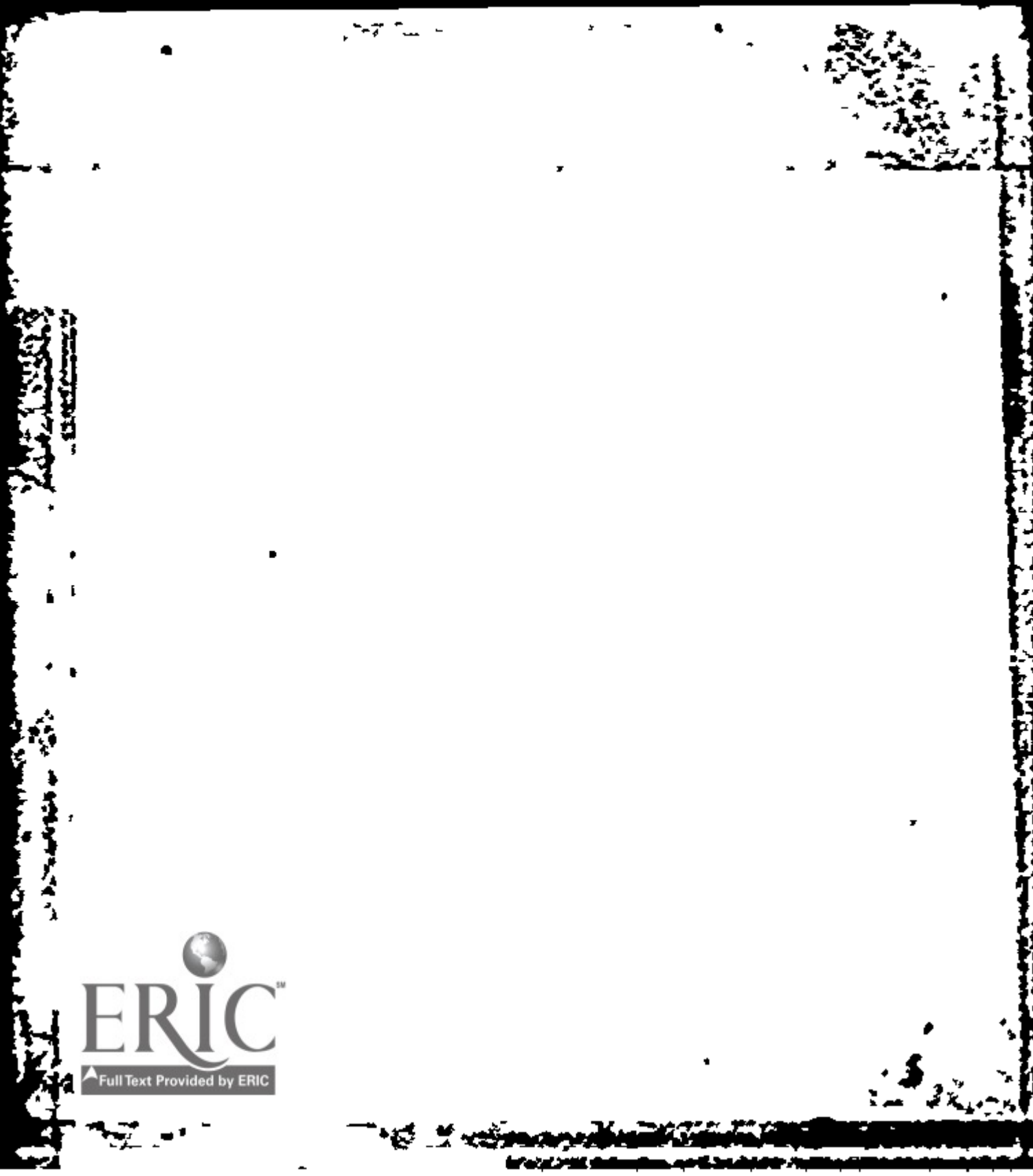

DOCUMENT RESUME

ED 107 713 TM 004 558

AUTHOR Peid, J. C.
TITLE A Survey of Evaluation Instruments Used in Clinical Clerkships in American Medical Schools.
PUB DATE Jul 74
NOTE 22p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; Feedback; Formative Evaluation; Measurement Techniques; Medical Education; *Medical Schools; *Medical Students; Objectives; Professional Education; *School Surveys; Student Evaluation; Summative Evaluation; Task Analysis; Testing; Test Reliability; Test Validity

