DOCUMENT RESUME

ED 329 594 TM 016 257

AUTHOR Kromrey, Jeffrey D.; Renfrow, Donata D.
TITLE Using Multiple Choice Examinatio. Items To Measure Teachers' Content-Specific Pedagogical Knowledge.
PUB DATE Feb 91
NOTE 20p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Boston, MA, February 13-16, 1991).
PUB TYPE Guides - Classroom Use - Teaching Guides (For Teacher) (052) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Content Validity; Elementary Secondary Education; Higher Education; *Instructional Effectiveness; *Knowledge Level; Licensing Examinations (Professions); *Multiple Choice Tests; Teacher Certification; Teacher Education; *Teacher Evaluation; Test Validity
IDENTIFIERS *Subject Content Knowledge; *Teacher Competency Testing

ABSTRACT

The use of multiple-choice test items measuring content-specific pedagogical knowledge (C-P) as a viable method of increasing the validity of teacher tests is described. The purposes of the paper are to: (1) present examples of multiple-choice test items used for the assessment of C-P and contrast these items with items used for assessing content knowledge and items used for assessing general pedagogical knowledge; (2) develop a working definition of C-P test items; (3) suggest a preliminary categorization of such items; and (4) describe practical considerations related to the development and use of C-P items in testing programs. Current controversies in teacher assessment are discussed, and a working definition of and categorization system for C-P items are developed. The categories include error diagnosis, communication with the learner, organization of instruction, and learner characteristics. The paper encourages researchers and practitioners to acknowledge that multiple-choice testing has applications which exceed its traditional use. A (14-item list of references and 2 figures are included. (Author/TJH)

Using Multiple Choice Examination Items to Measure Teachers' Content-Specific Pedagogical Knowledge

Jeffrey D. Kromrey
University of South Florida

Donata D. Renfrow
Advanced Technology, Inc.

A Paper Presented at the Annual Meeting of the Eastern Educational Research Association, Boston, February 13-16, 1991

Abstract

The use of multiple-choice test items measuring content-specific pedagogical knowledge as a viable method of increasing the validity of teacher tests is described. The purposes of the paper are (a) to present examples of multiple-choice test items used for the assessment of content-specific pedagogical knowledge, and to contrast these items with items used for assessing content knowledge and items used for assessing general pedagogical knowledge; (b) to develop a working definition of C-P test items; (c) to suggest a preliminary categorization of such items; and (d) to describe practical considerations related to the development and use of C-P items in testing programs. The paper encourages researchers and practitioners to consider the broad possibilities of multiple-choice testing, beyond the previous limits of measuring the lowest level of cognitive ability.

Teacher testing has been a topic of active discussion in education for many years. In their quest for valid, job-related measures of teaching knowledge and skill, researchers have enumerated the limitations of traditional multiple-choice tests and advanced a call for more authentic assessment techniques. While many features of authentic assessment are desirable, the time and resource gaps between legal mandates for testing and the development and validation of instruments suggest interim alternatives may be necessary. The most expeditious alternative is to substantively improve the multiple-choice tests currently being used -- to elevate the assessment beyond the knowledge-level items that mirror what is asked of students. These improved tests seek to measure the higher order thinking skills used by teachers to articulate their knowledge of content and teaching strategies with characteristics of the students being taught. This paper discusses the use of multiple-choice items for the measurement of certain higher order thinking skills applied to teaching, an area of knowledge called content-specific pedagogical knowledge.

The development and implementation of content-specific pedagogical test items provides a vehicle through which the validity and job-relatedness of existing teacher tests may be improved while honoring the constraints of time, budget, and other general realities of state testing programs. Content-specific pedagogical test items represent a reasonable step in the improvement of multiple-choice components of teacher assessment systems.

The purposes of this paper are (a) to present examples of multiple-choice test items used for the assessment of content-specific pedagogical knowledge (C-P items), and to contrast these items with items used for assessing content knowledge (C items) and items used for assessing general pedagogical knowledge (P items); (b) to develop a working definition of C-P test items; (c) to suggest a preliminary categorization of such items; and

(d) to describe practical considerations related to the development and use of C-P items in testing programs. The paper encourages researchers and practitioners to consider the broad possibilities of multiple-choice testing, beyond the previous limits of measuring the lowest level of cognitive ability (i.e., examinees' recall of content knowledge and general principles of pedagogy).

