DOCUMENT RESUME

ED 295 963 TM 011 748

AUTHOR Vu, Viet Nu; And Others
TITLE Validity and Accuracy in Methods of Scoring Clinical Performance.
SPONS AGENCY National Fund for Medical Education, Cleveland, Ohio.
PUB DATE Apr 88
GRANT 32-86A
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988). Also sponsored by NYNEX Foundation.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Abstract Reasoning; *Clinical Diagnosis; Evaluative Thinking; *Graduate Medical Students; *Scoring; *Validity
IDENTIFIERS Accuracy Measures

ABSTRACT

The process of generating and evaluating diagnoses and selecting appropriate investigative or management procedures is based on a constant acquisition, interpretation, and evaluation of critical findings. Common assessments of these processes are often limited to scoring the accuracy of the diagnoses, investigative procedures, or management decisions that a student selects or lists. The assumption is that if answers listed by a student are correct then the reasoning underlying those answers is also correct. The purpose of this study is to determine: whether this assumption is accurate, and the extent to which adding the students' reasoning processes into the scoring affects scores as compared to scores obtained solely based on the accuracy of the answers. Sixty-four second-year medical students were administered a patient problem to evaluate, and 67 fourth-year medical students were administered a different patient problem. The results show that taking into account the student's reasoning processes affected the subjects' scores differently; it reduced several of the students' scores and affected the scores of second-year students more than those of fourth-year students. This scoring process was found to provide a more accurate assessment of students' performance. (TJH)

Validity and Accuracy in Methods of Scoring Clinical Performance

Nu Viet Vu, Ph.D., Patricia A. Lee, and David E. Steward, M.D.

Southern Illinois University School of Medicine

ABSTRACT

Although it was found that the processes of generating and evaluating diagnoses, and selecting appropriate investigative or management procedures are based on a constant acquisition, interpretation, and evaluation of critical findings (Elstein et al., 1978; Barrows et al., 1982), common assessments of these processes are often limited to scoring the accuracy of the diagnoses, investigative procedures, or management decisions which a student selects or lists. The assumption is if the answers listed by a student are correct then the reasoning which underlie those answers is also correct. The purpose of this study is to determine whether this assumption is accurate, and to what extent adding the students' reasoning processes into the scoring of their answers would change the scores which they would obtain if they are solely based on the accuracy of the answers. Results based on 64 first-year and 67 fourth-year medical students indicated that taking into account the student's reasoning processes into the scoring of their answers affected differently the scores of the first- and fourth-year students, and suggested that this scoring process provided a more accurate assessment of students' performance.

This research is part of an NFME grant #32/86A, funded by the National Fund for Medical Education, and sponsored by NYNEX Foundation

This paper is present at the 1988 American Educational Research Association, New Orleans.

Validity and Accuracy in Methods
of Scoring Clinical Performance

Nu Viet Vu, Ph.D., Patricia A. Lee,
and David E. Steward, M.D.
Southern Illinois University
School of Medicine

Nu Viet Vu, Ph.D.
SIU School of Medicine
Dept. of Med. Education
P.O. Box 3926
Springfield, IL 62708

Introduction and Purpose

Written assessments of students' diagnostic workup of a patient case often consist of several tasks or measures. These typically have students select or list any of the following which pertain to the case: initial hypotheses, critical findings, investigative procedures, diagnoses, and management decisions. Existing studies have shown that there were positive but weak correlations between these measures (Elstein et al., 1978; Norman et al., 1985), suggesting that either these measures represent different and independent skills, or that the measurement itself may not assess the skills accurately, or both.