ABSTRACT

An analysis of the evaluation instruments of clinical clerkships from 54 medical schools was made. Instruments were classified as to purpose, format, and skills measured. Thirty-nine schools used a modified Likert format; a few schools also had a check list of adjectives or short answer questions. Nearly all instruments had some space for general comments. The most frequently evaluated concepts and skills were "knowledge," "getting along well with others," "hard worker," "ability," "dignity," "history-taking," and "performance." Several principles of the design of evaluation instruments were discussed. One of these was that the instrument should be part of an evaluation system, and should evaluate the specific tasks and objectives that have been identified in the first stages of the learning system. Other principles were that the instrument should be similar to the clinical skill, it should be derived from a content sampling map, it should not be used for two purposes that have conflicting goals, and it should be reliable and valid. Several suggestions were made to increase the reliability of clinical evaluation instruments. The use of comments as a replacement for the measurement of specific objectives and content sampling was discouraged. (Author/BJG)

A Survey of Evaluation Instruments Used in Clinical Clerkships in American Medical Schools

J. C. Reid, Ph.D.
July 1974

ABSTRACT

An analysis of the evaluation instruments of clinical clerkships from 54 medical schools was made. Instruments were classified as to purpose, format, and skills measured.

One purpose of all forms was to evaluate learners. Most forms gathered data that could be incorporated into internship letters. Forms of 16 schools also provided feedback on progress to the students.

Thirty-nine schools used a modified Likert format; a few schools also had a check list of adjectives or short answer questions. Nearly all instruments had some space for general comments.

The most frequently evaluated concepts and skills were "knowledge," "getting along well with others," "hard worker," "ability," "dignity," "history-taking," and "performance."

Several principles of the design of evaluation instruments were discussed. One of these was that the instrument should be part of an evaluation system, and should evaluate the specific tasks and objectives that have been identified in the first stages of the learning system. Other principles were that the instrument should be similar to the clinical skill, it should be derived from a content sampling map, it should not be used for two purposes that have conflicting goals, and it should be reliable and valid. Several suggestions were made to increase the reliability of clinical evaluation instruments. The use of comments as a replacement for the measurement of specific objectives and content sampling was discouraged.

Although five skills were recommended to be included in clinical evaluation instruments, and the influence of national Board examinations was pointed out, it was recommended that the objectives being measured be a function of the specific objectives and constraints of the local insitution.

A Survey of Evaluation Instruments Used in Clinical Clerkships in American Medical Schools

J. C. Reid, Ph.D.
July 1974

The evaluation of medical students' performance in clinical clerkships remains a major unsolved problem, even though several articles have reported efforts to evaluate clinical students at specific medical schools. (1-6) Evaluation is one part of a learning system. Although system designs differ somewhat, nearly all learning systems contain as components, the description of the existing system, specifications of objectives, execution of a task analysis, designing the instruction, conducting a formative evaluation, and revising the instruction. (7)

The formative and summative evaluations of a clinical clerkship should reflect the objectives (8) and task analyses (9-10) of that clerkship. Accordingly, one should be able to understand other schools' objectives and task analyses of clinical clerkships by analyzing their evaluation forms.

The purpose of this report is to analyze typical summative evaluation instruments of clinical clerkships in America, and to summarize the objectives or tasks being evaluated thereby. Since no single study summarizes either learning systems or typical evaluation forms of many medical schools in the nation, the need for such a survey was plainly evident. It was assumed that although many forms might be used to evaluate a student in a clerkship, the most pertinent data for a clerkship would be recorded on a single form for the student's permanent file. Finally, this report discusses the purposes of evaluation and some design and measurement principles that could be used to improve current evaluation practices.

The author appreciates the suggestions of Dr. Jack M. Colwill. Any faults are the sole responsibility of the author.

Dr. Reid is the Research Associate of the Evaluation Section, University of Missouri-Columbia Medical School.

Evaluations can be made for several purposes: to evaluate instruction, learners, or learning; (11) to provide a data base from which internship recommendations can be written, to give the student feedback on his weaknesses and strengths, and to answer research questions.

Different purposes suggest that different items be included in the evaluation form. If the intent is to evaluate instruction, then the items on the evaluation instrument should describe the teaching, demonstrations, instructional stimuli, clarity of objectives, and the instructional environment to a major extent. To describe the measurement of instruction was not a purpose of this study; a few schools thoughtfully sent forms, usually filled out by students, that did measure instruction, but these forms are not included in the present study.

If the purpose is to evaluate learners, then the items on the instrument are dictated by the behavioral objectives and by the results of the task analysis of the clerkship. Since the required knowledge base differs across clerkships, it is not apparent how one single form can evaluate the knowledge of learners or learning for several different clinical clerkships, unless it is a summary sheet to which results from other tests are transcribed. On the other hand, the ability to collect data and to solve problems could perhaps be measured by similar instruments. The differing goals and constraints of various medical schools prevent a blanket adoption of one specific evaluation system across medical schools. However, testing efforts by national Boards will unify educational objectives among medical schools to some degree.

