

February 1982

## A GUIDE TO CLINICAL PERFORMANCE TESTING

NEAL WHITMAN  
UNIVERSITY OF UTAH MEDICAL CENTER\*

### INTRODUCTION

College teachers often are faced with the task of evaluating the ability of students to apply knowledge and skills to real or simulated problems. The general term for this type of evaluation is performance testing. Performance tests are used in many areas such as music, art, physical education, foreign languages, drama, and the biological, physical, social, and medical sciences. McBeath and Lassen (1978) point out that the scope of performance testing is broad and far reaching, with applications in many aspects of university education which involve students in intellectual processes beyond memorization. This article addresses performance testing in the "natural setting." Specifically, attention is given to clinical evaluation in the health professions. Although this discussion and the examples provided relate directly to the health professions, college teachers of other disciplines should find many of the ideas generalizable to their own areas.

### The Competence Question

Evaluation of students seeks to answer the question, "Is this individual competent?" The obvious prior question must be, "What is competence?" Saying that someone is competent implies an agreement as to what is sufficient or adequate behavior. This understanding of competence goes to the root of evaluation. According to House (1980), at its simplest evaluation leads to a settled opinion that something is the case. In the clinical setting, teachers are responsible for judging whether students are competent according to agreed upon standards of performance. This is never an easy task.

### The Control Problem

The evaluation of health-profession students in a clinical environment or of other types of students in natural settings is one of the most difficult tasks faced by an evaluator. Its difficulty lies in the attempt to measure simultaneously cognitive, psychomotor, and affective changes, all of which are interwoven and performed together. There are also the difficulties of sampling student performance over time and of controlling test conditions. The health care delivery system itself dictates testing conditions (e.g., the type and number of patients, the facilities and equipment at the students' disposal, and the availability of auxiliary staff). No wonder Irby and Dohner (1976) noted, "Evaluation of student clinical performance is one of the most complex areas of the teaching/learning process in the health professions." (p. 208)

To help clinical educators who test students in the clinical setting, the three basic steps of performance testing will be described: 1) establishing the purposes, 2) describing the desired performance characteristics, and 3) measuring student performance.

### STEP ONE: CLARIFYING THE PURPOSES

The first step in performance testing is to establish the purposes of the evaluation. According to Scriven (1967) there are two types of program evaluation, formative and summative. These same terms can also be applied to the evaluation of students. Formative evaluation is conducted during the instructional process. Its purpose in performance testing is to improve student performance. Summative evaluation is conducted at the end of the instructional process. Its purpose is to rate. Often college teachers think of evaluation of students in terms of assigning a letter grade at the end of a course. However, college teachers should also evaluate students for the formative purpose of student improvement. Recognition that performance testing serves both functions has many implications for the teacher.

### The Need for Feedback

The justification for formative evaluation of student performance is well-timed feedback. In general, feedback is most useful at the earliest opportunity after the given performance. In the clinical setting, other students and the patient are likely to be present if feedback is given immediately after the student's performance. Feedback given in front of other students and/or a patient should be descriptive rather than judgmental. In fact, according to one study, third- and fourth-year medical students identified "correcting students when wrong without belittling" as one of the most helpful clinical teaching behaviors (Stritten, et.al., 1975, p. 879).

Propriety dictates that some feedback should be given away from patients, and sometimes even away from other students. Many times, even if feedback in front of others is appropriate, the press of time precludes giving immediate feedback, and later sessions should be scheduled. Regardless of when or where feedback is given, it should be specific. Telling a student nurse, "You should take twice as much time when you ask a newly diagnosed diabetic patient to show you how he draws his insulin," is more instructive than "You're too abrupt with diabetic patients."

Furthermore, feedback should be directed at behavior the student can do something about. This is particularly critical in the clinical setting where there are many providers of care. Feedback given to a student nurse about late meals is not helpful if the tardiness originates in the dietary department. Finally, a good rule of feedback is to ask the student to rephrase information to make sure it is understood.

\*Dr. Whitman was formerly with Michael Reese Hospital and Medical Center, Chicago.

