DOCUMENT RESUME

ED 360 385   TM 020 389

AUTHOR   Tittle, Carol Kehr
TITLE   Assessment Research in the Context of Practice.
PUB DATE   Aug 92
NOTE   17p.; Paper presented at the Annual Meeting of the
                American Psychological Association (100th,
                Washington, DC, August 14-18, 1992).
PUB TYPE   Reports - Evaluative/Feasibility (142) --
                Speeches/Conference Papers (150)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Educational Assessment; Educational Practices;
                Educational Research; Elementary Secondary Education;
                *Evaluation Research; Mathematics Tests; *Measurement
                Techniques; Research Design; Research Methodology;
                *Research Needs; Surveys; Test Construction; *Testing
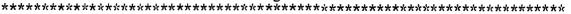   Programs; Test Results
IDENTIFIERS   Mathematics Assessment Questionnaire

ABSTRACT
                Commemorating the work of Anne Cleary, the author
considers the need for research on assessment in the practice
context, provides an example of research in context, and proposes
general areas of development for assessment research in the context
of practice. Research has shown that effects of testing programs on
practice are often not those that were intended. In addition,
assessments are becoming more complex. These factors and the
historical independence of test developers and measurement
practitioners from educational practitioners make research in the
context of educational practice extremely important. The development
of the Mathematics Assessment Questionnaire (MAQ), a survey of
students' thoughts and feelings about learning and doing mathematical
word problems in classroom activity settings, illustrates research in
the context of practice. A framework has been developed to describe
teacher change using the MAQ assessment. Research in the context of
practice will necessarily link more closely to research on teaching
and learning in subject matter areas. To foster this integration,
examples are needed of new organizational arrangements and research
that makes explicit use of theories of teaching and learning. New
procedures to extend and adapt assessment development are needed, as
are criteria for evaluating the meaning and use of assessments in
context. One table and one figure illustrate the discussion. (SLD)

ASSESSMENT RESEARCH IN THE CONTEXT OF PRACTICE

Carol Kehr Tittle
Graduate School and University Center
City University of New York

Assessment research in the context of practice

Carol Kehr Tittle
Graduate School and University Center
City University of New York

Anne Cleary valued research, she valued discovering and understanding troublesome findings about measurement applications, and I am sure she would engage in vigorous dialogue with her colleagues about the topic of this symposium. She would support the importance of the topic and have much to say from her perspective as a long-time contributor to trying to make measurement meaningful. She would be even more concerned about how to make measurement meaningful in view of the changing context for educational measurement. This changed context includes the likelihood that there will be national standards of achievement promulgated for the public education system, and potentially a quasi-national or national testing system.

Historically, the field of applied measurement, including educational measurement, has evolved into a profession that is based on general principles, principles modeled on a scientific paradigm that stresses generalizability across settings and thus functions more or less independently of the particularities of practice. The general principles also assume that measurement has an independence of local settings. However, assessment in the context of practice is highly dependent on the local and particular. My concern today is:

1. to direct attention to the need for research on assessment in its context--that of practice;

2. to provide an example of assessment research in the context of practice; and

3. to propose general areas of development for assessment research in the context of practice.

To talk about maximizing the meaningfulness of measurement means understanding and supporting the professional use of assessment and measurement. It means developing a research field focused on the assessment process itself, that is, research on the meaning of the process and outcomes to professionals and those with whom they work. It means assessment research that can contribute to policy based on the understanding of practice of professionals, whether in

schools or other educational settings, in community, family and mental health settings, or in business, industrial, military or other organizational settings. Today I will focus on assessment research in the context of professional practice in education. A priority for such research in education is assessment used by learners and teachers, since the context of teaching and learning is the largest context of practice in education. Other users of assessment are important--parents, other education professionals, policy makers, and so on--but are not considered here.

1. The need for assessment research

Why be concerned about assessment research for practice?

At least three points indicate the need for more systematic research on assessment in the context of practice:

  i. Research shows the effects of testing programs on practice, effects that are often unintended. And, research shows the beliefs about teaching and learning underlying most current testing practice do not match the prevalent views of educational psychologists.