A third reason for evaluation is the measurement of learning. Learning is commonly measured by a pretest-posttest design (12), although in fact the evaluation of learning is not simple (13).

If the evaluation form is to generate data for internship letters, then it probably should request some descriptive vignettes that characterize the student, as well as data that will predict future success. If the purpose of evaluation is to provide the student with feedback,

then the items should measure the objectives and provide direction to facilitate improvement. The purpose of evaluation as a research tool will not be discussed in this report.

Procedure

In December 1973, a request for a copy of the forms used to evaluate basic science and clinical medical students was sent to 98 schools of medicine listed in the AAMC directory. By February 1974, replies had been received from 63 schools. Since the purpose of the study was to obtain typical ideas rather than to describe the sampling distributions of characteristics of medical evaluation forms, no attempts were made to increase the percent of replies beyond the 64% obtained by the first request.

Of the 63 schools replying, 6 schools sent no evaluation form. These 6 schools typically indicated they used letter grades and/or a sheet of comments, and these 6 were not included in the analysis. Three schools sent only basic science forms. The present report is restricted to the analysis of the evaluation forms for clinical clerkship or single forms used for both clinical and basic science years that were sent by 54 medical schools. Some schools had several forms and other schools had a form of several pages; these instances were counted as one form.

The first step in analyzing the forms was to determine the purpose of the form. For the present study, the categories of evaluating instruction and gathering data for research were ignored. It turned out that forms could not reliably be placed into the learning category, so that category was dropped. If the form indicated that comments could be made about the student, then it was judged as capable of generating characteristic vignettes for internship letters. Finally, if the form had a copy marked "student's copy," or if a phrase on the form indicated that results could be shared with students, it was classified as capable of pro-

viding feedback to students.

The second step was to determine the format of the form, whether it was a checklist, short answer, etc. The format was pertinent since it would affect the specificity of the objectives or task analysis, and the ease of using data for the purpose the form was designed.

The third step was to determine what tasks (skills) or concepts were being measured by the evaluation forms. A frequency count was made of all words on the form that related to evaluation or measurement of performance.

Words were inspected for duality of meaning. For example, "rate" (and the same root with suffix -d) could mean "grade," as in "how would you rate this student," or "degree of growth," as in "rate of progress." Thus, homonyms and homographs were sorted into different classes; synonyms were grouped into similar tasks or concepts.

Few medical school forms evaluated specific tasks. Most rather evaluated more global concepts. Although lessened specificity is not desirable for both instructional and measurement reasons, the concepts were content analyzed along with a few tasks that were described. The analysis of concepts suggested classes of tasks that typified medical school concern.

Results

The first result discussed is the purpose of the clinical evaluation forms. If the evaluation was only used to evaluate learners, that is, to grade the student, then there would be no need to give the student detailed feedback on how he performed. On the other hand, if the evaluation was also intended to improve the student's performance or direct him into areas appropriate to his strengths, then the evaluation system should in addition provide the student with detailed feedback on his weaknesses and strengths. Every one of the 54 clinical forms evalu-

ated learners, that is, determined if a learner "passed". All 54 forms included a space for remarks or comments, and were therefore judged capable of providing some characterizing vignettes, so that faculty could write descriptive internship recommendations. In addition to evaluating learners and providing information for internship recommendations, thirteen schools used their evaluation to improve the students' performance by directing their attention to specific strengths and weaknesses, since they gave students a copy of the evaluation, or made a copy readily available. Three schools apparently made the feedback optional, as their evaluation sheets carried a question like: Check if this evaluation was discussed with the student: yes no. Forms (or instructions) from four schools indicated that students in danger of failing a clerkship had extensive counsel made available to them. This practice is probably fairly common. The remaining evaluation forms did not mention whether or not the student received other information from the evaluation other than just a grade, or a pass-fail note.

The second result of the analysis of the 54 clinical forms concerns the format of the instrument. Thirty-nine medical schools used a modified Likert format for their evaluation form. A Likert format consists of a series of phrases or statements each rated on a "strongly agree, agree, . . . , strongly disagree" scale. Four examples of modified Likert formats are in Table 1. Several instruments also had an adjective check list.

Insert Table 1 about here

Fifteen schools had only short answer questions, i.e., not rating scales.

The final result described will be the frequency of tasks or concepts from the content analysis of the 54 evaluation forms. Several

hundred terms were combined into 90 tasks or concepts: 24 dealing with knowledge, 6 with interpersonal relations, 31 with personality traits, 24 with specific skills, and 5 with miscellaneous.

Table 2 groups the most frequently used concepts by frequency of

---

Insert Table 2 about here

---

occurrence in different schools. The most frequently evaluated tasks or concepts were "academic" or "knowledge;" nearly all forms requested the rater to comment specifically on the student's knowledge or understanding.

