

## A Critical Elements Approach to Developing Checklists for a Clinical Performance Examination

Barbara G. Ferrell, Ph.D., The University of Texas Medical Branch

**Abstract:** A two-stage process was used to develop checklists for cases on a clinical performance examination for a clerkship in family medicine. Items generated for each checklist were reviewed by faculty members to determine which of them might be deemed "critical" to the case; those which were so important that less than mastery would result in failure of the exam. For each case, a score was determined based on completion of the items which judges rated as "critical," and weights for each of the items on the checklists were generated. This method yielded a scoring protocol for each case. The protocol is outlined and applied to a hypothetical student. The perceived advantages of the approach are discussed and suggestions made for further work. The protocol is being used to develop similar checklists for additional cases as they are written for the clinical performance examination.

As a part of student evaluation for a newly required four-week clerkship for third year medical students, a clinical performance examination (CPE) using standardized patients was developed. The exam, which uses a long-station focused encounter format, constitutes the final examination for the family medicine clerkship. A description of the CPE and a summary of its properties have been previously reported.<sup>1,2</sup>

To evaluate a student's performance on the CPE, faculty members used a behaviorally anchored form that was similar to the one used to rate a student's daily clinical performance. The rating form had been developed to assess the degree to which students had achieved the objectives of the clerkship.

Early examination of the data from the administration of the CPE revealed that the major source of variation in a student's grade was the examiner to which he/she was assigned.<sup>2</sup> Mean scores for examiners ranged from 84.3 to 93.7.

Clearly there were "hawks" and "doves" among the faculty members and the rating form allowed for a great deal of variability in the way in which student performance was being interpreted and evaluated. The need to develop case-specific checklists for scoring and setting standards in addition to the evaluation of the more generic

competencies being assessed by the rating form was identified.

Most commonly used procedures for standard setting involve the use of "expert" judges who estimate the probability that a "borderline" or "minimally acceptable" student would correctly complete the item on a checklist. A discussion of this concept with department faculty revealed that the notion of borderline was difficult to conceptualize for a clinical performance examination and that faculty felt more comfortable with identifying a set of items which they felt were critical to success on a given case.

The purpose of this paper is to describe a method used to develop case-specific checklists for the CPE which involves the identification of critical elements, checklist items which the faculty feel are so important that to omit them would constitute failure for the case.

The model which was selected to use with the clinical performance examination was one outlined by Berk<sup>3</sup> in which competence would represent 100% mastery of a subset of items, the all-or-nothing state, and a continuously distributed constellation of other items.

Checklist development would involve a review of the cases to determine if they contained tasks that might be deemed "critical elements," those

checklist items which in the judgment of the content experts are so important that less than mastery would result in failure of the exam. The number of critical elements might be few or many, depending on the nature of the case. This approach has been used to develop a clinical equivalency exam for a career-ladder nursing program.<sup>4</sup>

A minimally acceptable agreed upon proportion of other items on the checklist, such as 70%, would also be necessary for satisfactory performance on the case. This would represent the continuum, and would be appropriate when the universe of possible behaviors is large and multiple combinations of performance might be deemed acceptable, such as in a clinical performance assessment.

The model used to develop the checklists for the CPE followed this approach. The method employed reflected Shepard's<sup>5</sup> opinion that "in most knowledge areas there are very few 100% essential items" (p. 176). In assessing a complex behavior such as clinical competence, "ignorance on one point can usually be compensated for by success on other points" (p. 176).

### **Description of the Process**

Each case was written by one or more members of the family medicine faculty and was based on a problem commonly encountered in family medicine. A case script including the history of the present illness, psycho-social information about the patient, past medical history, and information about findings from the review of systems and physical examination was developed. Cases frequently were based on a real patient from the faculty member's practice. For some cases the standardized patient (SP) was the same person upon which the scenario was based. Some SPs had real physical findings; in other cases the SP was trained to elicit the symptoms.