Research on tests and testing systems has been sporadic. Early research was primarily concerned with teacher attitudes toward and knowledge about test scores (e.g., Goslin, 1967; Ward, 1980). With the use of large-scale testing systems for accountability and resource allocation, empirical research has examined the effects of such systems on teacher practice (Ellwein, Glass, & Smith, 1988; Haladyna, Nolen, & Haas, 1991; Salmon-Cox, 1981; Smith, 1991) and the model of teaching and learning underlying the test developer's views of education (Shepard, 1991). (Information on effects of tests used in evaluation research has also been described, cf. Campbell, 1977, on the corruption of test data). Thus, while assessment and measurement is used extensively in the work of educational psychologists, it has not been the focus of continuing research programs. There is not an extensive body of research on the meaning and use of external assessments to students and teachers, for example.

  ii. Assessment and research programs are developing more complex performance assessments. There are efforts to integrate cognitive and measurement theories.

  The changing views of teachers and learners provided by cognitive (and social) constructivist theory has already influenced assessment development (Tittle, 1990). The influence is felt in the development of complex performance assessments and in models of measurement (Mislevy, 1989). For example, assessments may have components and scoring

based on models of the reading process or expert models of historical writing. Since there is typically a time lag for the integration of theory into teacher development and practice, teachers are being provided assessment information based on theoretical models with which many teachers have little familiarity.

An example is provided by the professional and curriculum standards of the National Council of Teachers of Mathematics. These standards have a cognitive constructivist view of teaching and learning mathematics, and were the basis for changes in the California Assessment Program. A case study of the California assessment in mathematics (Peterson, 1990) included classroom observations. One observer (Cohen, 1990), reports that teachers may adopt some practices characteristic of a cognitive constructivist view of mathematics teaching and learning. However, adoption of practices did not mean the adoption of the underlying epistemology required for transforming practice. Thus, there is not a good fit at present between theory, changes in assessment, and the context of practice. This situation is not unexpected, since assessment is being changed with the purpose of bringing change in teaching practice.

iii. There is an historical independence of test development and measurement practitioners from educational practitioners.

Professional practice in test and assessment development has evolved to be functionally independent of practice in educational classrooms. This is not to say that teachers and students are not involved in test and assessment development. They are, but primarily to provide support for already-developed assessment plans or assessment procedures. Teachers provide judgements, sometimes write tasks, and students provide responses. They are not typically an integral part of the assessment development process. The Standards for Educational and Psychological Testing (1985) do not discuss roles for professional practitioners when developing tests and assessments.

Thus, research on unintended effects of testing programs, the changing psychological and educational theory underlying measurement, and the independence of test development from educational practice, all support the need for developing assessment research in the context of practice. Research is needed to develop an understanding of assessments in classroom settings and to support the intended meaning and effects of assessment processes and the resulting information.

2. An example of assessment research in the context of practice.

What is an example of assessment research in the context of teaching and learning?

Consider, for example, the teacher's task in constructing meaning from assessment results, for a class of from 15 to 40 students. For each of the students the teacher may consider: the student's instructional and personal history that may be relevant to her or his performance (responses) and to instructional planning; the classroom context in which instructional decisions are embedded in the learning and organizational environment of the classroom and school; and an array of assessment results (Tittle, 1989).

In our current work we--Deborah Hecht, myself, Ralph Smallberg (Programmer), and graduate students-- have developed the Mathematics Assessment Questionnaire (MAQ) that surveys students' thoughts and feelings about learning and doing mathematical word problems in the context of three classroom activity settings: during class, a teacher-led activity; working with other students in a problem solving group; and doing homework, an independent activity (Hecht & Tittle, 1990). As part of this work we have developed programs so that the survey can be computer-administered and the results compiled on a teacher program disk (Tittle & Hecht, 1992). There have been many interactions (and some collaborations) with teachers in the development process. Among the most interesting are the opportunities the teacher program disk now offers to examine the manner in which teachers explore an unfamiliar type of assessment information that is made available in an unfamiliar mode.

The assessment information is unfamiliar to teachers in the following way: there are 143 statements intended to elicit student beliefs about their self-regulatory thoughts and behaviors in each of the three activity settings (during class, with others, and homework). There are other statements in the areas of confidence, anxiety, interest, value, motivation, and attributions. These are areas complementary to the mathematics topics and skills typically assessed by teachers and by external testing.