Current Controversies in Teacher Assessment

Most states involved in certification testing assess beginning teachers with performance-based evaluations, multiple-choice tests, or both. Some tests are designed to measure basic academic skills (reading, writing, and arithmetic); others are developed to measure basic pedagogical knowledge; and others purport to measure content area knowledge. Certification testing has come under increased scrutiny as testing programs have received legal and scholarly challenges to increase the validity of teacher assessment (Jaeger & Bush, 1988; Madaus & Pullin, 1987).

While legal challenges rest principally on issues of test development, scholarly challenges focus on both the formats of test items and the content coverage of tests. Typically, basic literacy tests have been perceived as too elementary, the items covering the same level of content knowledge that students are expected to master. Conversely, some items have been faulted for covering esoteric content that evidences no relationship to classroom teaching.

Rudner (1988) reflected this critical view in the assertion that current certification tests are

> based on the logic that people who cannot pass a simple test of minimal, basic knowledge that is often acquired by eighth grade should not be placed in a position where they are responsible for the

> education of children. Such testing is a poor substitute for a valid test that measures the skills and attitudes needed to be a teacher.... (p. 19)

In an effort to broaden the conceptual base of the education profession, Shulman (1987) attempted to define a knowledge base of teaching. The seven areas of knowledge proposed by Shulman include (a) content knowledge, (b) general pedagogical knowledge, (c) curriculum knowledge, (d) pedagogical content knowledge, (e) knowledge of the learner, (f) knowledge of educational contexts, and (g) knowledge of educational goals. Of these, Shulman asserted that pedagogical content knowledge may best delineate the knowledge base of teaching:

> ...the key to distinguishing the knowledge base of teaching lies at the intersection of content and pedagogy, in the capacity of a teacher to transform the content knowledge he or she possesses into forms that are pedagogically powerful and yet adaptive to the variations in ability and background presented by the students. (Shulman, 1987, p. 15)

Several projects are underway to explore more "authentic" approaches to teacher assessment, using videotapes of classroom instruction, essay questions, portfolio evaluation, and simulation exercises (e.g., Leinhardt, 1990; Popham, 1988; Shulman, 1986, 1987, 1988). These new assessment approaches are appealing in their face validity; however, they are significantly more expensive to administer and score, and their psychometric rigor has not been thoroughly appraised. Although Rudner's critique accurately describes many tests that fall short of Shulman's map of the teacher knowledge domain, the press to abandon multiple-choice items is likely to be premature. The development of multiple-choice items that measure more

comprehensive aspects of teaching is largely unexplored in the literature.

Developing a Working Definition of C-P Items

Figure 1 presents a P item that measures examinees' knowledge of the presentation of concepts. This item measures a general awareness of "what to do next" that may apply to any content field. This is contrasted, in the figure, with a C-P item from the field of Specific Learning Disabilities. This item also measures "what to do next," but the application is specifically embedded in teaching mathematics to a learning disabled student. An examinee's ability to answer this item requires knowledge of the mathematics content, blended with general pedagogical knowledge and the specific pedagogical techniques used in teaching the learning disabled.

Similarly, Figure 2 presents two items from a subject-area test in Art. The C item measures knowledge of the use of media to achieve a desired artistic effect. In contrast, the C-P item measures the application of the use of media to a specific instructional setting. Again, the C-P item measures a blend of content knowledge with pedagogical knowledge.

Our working definition of C-P items originated with Shulman's (1986) functional distinction that these items measure the knowledge and skill that distinguish the biology teacher from the biologist. As we examined representative items from several content areas, we noted two exceptions that suggested that Shulman's distinction may be too narrow.

First, a separate discipline of practice, distinct from educative involvement, is not discernible in some content fields. In a field such as music, the distinction between the musician and the music teacher is obvious. However, a noninstructional parallel profession for elementary education or teaching the emotionally handicapped does not exist. In the latter fields, it

is necessary to distinguish between the many content areas that are taught and their corresponding noninstructional disciplines. When an elementary teacher is teaching music, the C-P items relevant to the teaching assessment are those that distinguish the teaching of music to primary-grades learners from the knowledge and skill used by the practicing musician. When the same elementary teacher is teaching arithmetic, the C-P items distinguish the teacher from the mathematician.