All teachers are familiar with summative (rating) evaluation. In the clinical setting, because teachers are responsible for certifying the competence of students, summative evaluations are important. Therefore, clinical teachers 1) must be committed to evaluations which are relevant, rigorous, and understandable, and 2) must be able to produce records that are valid, reliable, and believable (Patton, 1980). No wonder that evaluating students in the clinical setting strikes many teachers as an overwhelming task! It may be helpful to think of summative evaluation as the collection of evidence that allows teachers to have confidence in the judgments they make. The key to this confidence is a multiplicity of data sources. The more sources of data, the more confident faculty can be in the summative evaluation judgments they make about the clinical competence of students. The use of multiple data sources will be explored in the third step of performance testing-- Measuring Student Performance.

## STEP TWO: ESTABLISHING PERFORMANCE CHARACTERISTICS

In order to evaluate the clinical performance of students, the characteristics of the performance must be established. Characteristics of desired clinical performance are best described as goals and objectives.

### Setting Goals

Goals should serve as statements of the ideals toward which students are moving. Goals clarify what is to be measured by the evaluation. An example of a helpful goal is one provided by the medical director of an adolescent medicine program for pediatric residents (Klein, 1980). The adolescent medicine goal is "to provide pediatric housestaff with learning experiences so that they will incorporate primary care services to adolescents in their future practices" (no page number). There should be no doubt regarding what the adolescent medicine program is aiming for, or what the evaluation seeks to measure.

To reach its goal, the adolescent medicine program provides many classroom and clinical activities. One clinical activity is the adolescent medicine clinic. Its goal is: "To provide second year pediatric residents with learning activities so that they will become more proficient and more comfortable with outpatient adolescents." (Klein, 1980, no page number). Again, there should be no doubt regarding the clinic's intention.

### Defining Objectives

The purpose of clinical evaluation is to judge whether students are moving toward the specified goals. In this context, objectives are statements of evidence that there is movement toward a goal. Objectives describe cognitive, psychomotor, and affective changes which are indicators of students moving toward an ideal. In the adolescent medicine clinic previously described, cognitive and psychomotor changes were combined into "proficiency" objectives. In this way "knowing" and "doing" are interwoven when students perform. The performance objectives in the adolescent medicine clinic are: 1) Take an adolescent history; 2) Perform an adolescent examination and make appropriate assessments; 3) Perform pelvic examination when indicated and perform appropriate auxiliary tests; 4) Counsel adolescents regarding contraceptives; 5) Assess sexual maturation using Tanner stages; and 6) Diagnose and treat vaginitis, venereal disease, menstrual irregularities, and dysmenorrhea.

In addition, affective objectives were described as "comfort" objectives since the goal of the clinic called for the residents to become more comfortable with adolescents. The comfort objectives in the adolescent medicine clinic are for residents to become increasingly: 1) sensitive to adolescent needs, 2) independent of faculty, 3) self-directed to learn more about adolescent medicine, 4) interested in gaining more experiences as a resident with adolescents, 5) desirous of incorporating adolescents in future practice, and 6) able to develop a confidential relationship with clients and able to maintain and respect the privacy of clients.

### Using Objectives

Objectives constitute evidence of movement toward a goal. Thinking of objectives as evidence precludes thinking of them as ends. They are means to an end, namely the goal. In the adolescent medicine clinic, measures of proficiency and comfort objectives collectively constitute evidence that house staff are more proficient and comfortable with outpatient adolescents.

Thus, after establishing the summative and formative purposes of performance testing, the second step is to describe the desired performance characteristics in terms of goals and objectives. The objectives must be measurable and constitute evidence of goal attainment. Teachers know the objectives are comprehensive when attainment of all of them instills confidence that the student has successfully moved toward the goal.

## STEP THREE: MEASURING STUDENT PERFORMANCE

Whether evaluating for improvement or for rating, the third step in performance testing is the most difficult. The clinical teacher is faced with the task of measuring proficiency (cognitive and psychomotor) and affective objectives which are being performed simultaneously. Four tools which can help clinical teachers test performance are checklists, observation logs, critical incidence forms, and anecdotal records.