Student responses are available to individual mathematics teachers through a microcomputer-based program, also typically an unfamiliar way for them to view assessment information. (See Figure 1) A small set of teachers has volunteered to use the survey with a class and to meet with us individually to talk about their students' responses to the survey.

The procedures we have used with teachers are based on the think or talk-aloud procedures used in some cognitive research (Ericsson & Simon, 1984). In this procedure the teacher is asked to talk aloud as he or she explores the structure of the program, the structure of the assessment survey, the psychological constructs in the survey, and the responses of her or his individual students and class. (There are also help features and instructional strategies in the program that have not yet been systematically explored with teachers.)

We have used the transcriptions of these talk-aloud sessions to consider the meaning that teachers construct from assessment information. In speculating about the framework that might describe how teachers develop or change in their use of such an assessment tool and program, we have drawn on research from the field of evaluation (Hall & Loucks, 1977) and from current research on development and change of mathematics teachers (Schifter & Simon, in press; Franke, Fennema, Carpenter & Ansell, 1992). The evaluation work focused on examining Levels of Use (LoU) of an educational innovation-- in terms of particular strategies. The LoU identified eight categories including Nonuse, Mechanical Use, Routine Use, Refinement, Integration, and Renewal.

The work with mathematics teachers examined not only the teacher's adoption of new classroom techniques but also the change in "the epistemological perspectives that informed teachers' instructional decision making," that is, how instruction on a particular topic was thought about, planned for, and implemented (Schifter & Simon, in press). The concern with epistemology as evident in the context of particular instructional decisions is due to the explicitly constructivist orientation of the mathematics teacher training program. The decision to add to the LoU ratings (developed for evaluation research) was due to the lack of sensitivity of the LoU ratings of classroom behaviors to the developmental changes in teacher beliefs about learners and teaching that were reflected in qualitative differences in very specific aspects of teaching, e.g., selection of problem types.

The categories Schifter & Simon developed for the Assessment of Constructivism in Mathematics Instruction (ACMI) included: (0) does not hold a constructivist epistemology (CE); (I & II omitted); (III) holds a CE but has difficulty implementing instruction in response to it; (IVA) holds a CE and is comfortable with instruction; focuses on teaching behaviors; (IVB) focuses on student learning from a CE. The addition of a dimension to examine teacher understanding of a particular set of beliefs or model of instruction is relevant to our speculation about teacher change and development in relation to an unfamiliar

set of assessment constructs and an unfamiliar reporting mode.

We have developed a framework to describe teacher change using the MAQ assessment (Hecht & Tittle, 1992). This framework (Table 1) proposes four levels of understanding, levels of the teacher's constructing meaning from the assessment information, using the teacher computer-based program (TMAQ):

Table 1. <u>Teacher Levels of Understanding of Student Responses</u>

1. Acquiring conceptual structures and procedural skills

   i. acquiring facility with the computer program and the general structure of the assessment information

   ii. understanding the psychological constructs that comprise the assessment information (definitions)

2. Contextualizing student responses in the psychological domain

   i. accessing other, relevant information about the student

   ii. interpreting the student or class response conditional upon this other information

3. Using the contextualized information to select or develop specific instructional strategies

4. Internalizing and transforming the assessment information about the psychological domain into other instructional settings and practices.

So far we have been able to identify examples of the first three levels in the five teacher transcripts we have examined. It is also clear that we could identify other meanings that teachers construct from the assessment report, such as knowledge about the self as teacher (Tittle, 1991).

While the discussion so far is particular to one assessment survey, it provides an example of research that can support assessment for practice. The example also suggests that this area of research will necessarily link more closely to research on teaching and learning in subject matter areas.

3. General areas for development for assessment research in the context of practice--the practice of teaching and learning

What is needed for meaningful assessment research in the context of practice?

To maximize the meaningfulness of measurement requires moving assessment research into closer links with teaching and learning research. Four general areas need development to foster this integration.

1. Examples of new organizational arrangements that foster the integration of assessment research and practice. Educational psychologists have actively been concerned with more meaningful, complex theories of teaching and learning, and with more complex models of assessment. Organizational arrangements are needed to support the engineering of these ideas into assessment research conducted in educational settings. The term engineering is used deliberately, to indicate adaptation of ideas or theories in the setting of practice. Research at CRESST provides one example of such an orgnization, as perhaps does the QUASAR project at LRDC.