Although it was found that the processes of generating and evaluating diagnoses, and selecting appropriate investigative or management procedures are based on a constant acquisition, interpretation, and evaluation of critical findings (Elstein et al., 1978; Barrows et al., 1982), common assessments of these processes are often limited to scoring the accuracy of the diagnoses, investigative procedures, or management decisions which a student selects or lists. The assumption is that if the answers listed by a student are correct then the reasoning they use to derive those answers is also correct. One limitation of this scoring practice is that it would be difficult to assess the students' medical understanding accurately and in its entirety because the measurement used is limited to scoring students' answers without attempting to score the accuracy of the reasoning processes from which

---

This research is part of an NFME grant #32/86A funded by the National Fund for Medical Education, and sponsored by NYNEX Foundation. The paper is presented at the 1988 American Educational Research Association, New Orleans.

students derive those answers. By incorporating the scoring of the reasoning processes into the evaluation, the scorers may be able to assess more accurately students' understanding by determining whether the answers they listed are derived from appropriate lines of reasoning or whether they are based on incorrect or incomplete knowledge, or pure guessing. The purpose of this study is to determine the extent to which adding the students' reasoning processes to the scoring of their answers would affect the scores they would obtain when the accuracy of the answers alone is considered in the scoring.

Method

Two classes of students were included in this study: 64 second-year medical students were administered Patient Problem 1 (P1) at the beginning of their Introduction to Clinical Medicine course, and 67 fourth-year medical students were given Patient Problem 2 (P2) as part of their Senior Comprehensive Exam.

Patient Problem 1 (P1): The patient in P1 was a 36-year-old man who came into the office for routine history and physical examination. The patient presented with a cough and occasional shortness of breath. For this case, students had a 45-minute history and physical examination encounter with a live simulated patient (a person trained to simulate accurately a real patient), which was followed by a 15-minute written evaluation consisting of two short-answer questions. Question 1 had the students list the history and/or physical findings which they believed were critical and significant in assessing the patient's problems. Question 2 asked the students to list up to three most likely diagnoses, and for each listed diagnosis, to indicate the

specific findings which specifically suggested or supported it. Two scores were generated for this question: one score which credited one point to any correct diagnoses that were listed, and one which credited one point to a correct diagnosis only if a minimum number of correct supportive findings were listed for that diagnosis. For this study the minimum number of findings required for each diagnosis was derived from the class distribution using the mode of listed findings. The scoring key was generated by the case physician-author and reviewed by a second physician for accuracy.

Patient Problem 2 (P2): The patient in P2 was a 65-year-old woman whose primary complaint was recurrent high blood pressure. For this case the students had a 15-minute history and physical examination encounter with a live simulated patient followed by a 20-minute written evaluation. This evaluation was a two-part examination with each part consisting of two short-answer questions. In the first part, question 1 required the students to give their best evaluation of the patient's primary medical condition and to list three problems which might cause that condition. For each problem, students had to indicate the specific pathophysiological mechanisms which could explain the patient's medical condition. Question 2 asked the students to list four laboratory and/or diagnostic procedures they would order, and to include how the obtained results would help them to further evaluate or initially manage the patient's medical condition. The students had to return Part I before they received Part II.

For Part II, all students, regardless of the laboratory and/or diagnostic procedures they ordered in Part I, received the same laboratory results on the

patient. Question 3 asked the students to list three main problems or reasons which could cause the low level of potassium observed in the laboratory results of this patient. For each problem or reason listed, the students had to indicate the mechanisms which could cause the low level of potassium in the patient. Finally, Question 4 asked the students to list two investigative procedures they would order at this point to further investigate the patient's problems. Again, for each procedure, the students had to indicate how the obtained results would help them in differentiating the patient's problems. Two scores were again derived for each of the four questions: one score which credited one point to any correctly listed problems and investigative procedures, and one score which credited one point to a correct problem or investigative procedure only when it had respectively, the correct pathophysiological mechanisms or use of results listed. Again, the scoring key for the case was generated by the case physician-author and reviewed by a second physician for accuracy.

Results and Discussion

For problem P1 question 1, it was found that with a total of 26 possible significant findings, the students listed 0 to 14 findings (0% to 54%), with a mean of 8 findings (31%). The results from question 1, although not used for this study, help to better understand the results of question 2, where it was found that students used an average of 37% of the findings they listed, or a small proportion of their elicited data, to evaluate and support their hypotheses.