Second, many C-P items can be answered by either teachers of the content or practitioners of the field. The mastery of the knowledge to answer an item is not the critical distinction. Rather, the critical issue is the relationship between the knowledge assessed by the item and (a) the performance of the act of teaching the discipline, and (b) the practice of the discipline. C-P items reflect the process of teaching the content, not the noninstructional practice of the discipline.

With these two clarifications of C-P items, the following working definition is proposed:

> The class of C-P items includes those items for which the examinee's determination of the correct response depends upon knowledge of the treatment of content in educational situations.

This definition excludes items that solely address content, without an educational context, and items that address general pedagogical principles in the absence of content-specific interpretations. Additionally, the definition forces attention on test items themselves, rather than on the conceptual domain from which items will be developed. Such a focus maintains a practical orientation and avoids potential distraction into arguments about whether C-P items arise from a separate domain or from the intersection of two or more extant domains (Reynolds, 1990).

A Proposed Categorization of C-P Items

Based upon a review of C-P items developed for a variety of subject-area tests, four major categories of items have been identified.[1] By the nature of its development, this list of categories must be viewed as incomplete. The C-P items reviewed in the process of developing this framework were not written to determine the number of different ways such items can appear, but were written to measure specific skills identified as important for inclusion on teacher subject-area certification examinations. As more testing programs gain experience in developing this type of item, it is anticipated that new categories and new variations on these four categories will appear.

Category 1: Error Diagnosis

One of the most commonly occurring categories of C-P items is error-diagnosis. The stimulus presents an example of student work. For example, 10 measures of a musical score are played and the measures are printed; a student's solution to a series of mathematical problems is presented; or several paragraphs of text with oral reading errors are marked. The examinee is required to respond to the example. Problem analysis can occur in different ways:

- Identify the manifest error (e.g., violins played in C natural instead of in C sharp).

- Identify the student's logical error (task analysis) either by naming the problem or by replicating it; e.g., the

---

[1] Previous efforts have been directed at the development of a framework for teacher knowledge (e.g., Shulman, 1987; Tamir, 1988; Smith & Neale, 1989) but, aside from Carlson's (1989) list of item types, this is probably the first attempt to develop a framework to classify test items designed to measure content-specific pedagogical knowledge.

student did not convert to a common denominator; the student misinterpreted 2/2 time and played half notes as two beats instead of as one beat.

Category 2: Communicating with the Learner

The second major category of C-P items deals with appropriate communications between teachers and learners. These types of items appear in the following major ways:

- Evaluate student homework. For example, which feedback is most appropriate for a six-year-old first grader who wrote a story about "nites in shng armr ftng dragnz?"; how should a teacher provide feedback regarding a student's customer letter responding to a delayed order for a business communication class).

- Simulate a dialogue between teacher and student(s) as the item stimulus, to show student confusion. The response required is a "next step" activity or query that would best lead the student(s) to understanding the problem and resolving the confusion.

Category 3: Organization of Instruction

This category of test items focuses on teacher plans for instruction. For instance, an item stimulus may describe a group of students and an instructional objective. The item response options would be teaching activities, one of which is most appropriate for the group and the objective. Variations on this basic item type are:

- An activity is described that did not result in successful instruction and the examinee provides an alternate activity.

- A failed activity is described, and the respondent provides a plausible reason for the failure.

- An activity is described, some part of which is inappropriate. The respondent identifies how the activity can be corrected or why it is inappropriate.

- A failed activity is described and a successful corrected activity is described. The respondent identifies a reason the correction worked.

- A set of available classroom resources is described, and the respondent provides a plausible activity or a means to compensate for a limit in the resources. For example, the first violinist contracted mononucleosis two days before a concert. How would you compensate for this absence? A chemistry teacher is out of compound X, but has plenty of U, V, and W. Which, if any, of these can be substituted to complete the planned lesson?.

Further variations on this theme include items that describe a group of students and instruct the examinee to:

- Identify the objective, given a class activity, .

- Order a set of activities or skills in the most appropriate manner.

- Translate material to a different level (e.g., the trumpets can't play this part. How should it be simplified for rehearsal?).