After knowledge, the next two most frequently used concepts were "gets along well with others" and "hard worker." The concept of getting along well was expressed variously: "acceptable to others," "works well on a team," and "liked by coworkers and hospital personnel" were common expressions. The phrases "human relations" and "personality" may partly bear on this concept. Being a "hard worker" was expressed by words ranging from "industrious" and "does more than his share of work" to "lazy" and "apathetic."

The next most frequently evaluated skills or concepts were "ability," "dignity," "history taking," and "performance." The concepts of ability and achievement or knowledge may overlap (14-15). The somewhat less specific words of "behavior," "bearing," "emotional maturity," and "manner" probably relate for the most part to the concept of dignity. The skill of what some schools term "history taking" may be part of what others describe as "performance." Within the label of history taking is the task of keeping charts and records.

These seven concepts complete those used by at least two-thirds of the responding medical schools. Of the eight concepts reported by 20 to

29 medical schools (about half of the schools responding) the most important is probably conducting a physical examination.

The concepts mentioned by about 20 to 40% of the responding medical schools, listed in Table 2 under the 10-19 heading, also are used to describe an ideal medical student. Such a medical student seems to be prompt and present rather than absent. When he is present, he participates. Some schools expect the ideal student to be neat in appearance. He asks questions (but, admonish some forms, asks respectfully), accepts criticism well, and is willing to do what is asked. He communicates well in spoken and written language, and he presents data or cases well. This skill in communication may stem partly from the fact that he is well organized, he reads and uses the library, he has knowledge of facts learned in basic science, and synthesizes ("correlates") and applies knowledge well. He does well at management of the patient, patient care, and patient problems. Probably most important, he is good at analyzing and solving problems.

Terms that appeared only infrequently can be pointed out. These include adaptability, anxiety, being aware of patient change, being aware of economic factors (costs) and outside agencies, discernment (this concept may be subsumed within others discussed above), having a sense of humor, being prepared (specifically mentioned by only one school), and the two skills of following up and listening, both quite important.

Discussion

Six purposes of evaluation were described. No form was classified as evaluating learning, instruction, or as being used for research purposes for the reasons given above.

Evaluation instruments for learners should approximate the specific skills being measured. An evaluation may be potentially destructive to the educational process, particularly if students study for the examina-

tion rather than for the acquisition of the skill. Consequently several studies have used simulation techniques, actors as patients, and special testing methods (3, 16-28) so that the examination closely approximates the performance. Further support to the idea that examination should resemble performance is lent by reports that physician performance is predicted better by length and type of internship and residency than by medical school ratings. (29, 30) As the evaluation format approaches in similarity the actual clinical skill, then students increase their efforts toward acquiring the skill, and decrease sycophantic behavior toward the persons whom they perceive will subjectively rate them.

The measurement of learning necessarily requires assessments at two or more points over time. None of the 54 clinical forms were classified as measuring learning because such classification could not be done reliably. Nevertheless, it seemed that the great majority of the forms measured learners, and few measured learning. Surely the evaluation of learning would be a priori as important as the evaluation of learners. David P. Ausubel wrote on the first page of one of his books, ". . . the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly." (31) Several evaluation forms called for a subjective estimate of students' "growth," which is probably a request for the evaluation of learning. To be useful and reliable, learning should be measured objectively, not subjectively. Again, instead of a subjective estimate of students' "potential," a more reliable, objective estimate might be obtained from a discriminant analysis, comparing scores of present students with scores on that same instrument of previous students who later became satisfactory and satisfied physicians in a certain career. It is possible of course that a subjective rating of learning on an evaluation form might in actuality be a summary of objective scores.

If the purpose of the summative evaluation form is to provide data from which internship recommendations will be written, then the evaluation might result from compiling critical incidents (6, 26, 32-41) de-

rived from faculty or physician perceptions which are then categorized by content area. The construction of a content sampling map will assure that all important objectives are discussed in their proper proportion. It should be emphasized that getting a "wide variety of important topics" (42) does not satisfy the requirements of content sampling in terms of objectives and tasks.

If the goal is to give the student feedback on his strengths and weaknesses, then the items need to be translated back into behavioral objectives so the student can readily tell which objectives he achieved and which he did not.

If a single evaluation form is to have several purposes, then it probably will have several sections, each of which is processed differently. A single instrument that has two purposes (such as the evaluation of learners and the providing of student feedback) may serve neither of them well. Years ago the army discovered that captains were reluctant to give lieutenants low ratings, or to indicate a poor performance, because the lieutenants would find out who had rated them low. One attempt to overcome this was described by Sisson (43). If an evaluation form is used for the two purposes of evaluation of learners and for student feedback, then medical faculty may not make wholly honest ratings, and students who would benefit from frank counsel may never receive it.