### Using the Checklist

A checklist is a breakdown of a performance into specific steps. Common errors are also listed to make observation easier and documentation less time-consuming. Irby and Dohner (1976, pp. 216-217) list seven characteristics of effective checklists:

1. Clear and adequate instructions for administration.
2. Concisely stated terms written in observable, behavioral terms.
3. Only critical items used in order to keep the form short.
4. Option for non-observed items provided by adding a "not observed" category.
5. Objective criteria used for items.
6. Unit objectives accurately reflected in the items.
7. Items arranged in a natural order of sequence for performing the skill.

In addition to using checklists for proficiency objectives, these also can be used for affective objectives. An affective checklist consists of cues which are observable behaviors expected of someone who is attaining the objective. For example, one of the comfort objectives of the adolescent medicine clinic was "the resident should become increasingly self-directed to learn more." The checklist of cues includes: 1) brings relevant articles to the attention of the teacher, 2) identifies own areas of weakness, 3) asks in-depth questions, and 4) goes beyond the individual patient and generalizes to other patients.

#### Using Observation Logs

Observation logs allow each objective to be pre-printed with additional space for writing notes (Knopke and Diekelman, 1978). Completeness and accuracy of performance should be noted. The difference between a checklist and an observation log is that the checklist pre-establishes what to observe. How does a teacher know which to use? Generally, checklists are advantageous when performance skills can be broken down into a small number of steps (e.g., drawing insulin into a syringe). Observation logs are advantageous when the performance skills are more complex (e.g., teaching a patient how to draw insulin into a syringe). One can anticipate that teaching is a complex performance which cannot be broken down easily into a small number of specific steps. Both checklists and observation logs are used to assess objectives.

#### Using Critical Incident Forms

Critical incident forms are used to identify one specific aspect of performance (e.g., "introduces self to patient with courtesy and warmth"). When a positive or negative example is observed, a description of the situation is documented on the critical incident form. (See Igalbe and Spears, 1979, for one example.)

#### Using Anecdotal Records

Anecdotal records provide brief descriptions of any situations which seem relevant to the evaluation of students (Knopke and Diekelman, 1978). The anecdotal record is similar to the critical incident form, but aspects of performance to be observed are not pre-established. The critical incident forms and anecdotal records supplement the checklists and observation logs which cannot cover all aspects of performance.

What all four tools have in common is that these are concurrent measures. That is, these data are collected and recorded at the same time that the student is performing. If the record keeping cannot be done at the time of performance, it should be done soon after. Reliability is lost as time passes.

The relationship among these four concurrent tools is shown in Table 1. Checklists and observation logs assess objectives; critical incident forms and anecdotal records assess other aspects of student performance. Checklists and critical incident forms use pre-established behaviors to be observed; observation logs and anecdotal records are open-ended.

Table 1. Concurrent Tools

	Specifics Pre-established	Specifics Not pre-established
O B J E C T I V E S	CHECKLISTS	OBSERVATION LOGS
O T H E R	CRITICAL INCIDENT FORMS	ANECDOTAL RECORDS

#### Using Retrospective Methods

In addition to those four concurrent tools, there are two retrospective tools which measure student performance. These are medical records and faculty surveys.

Medical records provide quantitative evidence that certain actions have been taken (e.g., a medical student has performed a physical examination). However, one limitation of medical records is that they do not indicate the quality of performance.

Scott and Sniderman (1973) report the use of medical charts to evaluate the contributions of individual residents in general medical service. Checklists were established and the chief resident and attending physician on service independently scored each record for the number of problems identified at the time of admission; omissions in the medical history or physical examination; or problems not detected, problems not adequately resolved, or major errors.

The authors report that information from studies such as the one they described have at least four implications for evaluation: 1) errors in patient care can be detected and corrected; 2) medical students and residents can receive feedback regarding their clinical deficiencies; 3) clinical teachers can assess strengths and weaknesses in the teaching program; and 4) certifying bodies can be better informed regarding the clinical experience of their candidates.