Following a training period for the SP, the case was "tried out." The try-out was not an actual piloting, because faculty members were used instead of students. The purpose of this phase was two-fold: 1. To provide an opportunity to reveal problems with the case and assess its appropriateness for use in the exam, and 2. To generate a list of preliminary items to begin the checklist development process.

During this try-out phase, physician family medicine faculty members interviewed and examined the SP using the same parameters given the third-year medical student; the faculty member was given the mock chart and told that he/she had 30 minutes in which to review the chart, interview and examine the patient. Following the encounter, each faculty member was interviewed. Questions were directed toward level of case difficulty, ability to complete the examination of the patient within the 30 minute time frame, appropriateness of the information in the chart, and inconsistencies in the medical information being given in the case.

An additional set of questions was used to begin the checklist development process. Each faculty member was asked to generate a list of items to be gathered by the third-year student during the encounter. This list included not only history and physical examination data but also psycho-social and health maintenance issues which might be important to the cases. A minimum of two physician faculty participated in the try-out phase for each case and from this a list of case-specific behaviors was developed.

Development of the final checklists took place in two steps. In the first step, a preliminary checklist was developed from the behaviors generated during the case try-out. These checklists and copies of the cases were given to faculty members who were asked to indicate whether each of the behaviors was "critical," "appropriate to the case but not critical," or "not appropriate to the case." Faculty members were also given a chance to add elements which they felt had been omitted from the initial list. Each case and accompanying checklist was reviewed by a minimum of six physician faculty members.

Those behaviors for which there was a 67% agreement among faculty that they were "critical" were used in a second rating process patterned after a modified Angoff method used by Jaeger<sup>6</sup>. Each faculty member was given three cases, randomly presented. Faculty judges in the second round were asked to respond to each item on the checklist by answering "yes" or "no" to two questions: 1. "Upon completion of the family medicine clerkship, should a medical student include this item in his/her focused encounter?" and, 2. "If the medical student does not include this item, should he/she fail that case on the CPE?" The final checklists were comprised of items for

which there was 100% agreement that the item should be included by a third-year medical student performing the focused encounter. Critical elements were defined as those for which at least 80% of the judges had indicated that the medical student should fail the case if he/she did not include it in the focused encounter.

## Results

Each case was evaluated by six faculty members in both phases of checklist development. The protocol developed was applied to five cases initially developed for the CPE. The cases represented five common problems: 1. Abdominal Pain, 2. Chest Pain, 3. Hypertension, 4. Low Back Pain, and 5. Upper Respiratory Infection (URI).

Following the two-stage process, the number of items on the final checklists for the 5 cases ranged from 13 for the URI case to 29 for the Abdominal Pain case (see Table 1). Final checklists contained approximately 1/3 the number of items initially generated for all cases.

Items which all faculty members felt were so essential that failure to include the item should be the basis for failure on the exam were few. However, there were items for all cases but the Hypertension case for which 5 of the 6 faculty members (>80%) agreed. These were operationally defined as "critical elements." The number of critical elements ranged from 0 for the HTN case to 7 for the Chest Pain case. (See Table 2) The more acute the presenting problem appeared to be, the greater the number of critical elements generated. Thus, a chronic problem such as hypertension yielded no items for which omission represented failure, whereas an acute and potentially life threatening one such as chest pain had several.

Checklist items were categorized based on the four objectives for the clerkship; family and bio-psychosocial issues in patient care, diagnosis and management of problems common to family medicine, continuity of care, and health maintenance and patient education. While critical elements for the cases were much more likely to fall under Objective II, Diagnosis, the focus of each case varied with family/bio-psychosocial issues, continuity of care and health maintenance/patient education more important in some cases than others. Faculty members appeared to be less

willing to fail a student based on the omission of family/bio-psychosocial and continuity of care issues than for lack of skill in diagnosis.

The scoring protocol which was developed from this procedure consisted of a small number of critical elements for each case and weights for both critical elements and those items deemed important but not critical. To illustrate the procedure, this protocol has been applied to a hypothetical student for the Hypertension case. (See Figure 1, Sample Checklist.)