Before closing the discussion it is well to emphasize two important principles of measurement, reliability and validity. If a test measuring knowledge is reliable, then if Pete scored high in knowledge on Monday, then he will also score high in knowledge on Wednesday. Good measurement suggests that whatever is measured be measured reliably. Five principles affecting reliability were often violated on the evaluation forms. The first principle is specificity. Vague concepts such as "personality," "human relations," "habits," "bearing," "manner," and "social" cannot be measured reliably by ratings, since Dr. Jones' notion of "personality" differs from Dr. Smith's. Forms using these vague

terms will produce unreliable data. "Ability," "performance," and "skills" are useful concepts, but are wholly unsatisfactory when used as stems for single-item rating scales.

The second principle is that measurement should be as objective as possible. It is surprising that knowledge, which is not difficult to measure by multiple-choice tests, would so frequently occur as a subjective estimate. A subjective rating is probably the least reliable measure of knowledge; nevertheless, several evaluation forms required faculty members to indicate their subjective appraisal of the amount of soneone's knowledge. A more reliable measure of knowledge than subjective rating is an objective measurement: a written test, or the performance of a standard clinical task before a trained observer, or the student's reaction to simulated stimulus. (4) In fact, relevant objective measures should be obtained instead of subjective ratings wherever possible. (44, 45) It is possible that the "knowledge" scale on some evaluation forms is a summary index of students' weighted total scores on objective knowledge instruments. Two or three schools enclosed samples of their objective instruments to measure knowledge. It is important to recognize that the following statement is not true: "any method producing more objectivity with respect to the evaluation of medical students provides measurable advantages. . . ." (46, p. 345). Objectivity is better only if it is pertinent to the behavioral objectives and the results of the task analysis. Height of students is an objective measure, but it is not relevant to clinical tasks.

Third, a person should not rate a quality or skill unless he has observed it first hand. (41, 47) Courts recognize this recommendation by distinguishing between acceptable evidence and hearsay. Some rating forms have a column "don't know" or "no chance to observe." Several evaluation forms correctly instructed the rater not to evaluate any trait or skill that he had no personal knowledge of.

Fourth, questions or scales can be most reliably responded to if

each scale measures only one trait. Thus a more reliable form would have a format such as example I, III or IV in Table 1, and would not have a format such as example II in Table 1, in which several concepts are represented by a single scale, and the anchors (descriptors) at one end of the scale are not opposites in meaning of the anchors at the other end of the scale. An exception would occur if reliable factor analyses indicated that different scales loaded highly on the same trait.

Fifth, the skills being measured should have been previously well defined, and evaluation be limited to that well-defined set. (2) An evaluation instrument, after it has passed the developmental stage, should consist of scales or items that reflect the well-defined content being measured, rather than consist of large blank spaces for unfettered English comments. The use of printed scales will assure that the agreed-upon objectives are being considered by raters, and a weighted score can be quickly computed and compared against previously-agreed upon criteria of pass, fail or A, B, C, D. Although it is meritorious to measure unobtrusively where possible, (48) particularly in the affective domain, the use of patient records to evaluate clinical performance has not been satisfactory (36, 49) probably because of the wide variability in record keeping.

The widespread use of comments is difficult to understand when it is generally acknowledged that comments are time-consuming for faculty to make and are time-consuming to analyze and evaluate. Although comments are widely used in the developmental stages of an instrument, they are typically not retained in the final revision.

Comments could perhaps be justified in a final version of the instrument in three instances. First, comments might record highly unusual events (e.g., "student is deaf but nevertheless achieving satisfactorily"). Second, if a copy of the instrument is to be given to the student, then comments could be included to advise a particular student, e.g., "review Peters Chapter 2, particularly the 2nd and 3rd sections." Third, comments may provide descriptive vignettes for an internship recommenda-

tion letter. Again, higher reliability would result if these vignettes could be related to a content map of agreed-upon critical incidents that described unacceptable or acceptable practices.

Other than these three instances, comments should not be encouraged. A student will learn as much by seeing himself checked high on a scale of "hard-working and industrious" as he will be seeing a hand-written comment, "Joe is hard-working and industrious." A faculty member might not write that phrase in longhand, even though it might be appropriate.

A second important measurement principle is validity. A validity index will indicate the degree to which the instrument measures what it should. Presumably, if the instrument has been designed as part of a learning system as outlined herein, the instrument will have at least respectable content and construct validity, although this presumption should be tested rather than assumed. In general, measures of clinical evaluation have not predicted physician performance, partly because few clinical performance instruments are either reliable or valid, partly because it is difficult to decide what constitutes satisfactory physician performance, and partly as Price, Taylor and others point out, physician performance is not unidimensional. (50, 59) Some Boards have made an effort to improve the reliability and validity of their assessment techniques (23, 26, 30, 35, 42, 60-64).