Another subcategory of instructional organization questions addresses content-specific methods, materials, and evaluation:

- Content-specific methods and materials. For example, the stimulus presents a musical score and asks about its use with a particular group of students; the stimulus presents a description of a reading activity which the examinee must identify as one particular method of teaching reading.

- Questions on formal and informal evaluations in the content area include selecting an appropriate evaluation procedure, interpreting the results, drawing reasonable conclusions from evaluation results, and predicting appropriate instructional directions and next steps.

Category 4: Learner Characteristics

The final category of C-P items is that which includes items addressing the examinees' knowledge of developmental norms within the content area or the expected sequences of skill development and the progression of competence in the discipline (e.g., a teacher is having trouble teaching addition of fractions to first graders. Why?).

Practical Issues

Carlson (1989) pointed out that C-P items are typically more difficult to write than either C items or P items. The authors' experiences confirm that these items require more planning, writing, and editing than others. Typically, content items are explicit, requiring fewer words. While it is relatively easy to produce these single-fact recall items, it is much more difficult to "freeze frame" a teaching situation. Instead of selecting literal information, writers are asked to call up to conscious awareness all the elements and relationships that impact a teacher's decisions (e.g., Should this choral piece be accepted or rejected? How should I simplify this complex set of rhythms for the clarinets? What rehearsal activities need to be planned

for this band piece?). C-P items require a metacognitive awareness of the teaching process, that is not needed for the composition of items measuring only content knowledge.

Veteran teachers working as item writers often function on an intuitive level, and work so quickly that it is difficult for them to be conscious of the variables that impact their decisions. Part of the frustration for item writer trainers is not knowing how to give the proper guidance. The authors' past two years of experience have helped pool many examples and questioning strategies such as those described earlier. These could now be used to guide item writers through this process with less frustration and more productivity.

Once a viable context is described, and a question is formulated, additional problems arise. While a factual item is likely to have a single, unambiguously correct response, writers can easily select from many reasonably correct options for a C-P item.

From a communication theory standpoint, this apparent ambiguity is easy to explain. Because the C-P item captures an actual "frame" from the classroom "reel," writers must contend with the focus of the examinee's attention, the interpretation of the symbols on which the focus is directed, the relationship of these symbols with the entire teaching context, and the words chosen to present the context (words that reflect hidden -- or not so hidden -- biases).

More writing talent is required to compose C-P items. Creating a scenario to describe the "freeze frame" necessitates the best imagination and recall, noting all details necessary for the examinee to consider. Many early drafts of items contain sketchy information, the writers not realizing how much of their mental image never made it to the paper.

A good C-P item evolves out of many iterations through which the scenario is refined. Qualifiers are added to limit the acceptability of otherwise plausible responses. Because these

items are more complex, field testing provides critical feedback to produce workable revisions of the items. Without the luxury of adequate field testing (a problem especially common in low-incidence content areas), test developers must increase the frequency and intensity of item reviews by the test development committees, editors, and psychometricians.

An essential consideration for test developers is the psychometric performance of C-P items on operational test forms. Initial data suggest generally favorable performance of these items. The items appear to yield reasonable values on indices of item difficulty and point-biserial correlation. However, a systematic examination of the psychometric properties of C-P items is just beginning (Delandshere & Guiton, 1990; Renfrow et al, 1990). Research on item performance in low incidence content fields is likely to be constrained by the small number of examinees.

Teacher educators sometimes react negatively to the concept of C-P items. The most frequent arguments are that no single correct answer is appropriate and that the items negate creative responses. Scholars have voiced a similar concern suggesting that there exists no body of content-specific pedagogy to assess.

Our perspective is that C-P items can meet both challenges. Questions are most often developed with accompanying scenarios that describe a teaching problem. Once the item's scenario and response options are refined, effective foils offer choices with flaws that make them unsuitable given the variables presented in the scenario. Foils must not present equally defensible alternatives that reflect only individual or philosophical preferences.

We suggest that this approach negates the second argument -- there is no need for a body of content-specific pedagogy to be delineated. The items do not seek recall of specific details; they elicit examinees' use of the information they have about students and content to be problem solvers. That is nothing more

than we ask of our teachers.

Conclusions

The assertion of firm conclusions on the success of using multiple-choice items to measure content-specific pedagogical knowledge would be premature at this time. However, tentative conclusions are appropriate for this interim progress report.