A student's completed checklist would first be reviewed for the critical elements identified for that case. These are indicated in bold print on the checklist. Minimal competency for the case constituted completing the "critical elements." For example, there were four critical elements for the HTN case. The hypothetical student completed all four, representing minimal competence. A student who was not observed completing all of these would be given a failing score regardless of the number of additional checklist items completed.

After determining that a student had completed the critical elements, each of the items completed, including the critical elements, was assigned a weight representing the proportion of judges in the standard setting process who indicated that a student should fail if not completing the item. The sum of these weights would constitute the student's score for the case as determined by the checklist. Summing the weights for the hypothetical student's completed items would result in a total of 7.49. The student's score for the checklist would be 7.49/9.32, or 80%.

## Discussion

This paper outlines an attempt to address a problem identified as one of the needed areas of research in using standardized patients to assess a student's clinical performance. A protocol to develop case-specific checklists based on expert judgment was developed. The protocol used a "critical elements" approach to establish the cut score; to identify those tasks which are essential to a specific clinical simulation. As such, it represents a departure from the commonly used checklist development protocols.

The method was conceptually easier for faculty members and made the tedious process of case

review less arduous. Because faculty members have been involved in the process in all of its phases, there is a feeling of ownership of the checklists which have been developed. Both "hawks" and "doves" have a say in the process, and the items contained in the final checklists represent a form of consensus between the two extremes.

The process also made possible the development of cases which could directly assess areas which are not usually contained on checklists in clinical performance assessment, such as integrating family and bio-psychosocial issues into assessment. With the method developed it was possible to weigh parts of the focused encounter and this has led to the development of some cases in which the focus is on diagnosis and others in which the focus is on the other clerkship objectives such as health maintenance or psycho-social issues. It has also enabled us to examine the total set of cases to determine if the emphasis of exam cases fits with faculty expectations. Having a bank of cases with multiple emphases will make the exam more representative of the content of the clerkship and thus increase its content validity.

It is hoped that the checklists can be used along with the more generic rating form to assess the student's performance on each case. Before the checklists can be used to make pass/fail decisions and assign grades to students in the family medicine, however, clerkship issues of validity must be more fully addressed. The "critical elements," minimal competence level and scoring method must be determined to meet the criteria for defensibility outlined by Berk<sup>3</sup> for performance standards for criterion-referenced tests; technical adequacy and practicability.

Steps have been taken to begin this task. The checklists developed for the cases are being validated for technical adequacy by comparing student's scores on the checklists with the more subjective rating forms currently being used to grade students. This process will yield information about the ability of the checklists to classify students into pass/fail categories.

Practicability of the checklists is also being assessed. Checklists are being used by patients, faculty and researchers to determine ease of implementation. Credibility of the checklists for

both new faculty members who may not have been involved in the standard setting process and students must also be assessed.

While additional work is necessary to improve the examination process, the method outlined here represents a beginning of a new method of checklist development for our clinical performance measure.

## References

1. Ferrell B.G. & Thompson B.L. A clinical performance assessment for a third-year clerkship in family medicine. *Fam Med* 1993;25(4):256-7.
2. Ferrell B.G & Thompson B.L. A long-station format clinical examination with standardized patients. *Med Educ* 1993;27:376-81.
3. Berk R.A. A consumer's guide to setting performance standards on criterion-referenced tests. *Rev Educ Res* 1986;56(1):137-72.
4. Ferrell B.G. A Rationale for a Procedure To Evaluate the Competencies of Non-LPN Applicants To the SICCM/ADN Program. Carterville, IL: Southern Illinois Collegiate Common Market, 1979.
5. Shephard L.A. Setting performance standards. In Berk R.A. (Ed.), *A Guide To Criterion-Referenced Test Construction*. Baltimore: Johns Hopkins University Press, 1984.
6. Jaeger R.M. A proposal for setting a standard on the North Carolina high school competency test. Paper presented at the spring meeting of the North Carolina Association for Research in Education, Chapel Hill, 1978.