The results of this survey combined with the published literature suggest that the following dimensions are important in the measurement of clinical competence: knowledge about disease, ability to collect data (including physical examinations and taking histories), ability to identify and solve problems, maintenance of an appropriate relationship with patients and colleagues, and commitment to get the job done (a combination of hard work and efficiency).

Nearly every school in the survey stressed the importance of knowledge. The evaluation of knowledge should be based on the task analysis and the behavioral objectives, and should be objective rather than subjective. The ability to collect data also appeared on many evaluation

instruments; relevant literature is cited elsewhere in this survey. Data collection skills have been measured by observation and critical incidents techniques; such skills should be rated objectively against pre-specified criteria (2, 34, 65).

Several schools did not explicitly measure problem-solving ability, yet this seems to be a crucial trait for successful physicians to have. Although important work has been done by Rimoldi, Elstein, McGuire, and others as cited herein, in general the large literature on the measurement and evaluation of problem-solving is not apparently implemented in clinical settings.

The maintenance of an appropriate relationship with patients and coworkers, and a commitment to get the job done occurred quite frequently in the present sample of clinical evaluation instruments and is supported in the literature (4, 34, 65, 66). Sociometric ratings and peer ratings on checklists will produce more reliable results than will random observations.

The purpose of this study was to describe some common existing evaluation processes of clinical blocks of American medical schools. The cooperation of those schools participating in the study is greatly appreciated since both the satisfactory and the unsatisfactory practices helped to clarify and sharpen issues that were not universally recognized in the medical evaluation literature. By reporting the clinical evaluation processes in over fifty medical schools, this study should make it easier for an institution to make the first step in a learning system, that of analyzing their existing system. Sufficient evidence has been cited to demonstrate that those who design evaluation instruments independently of a learning system that encompasses steps of specifying objectives and analyzing tasks will fail. This study has reported 5 clinical competencies that appear to be regarded as basic, as well as numerous ancillary concepts. Objectives of insitutions (1, 2, 4-6) and of national Boards may also be useful, but an instrument should reflect the objectives and constraints of a local institution and not someone else's.

Table 1

Examples of the format of evaluation forms

I.

	Don't Know	Superior	Excellent	Average	Adequate	Inadequate
1. Knowledge						
2. Clarity of thinking						
.						
.						

II. 1. Intellectual process of obtaining information by history, physical exam and lab studies

Don't know Poor 1 2 3 4 5 6 7 8 9 Excellent

Ignores leads, uses time inefficiently, has no clear rationale for procedures, questions, or orders . . . .

Avoids stereotyped approaches, correlates facts into meaningful whole, discriminates in choice of questions, tests and procedures . . . . .

. . .

III.

	Low				High
	1	2	3	4	5
1. Knowledge					
.					
.					
.					

IV.

Medical Knowledge					
☐ Not observed	☐ shows very inadequate knowledge of medical situations related to the patient's problems.	☐ Shows less than minimal amount of knowledge of medical situations related to the patient's problems.	☐ Shows adequate comprehension of basic medical principles and relates them to the patient's problems.	☐ Shows complete knowledge of the basic medical principles and displays great insight in relating them to the patient's problems.	☐ Demonstrates intellectual agressiveness which qualifies him as the local expert on the patient's problems.

Table 2

Frequency of Different Schools Using Certain
Concepts and Skills of Clinical Clerkship Evaluation Forms

>50	40-49	30-39	20-29	10-19
academic	gets along well with staff	ability	interest	absent
	hard worker	dignity	physical exam	analyze
		history taking	potential	appearance
		performance	progress	application of knowledge
			reliable	asks
			responsibility	basic science, knowledge of
			skills	cognition
			strength & weakness	communicate
				complete
				correlation
				criticism
				data, gather
				emotional maturity
				important
				integrity
				judgment
				lab skills & tests, knows
				management of patient
				organized
				participation
				patient care
				present data or case
				problems, patient
				problem solving
				reads
				willingness

References

1. Geertsma, R. H. and Chapman, J. E. The Evaluation of Medical Students. J. Med. Educ., 42:938-948, 1967.

2. Hinz, C. F., Jr. Direct Observation as a Means of Teaching and Evaluating Clinical Skills. J. Med. Educ., 41:150-161, 1966.

3. Rimoldi, H. J. A. Rationale and Applications of Test of Diagnostic Skills. J. Med. Educ., 38:364-368, 1963.

4. Oaks, W. W., Scheinok, P. A., and Husted, F. L. Objective Evaluation of a Method of Assessing Student Performance in a Clinical Clerkship. J. Med. Educ., 44:207-213, 1969.

