System	42	.802	16	.562
Skin Diseases	18	.456	18	.618
Nutrition & Metabolism	21	.724	38	.835
Injury, Poison, & Immunology	43	.861	48	.772
Neonatal	16	.614	12	.691
General Pathology	38	.807	30	.726
Techniques	15	.706	36	.766
Neurology	-	-	10	.506
Radiology	12	.490	10	.693
Behavioral Science	11	.465	17	.426

[1]Computed by Kuder-Richardson 20 Formula

MEAN AND RANGE OF PERFORMANCE OF VARIOUS GROUPS ON AUGUST '73 EXAM