The use of C-P items has increased the viability of multiple-choice testing for teacher licensure examinations. The teachers who review the items commend them; the administrators who must stand by the test results support them; even the measurement specialists appear to be nodding their approval.

More research is needed on the statistical properties of C-P items. The introduction of C-P items may raise concerns about the unidimensionality of the tests, concerns that must be addressed through empirical inquiry. Normative information is also needed on the proportion of C-P items on tests. Tests for content areas that emphasize the structuring of content for learners (e.g., elementary education) are likely to have more C-P items than content-area tests in fields such as high-school mathematics, in which teacher mastery of the content itself is a critical concern.

This paper extends Shulman's ideas and operationalizes them by describing how two states have begun developing C-P items. The categorization proposed has been elaborated with sample items, to a greater extent than has been presented previously.

The development of C-P items presents unique demands on test developers. Training item writers is more difficult; more time must be allocated for item writing and review; and several cycles of piloting and revision of items may be required to produce an item of quality. We have found, however, that we can construct multiple-choice items that lead to consensus on the correct

response. More importantly, teachers involved in the projects agree that these items accurately reflect the process of teaching the content.

References

Carlson, R.E. (1989). The intersection of subject matter and pedagogical knowledge in teacher assessment: Practical considerations for item development. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Delandshere, G., & Guiton, G. (1990). Content-specific assessment -- measurement concerns. Paper presented at the meeting of the National Council on Measurement in Education, Boston, MA.

Leinhardt, G. (1990). Capturing craft knowledge in teaching. Educational Researcher, 19, 18-25.

Madaus, G.F., & Pullin, D. (1987). Teacher certification tests: Do they really measure what we need to know? Phi Delta Kappan, 31-38.

Millman, J. (1988). Review of test development and score reporting procedures. In R. Jaeger & J.C. Bush (Eds.), An Evaluation of the Georgia Teacher Certification Testing Program.

Popham, W.J. (1988). The field-testing of items: Traditionalism's trap for innovative teacher assessment. Paper presented at the Directors of State Testing Programs, New York.

Renfrow, D.D., Gabrielson, S., Spaner, S., & Cramer, S.E. (1990). Multiple-choice items on a teacher certification test: Can they be used to assess content-specific pedagogy. Paper presented at the meeting of the National Council on Measurement in Education, Boston, MA.

Reynolds, Anne (1990). Content-specific assessment: Domain perspectives. Paper presented at the meeting of the National Council on Measurement in Education, Boston, MA.

Rudner, L. (1988). Teacher testing -- an update. Education Measurement: Issues and Practice, 7, 16-19.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15, 4-14.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 54, 1-22.

Shulman, L.S. (1987). Assessment for teaching: An initiation for the profession. Phi Delta Kappan, September, 38-44.

Smith, D.C., & Neale, D.C. (1989). The construction of subject matter knowledge in primary science teaching. Teaching and Teacher Education, 5, 1-20.

Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. Teaching and Teacher Education, 4, 99-110.

Figure 1

Items to Measure the Presentation of Concepts

P item	C-P item
According to current research, the most effective method for teaching concepts is to provide A. definition, examples, and non-examples. B. verbal drill and practice. C. visual, auditory, and kinesthetic activities. D. work sheets for written practice.	Mrs. Stevens will introduce addition to her first-grade SLD class. The best hierarchy for her to follow is to have the students A. recognize the words addend and sum; understand the "+" sign; compute sums less than ten; understand place value concerning regrouping tens and ones. B. estimate sums; understand the "+" sign; understand place value of ones and tens; compute sums less than ten. C. find missing addends; understand place value of ones and tens; understand the "+" sign; understand place value concerning regroupings of tens and ones. D. recognize the words addend and sum; estimate sums; understand place value of ones and tens; compute sums less than ten.

Figure 2

Measurement of Content Knowledge and Pedagogical Content Knowledge

C item	C-P item
An artist drawing illustrations for a book with a somber mood would most likely use A. pen and ink washes. B. pastels, wet and dry technique. C. thick and thin markers. D. colored pencils and watercolor washes.	To introduce gesture drawing to a class of first-grade students, the best material is A. crayon. B. vine charcoal. C. oil pastels. D. India ink.