5. Odom, G. L. Neurological Surgery and the Assessment of Accomplishment. J. Med. Educ., 44:784-790, 1969.

6. Graham, J.R. Systematic Evaluation of Clinical Competence. J. Med. Educ., 46:625-629, 1971.

7. Davis, R. H., Alexander, L. T., and Yelon, S. L. Learning System Design. New York: McGraw-Hill, 1974.

8. Merrill, M.D. Content and Instructional Analysis for Cognitive Transfer Tasks. AV Communication Review, 21:109-125, 1973.

9. Okey, J. R. Developing and Validating Learning Hierarchies. AV Communication Review, 21:87-108, 1973.

10. Gagné, R. M. The Conditions of Learning. 2nd ed. New York: Holt Rinehart and Winston, 1970.

11. Wittrock, M. C. The Evaluation of Instruction: Cause-and-effect Relations in Naturalistic Data. In The Evaluation of Instruction. Wittrock, M.C., and Wiley, D. E. (Eds.). New York: Holt Rinehart and Winston, 1970. Pp. 3-24.

12. Taranger, J. Comparison of the knowledge of Students at the Beginning and the End of a Course in Pediatrics. Brit. J. Med. Educ., 7:122-128, 1973.

13. Harris, C. W. (Ed.). Problems in the Measurement of Change. Chicago: University of Chicago Press, 1963.

14. Ferguson, G. A. On Learning and Human Ability. Can. J. Psychol., 8:95-111, 1954.

15. Ferguson, G. A. On Transfer and the Abilities of Man. Can. J. Psychol., 10:121-131, 1956.

16. Jason, H. Self Instruction in Medical Education: Principles, Practices, Prospects. Brit. J. Med. Educ., 2:20-23, 1968.

17. Elstein, A. S. et al. Methods and Theory in the Study of Medical Inquiry. J. Med. Educ., 47:85-92, 1972.

18. Hoffer, E. P., Barnett, G. O., and Farquhar, B. B. Computer Simulation Model for Teaching Cardiopulmonary Resuscitation. J. Med. Educ., 47:343-347, 1972.

19. Levine, H., and McGuire, C. Role Playing as an Evaluative Technique in a Certifying Examination. J. Med. Educ., 42:264-265, 1967.

20. Barrows, H. S., and Abrahamson, S. The Programmed Patient: A Technique for Appraising Student Performance in Clinical Neurology. J. Med. Educ., 39:802-805, 1964.

21. Abrahamson, S., Denson, J. S., and Wolf, R. M. Effectiveness of a Simulator in Training Anesthesiology Residents. J. Med. Educ., 44:515-519, 1969.

22. Freeman, R. B. A Computer Program for Simulating the Patient-physician Encounter. J. Med. Educ., 48:92-97, 1973.

23. Goran, M. J., Williamson, J. W., and Gonnella, J. S. The Validity of Patient Management Problems. J. Med. Educ., 48:171-177, 1973.

24. Penta, F. B., and Kofman, S. The Effectiveness of Simulation Devices in Teaching Selected Skills of Physical Diagnosis. J. Med. Educ., 48:442-445, 1973.

25. Harless, W. G. et al. CASE: A Computer-aided Simulation of the Clinical Encounter. J. Med. Educ., 46:443-448, 1971.

26. Levine, H. G., and McGuire, C. H. The Use of Role-Playing to Evaluate Affective Skills in Medicine. J. Med. Educ., 45:700-705, 1970.

27. Lamont, C. T., and Hennen, B. K. E. The Use of Simulated Patients in a Certification Examination in Family Medicine. J. Med. Educ., 47:789-795, 1972.

28. Andrew, B. J. An Approach to the Construction of Simulated Exercises in Clinical Problem-solving. J. Med. Educ., 952-958, 1972.

29. Peterson, O. L., Lyden, F. J., and Geiger, H. J. Appraisal of Medical Students' Abilities as Related to Training and Careers After Graduation. N. Engl. J. Med., 269:1174-1182, 1963.

30. Clute, K. F. The General Practitioner: A Study of Medical Education and Practice in Ontario and Nova Scotia. Toronto: University of Toronto Press, 1963.

31. Ausubel, D. P. Educational Psychology. New York: Holt Rinehart and Winston, 1968.

32. Flanagan, J. C. The Critical Incidents Technique. Psychol. Bull., 51:327-358, 1954.

33. Diederich, P. B. The "Critical Incidents" Technique Applied to Medical Education. Princeton, N. J.: Educational Testing Service, May 17, 1954. RM-54-9.

34. Hubbard, J. P. et al. An Objective Evaluation of Clinical Competence. N. Engl. J. Med., 272:1321-1328, 1965.

35. Miller, G. E. The Orthopedic Training Study. J.A.M.A., 206:601-606, 1968.

36. Brumbach, G. B., and Howell, M. A. Rating the Clinical Effectiveness of Employed Physicians. J. Appl. Psychol., 56:241-244, 1972.