*Dr. Ferrell is an Associate Professor of Family Medicine and Senior Medical Educator in the Office of Educational Development at The University of Texas Medical Branch. She can be reached via e-mail at bferrell@utmbgalv@mhost.utmb.edu*

Critical Elements Approach  
To Developing Checklists

13

OBJECTIVE	CASE											
	#1 AB. PAIN		#2 CHEST PAIN		#3 BACK PAIN		#4 HTN		#5 URI		TOTAL	
	N	%*	N	%*	N	%*	N	%*	N	%*	N	%*
I. Family/ Biopsychosocial	2	7	1	5	1	5	3	18	0	0	7	7
II. Diagnosis	11	38	11	52	7	37	4	24	2	15	35	35
A. History	7	24	4	19	7	37	5	29	6	46	29	29
B. Physical Exam	2	7	1	5	1	5	2	12	0	0	6	6
C. Laboratory/ Diagnostic	2	7	2	10	2	11	1	6	2	15	9	9
D. Differential Diagnosis	4	14	1	5	0	0	1	6	2	15	8	8
III. Continuity of Care	1	3	1	5	1	5	1	16	1	8	5	5
IV. Health Risk Appraisal/ Health Maintenance/Patient Education	29		21		19		17		13		99	
TOTAL												
*Rounded to Nearest Whole %												

Table 1. Number and Percent of Total Items on Final Checklist  
By Objective and Case.

Critical Elements Approach  
To Developing Checklists  
14

OBJECTIVE	CASE					TOTAL
	#1 AB. PAIN	#2 CHEST PAIN	#3 BACK PAIN	#4 HTN	#5 URI	
I. Family/ Biopsychosocial	0	0	0	0	0	0
II. Diagnosis						
A. History	1	4	0	0	0	5
B. Physical Exam	1	1	0	2	0	4
C. Laboratory/ Diagnostic	0	0	0	0	0	0
D. Differential Diagnosis	1	1	0	1	1	4
III. Continuity of Care	0	0	0	0	0	0
IV. Health Risk Apprai- sal/Health Main- tenance/Patient Education	1	1	0	1	0	3
TOTAL	4	7	0	4	1	11

Table 2. Number of Critical Elements by Case and Objective.

**Critical Elements Approach  
To Developing Checklists**  
15

Student Sam Club ID# 020815  
Case #1 Examiner Dr. Benson

Department of Family Medicine  
The University of Texas Medical Branch  
Clinical Performance Examination  
**CASE STANDARD CHECKLIST**  
Scenario #4  
Hypertension

*For this case the student must include the 4 highlighted items (critical elements) at least 5 of the additional 13 items on the checklist.*

I. During the directed history and physical, the following family/biopsychosocial issues should be addressed:

- A. Out of Work 0.50
- B. Financial Strain 0.67
- C. Single Parent, Three Children 0.50

II. During the directed history, the following information should be elicited:

- A. Interval Symptoms Since Last Visit 0.33
- B. Side Effects of Medications 0.67
- C. Headaches 0.33
- D. Family History of Cardiovascular Disease 0.50

III. During the directed physical, the following should be collected:

- A. Blood Pressure, One Arm 0.83
- B. Heart Exam 0.83
- C. Lung Exam 0.33
- D. Fundoscopic Exam 0.33
- E. Pulse 0.33

IV. Based on the focused history and physical, the student should request the following laboratory/diagnostic data:

- A. Cholesterol 0.17
- B. 6/60 0.67

V. The student should include the following in his/her differential diagnosis:

- Hypertension 0.83

VI. The following should be included in a management plan for this scenario:

- Refill Medications 0.67

VII. The following health maintenance/patient education issues should be addressed in this case:

- Smoking 0.83

STUDENT TOTAL  $\frac{7.49}{9.32} = 80\%$