For P1, question 2, it was found that 14 out of 64 students did not list any correct answer or diagnosis, and 50 had at least one correct diagnosis

among their answers. By taking into account the accuracy of the supportive findings in order to credit a correct diagnosis, it was found that 27 (42%) of the students had their score changed and reduced (Table 1). With a total maximum score of three, 20 students had their score reduced by one point, 6 by two points, and 1 by three points. In other words, 74% of the 27 students did not provide a minimum number of correct findings for one of their correct diagnosis, 22% for two diagnoses, and 3% for all three diagnoses.

If students' performance in each of the four questions in P2 was assessed separately, it was found that by taking into account students' reasoning into the scoring of their answers, there were a total of 53 score changes: 18 (27%) in question 1, 16 (24%) in question 2, 14 (21%) in question 3, and 5 (7%) in question 4 (Table 1). A breakdown of the number of scores which changed or decreased by one, two, or three points for each of the questions was provided in Table 1. Overall, most scores changed by one point, then by two and three points respectively. The decrease in the number of score changes from questions 1 to 4 suggested that as the students progressed through a case and as they were given updated information on the patient, their knowledge of the patient's problems might increase and also got corrected for accuracy; consequently their correct answers might tend to be more grounded on correct reasoning.

If students' performance on all four questions in P2 was assessed altogether, it was found that 42 (63%) of the students had a score change: 32 (76%) had a score change in one of the four questions, 9 (21%) had a score change in two questions, and one (2%) had a score change in three questions.

Comparisons of the numbers and percentages of score changes in the two problems P1 and P2 indicated that the scoring which considered both the accuracy of the students' answers and of their reasoning affected more the scores of the second-year than of fourth-year students. This finding may be explained that because the knowledge of second-year students is less well structured than of the fourth-year ones, the scoring which incorporated both the accuracy of the answer and of the reasoning allows better discrimination in the second-year students' performance since not all of their correct answers are based on accurate reasoning.

Conclusions

Given that the process of assessing and managing a patient's problems involves a constant acquisition, interpretation, and evaluation of critical findings, the present study attempts to determine to what extent taking into account students' reasoning into the scoring of their answers affect the scores they would obtain when the accuracy of the answers alone is considered in the scoring. Results from this study suggested that scoring which incorporated students' reasoning into the scoring of their answers affected and reduced several of the students' scores; in addition, it affected the scores of second-year students more than those of fourth-year students. The present findings are still preliminary and need to be further replicated with a larger sample of patient problems, with similar types of questions included in the problems so that the validity and usefulness of the two types of scoring can be better assessed.

Finally, the type of assessment and scoring presented in this study is most useful to faculty who want to use evaluation for diagnostic purpose: that is to identify students who need special remediations and to determine the kind of reasoning errors students often commit.

Selected References

Barrows, H.S., Norman, G.R., Neufeld, V.R., and Feightner, J.W. The Clinical Reasoning of Randomly Selected Physicians in General Medical Practice. Clinical and Investigative Medicine, 5, 49-55, 1982.

Elstein, A.S., Shulman, L.S., and Sprafka, S. Medical Problem-Solving: An Analysis of Clinical Reasoning. Harvard University Press, Massachusetts: Cambridge, 1978.

Norman, G.R., Tugwell, P., Feightner, J.W., Muzzin, L.G., and Jacoby, L.L. Knowledge and Clinical Problem-Solving. Medical Education, 19, 344-356, 1985.

6136m/14-22
03/03/88/bfm

Table 1
Number of Score Changes in Problems 1 and 2
and Breakdown of Those Changes by Number of Points Reduced

	Problem 1 2nd year (n = 64)	Problem 2 4th year (n = 67)			
	Question 2: Diagnosis/ Findings	Question 1: Problem/ Mechanism	Question 2: Procedure/ Use of Results	Question 3: Problem/ Mechanism	Question 4: Procedure/ Use of Results
Maximum Points	3	3	4	3	3
Number of Score Changes (%)	27 (42%)	18 (27%)	16 (24%)	14 (21%)	5 (7%)
Breakdown of Score Changes (%) by Number of Points:					
1	20 (74%)	16 (88%)	12 (75%)	13 (93%)	5 (100%)
2	6 (22%)	2 (11%)	3 (19%)	1 (7%)	
3	1 (3%)		1 (6%)		

6136m/23
03/03/88/bfm