37. Cowles, J. T. A Critical-comments Approach to the Rating of Medical Students Clinical Performance. J. Med. Educ., 40:188-198, 1965.

38. Newell, F. W. et al. The Multiple-choice Question Test of the American Board of Ophthalmology. Am. J. Ophthalmol., 71:57-61, 1971.

39. Sanazaro, P. J., and Bates, B. A Joint Study of Teaching Programs in Comprehensive Medicine. J. Med. Educ., 43:777-789, 1968.

40. Sanazaro, P. J., and Williamson, J. W. Physician Performance and its Effects on Patients: A Classification Based on Reports by Internists, Surgeons, Pediatricians, and Obstetricians. Med. Care, 8:299-308, 1970.

41. Cowles, J. T., and Kubany, A. J. Improving the Measurement of Clinical Performance of Medical Students. J. Clin. Psychol., 15:139-142, 1959.

42. Burg, F. D., and Wright, F. H. Evaluation of Pediatric Residents and their Training Programs. J. Pediatr., 80:183-189, 1972.

43. Sisson, E. D. Forced-choice--the New Army Rating. Personnel Psychol., 1:365-381, 1948.

44. Tinkelman, S. N. Planning the Objective Test. In Educational Measurement. Thorndike, R. L. (Ed.). 2nd ed. Washington, D. C.: American Council on Education, 1971. Pp. 46-80.

45. Wesman, A. G. Writing the Test Item. In Educational Measurement. Thorndike, R. L. (Ed.). 2nd ed. Washington, D. C.: American Council on Education, 1971. Pp. 81-129.

46. Printen, K. J., Chappell, W., and Whitney, D. R. Clinical Performance Evaluation of Junior Medical Students. J. Med. Educ., 48:343-348, 1973.

47. Barro, A. R. Survey and Evaluation of Approaches to Physician Performance Measurement. J. Med. Educ., 48:1047-1093, 1973.

48. Webb, E. et al. Unobtrusive Measures. Skokie, Ill.: Rand McNally, 1966.

49. Fessel, W. J., and Van Brunt, E. E. Assessing Quality of Care from the Medical Record. N. Engl. J. Med., 286:134-138, 1972.

50. Howell, M. A., Cliff, N., and Newman, S. H. Further Validation of Methods for Evaluating the Performance of Physicians. Educ. Psychol. Meas., 20:69-78, 1960.

51. Turner, E. V., Helper, M. M., and Kriska, S. D. Predictors of Clinical Performance. J. Med. Educ., 49:338-342, 1974.

52. Price, P. B. et al. Attributes of a Good Practicing Physician. J. Med. Educ., 46:229-237, 1971.

53. Korman, M., and Stubblefield, R. L. Medical School Evaluation and Internship Performance. J. Med. Educ., 46:670-673, 1971.

54. Clute, K. F. The General Practitioner: A Study of Medical Education and Practice in Ontario and Nova Scotia. Toronto: University of Toronto Press, 1963.

55. Price, P. B. et al. Performance Measure of Physicians. Salt Lake City: University of Utah, 1963.

56. Peterson, O. L. et al. An Analytical Study of North Carolina General Practice: 1953-1954. J. Med. Educ., 31:1-165, 1956.

57. Fisher, A. W. Patients' Evaluation of Outpatient Medical Care. J. Med. Educ., 46:248-244, 1971.

58. Crosbie, S., and Gilberstadt, H. Contrasts between Several Means of Appraising Physicians. J. Med. Educ., 36:1310-1315, 1961.

59. Price, P. B. et al. Measurement and Predictors of Physician Performance. Salt Lake City: Aaron Press, 1971.

60. Levine, H. G., and McGuire, C. H. The Validity and Reliability of Oral Examinations in Assessing Cognitive Skills in Medicine. J. Educ. Meas., 7:63-74, 1970.

61. Gregory, L. Orthopaedics and the Impact of Learning Theory. J. Med. Educ., 44:777-783, 1969.

62. Levine, H. G., and McGuire, C. H. Rating Habitual Performance in Graduate Medical Education. J. Med. Educ., 46:306-311, 1971.

63. Levit, E. J. Intraining Evaluation of Learning. Arch. Dermatol. 99:342-349, 1969.

64. Petersdorf, R. G. et al. The New Procedure for Evaluating the Clinical Competence of Candidates to be Certified by the American Board of Internal Medicine. Ann. Intern. Med., 76:491-496, 1972.

65. Elrick, H. The Clinical Evaluation of the Medical Student. J. Med. Educ., 42:453-459, 1967.

66. Hess, J. W. A Comparison of Methods for Evaluating Medical Students Skills in Relating to Patients. J. Med. Educ., 44:934-938, 1969.