Faculty surveys provide an accumulation of qualitative evidence. Teachers who observe medical students conducting physical examinations can rate the manner of performance according to the characteristics faculty deem desirable. Faculty surveys provide a measure of the amount of agreement among faculty about student perfor-

mance. An additional advantage of faculty rating surveys is that they can be developed to measure rather complex aspects of student performance. Rating scales are very helpful in collecting faculty survey data and provide an indication of the degree to which student performance is successful or unsuccessful. Rating scales are easy to use and to score, and take little time to complete. (See Irby and Dohner, 1976, for an example).

A limitation of faculty survey data for summative purposes is pointed out by Marienfeld and Reid (1980) who compared the Internal Medicine faculty ratings received by medical students at the end of their third year to the scores they received on the Internal Medicine subsection of the National Board of Medical Examiners Test. In general, they found that many faculty tend to overestimate the knowledge level of students. One explanation may be that faculty tend to rate students on characteristics not related to the performance being assessed. For example, "being friendly" is not the same as "instilling confidence." Yet, a teacher may overestimate a student's ability to instill confidence because she/he is friendly.

#### Using Multiple Data Sources and Instruments

The key to feeling confident about judgments of student performance is multiple data sources (Guba and Lincoln, 1981). Every evaluation tool has both advantages and disadvantages. Faculty tend to "see" clinical performance through their own clinical "spectacles." And, individual teachers "see" differently from day to day. Therefore, summative evaluations from a single faculty member should not be relied upon. Open-ended data tools, such as observation tools and anecdotal records, allow flexibility in recording relevant, observable behaviors. Close-ended tools, such as checklists, critical incident forms, and faculty rating surveys, alert evaluators to critical aspects of student performance. The use of both types of instruments supply the needed balance.

#### SUMMARY

Many college teachers are faced with the task of performance testing. Evaluation of student performance in the clinical setting is a difficult task. Three steps will help organize the faculty effort. First, the purposes of performance testing must be defined. Faculty should keep both student improvement as well as student rating in mind. Evaluation for the purpose of improving student performance depends upon timely, descriptive, specific feedback. Evaluation for the purpose of rating students requires reliable and valid measures of student performance. Second, the desired characteristics of student performance need to be defined. Goals and objectives are essential. Goals are statements of ideals, constituting the purpose of the entire educational endeavor.

#### REFERENCES

- Guba, E. G., & Lincoln, Y. S. Effective evaluation. San Francisco: Jossey-Bass, 1981.
- House, E. H. Evaluating with validity. Beverly Hills: Sage Publications, 1980.
- Ingalsbe, N. & Spears, M. C. Development of an instrument to evaluate critical incident performance. Journal of American Dietetic Association, 1979, 74, 134-140.
- Irby, D. M., & Dohner, C. W. Student clinical performance. In C. W. Ford & K. Morgan (Eds.), Teaching in the health professions. St. Louis: C. V. Mosby, 1976.
- Klein, J. Goals and objectives for adolescent medicine. Unpublished manuscript. Michael Reese Hospital and Medical Center, 1980.
- Knopke, H., & Diekelman, N. L. Approaches to teaching in the health professions. Menlo Park: Addison-Wesley, 1978.
- Marienfeld, R.D., & Reid, J. C. Subjective versus objective evaluation of clinical clerks. New England Journal of Medicine, 1980, 302, 1036-1037.
- McBeath, R. J., & Lassen, H. Performance testing, San Jose State University, 1978.
- Patton, M. W. Qualitative evaluation methods. Beverly Hills: Sage Publications, 1980.
- Scott, H. M., & Sniderman, A. Evaluation of clinical competence through a study of patient records. Journal of Medical Education, 1973, 48, 832-836.
- Scriven, M. The methodology of evaluation. AERA Monograph Series in Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.
- Stritten, F., Hain, J. D., & Grimes, D. A. Clinical teaching reexamined. Journal of Medical Education, 1975, 50, 876-882.

IDEA Papers may be ordered from the Center. Copies are 20¢ each, and the minimum order is 25 copies. Orders of fewer than 125 copies must be prepaid. Shipping and handling costs will be paid by the Center on all prepaid orders.

Center for Faculty Evaluation and Development  
Kansas State University  
1623 Anderson Avenue  
Manhattan, Kansas 66502