Engineering requires new approaches to applied research. As Bevan (1991) has argued in a science context, organizational arrangements, networks, or other structures need to establish what are, in essence, communities. The Institute for Research on Teaching, established in the 1970s at Michigan State University by Shulman, and Shulman's research project on teacher performance assessment come to mind as collaborative examples. For Bevan, what he calls the three Cs are paramount-- communication, collaboration, and colleagueship. In educational psychology, as an applied field, this translates into educational psychologists working with teachers and subject matter specialists, building communication, collaboration and colleagueship. Because of the nature of our goals and interests, these arrangements must involve close work with teachers and learners in schools or other educational settings. These arrangements will take time and effort to establish, for the three Cs to occur.

2. Examples of research that make explicit use of theories (models) of teaching and learning to guide studies on the meaning and use of assessment information. These examples should cover a wide range. Some examples will be efforts such as those designed to characterize the teacher's levels of understanding of assessments and student responses. Other examples will study students' understandings of assessment tasks and students'

9

explanations of their own results. Can we characterize student levels of understanding and use them in research on assessment tasks? Still other examples of research will be descriptions and analyses of the assessment tasks, the procedures used for rating/scoring, the reporting of results, and other components of reporting systems, for any and all levels of aggregation. These analyses would examine the extent to which aspects of the assessment system -- tasks, administration procedures, scoring, reporting, aggregating, are congruent with an intended model of teaching and learning.

3. Development of a set of procedures to extend and adapt the existing set of assessment development procedures.

The discussion here is based on work with Garlie Forehand (Tittle & Forehand, 1992). We have described a constructivist perspective on assessment, starting from a particular point of view. Think of three overlapping circles, a Venn diagram. For assessment research for teachers and learners there are three sets of representations from which meaning is constructed: the learner, the teacher, and the assessment developer. The recognition of three sets of representations leads to suggestions for assessment development. For example, assessment development procedures might well include examining the meaning teachers construct from an assessment task, to answer questions such as:

"What existing representations do teachers have about a learner's performance in problem solving? How are representations modified by particular assessment information? How do teachers explain a learner's assessment performance? What instructional actions do teachers suggest based on assessment?..."(Tittle & Forehand, 1992, p. 9).

Assessment development procedures might examine questions about students such as:

"What meaning do students assign specific messages about their performance? What features influence external versus internal attribution? What is the relation between the instructional function of feedback and the effect on self concept? How do students construct meaning from holistic analyses of complex performance? How much do students understand about the process of judging performance?..."(Tittle & Forehand, 1992, p. 11).

Assessment development procedures might examine questions about assessment developers such as: What is the underlying structure of knowledge assumed by an assessment developer for particular tasks? What do the task and other aspects of an assessment program assume about the

characteristics of learners and teachers? --about classroom instructional activities?

A framework that considers all three sets of representations--students, teachers, and developers-- can help to define questions that are relevant to new assessment development procedures. Current procedures examine student responses to develop indices of task characteristics and charactize a learner's development. New procedures are needed to describe types of task meaning and explanations of performance that are reported by representative groups of teachers and students, as well as those intended by assessment developers.

4. Establishment of a set of criteria for evaluating the meaning and use of assessment in the context of practice.

Validation criteria for complex, performance-based assessment have been proposed by Linn, Baker, and Dunbar (1991). Specific criteria for the rating schemes used to evaluate a performance task are described by Quellmalz (1991), and Valencia and Calfee (1991) have discussed concerns that need specification for portfolios used in assessment.

The criteria proposed by Linn and his colleagues, and others, will extend existing practice in validation research. They are concerned with consequences of assessment (corruption of the meaning of the assessments due to coaching and other unintended consequences), fairness, transfer and generalizability, among other criteria. The criteria arise from experiences with existing multiple choice tests and preliminary experiences with more complex performance tasks, and are extrapolations to the use of these performance tasks in large scale assessment programs.

As these validation criteria are developed, they can provide for evidence of the interpretive meaning or narratives that teachers and students construct from and the uses they make of assessment--and be integrated formally into assessment validation procedures.

Summary

We can safely predict that assessment results will be more complex as cognitive constructivist views and assessment become better integrated. Thus, we need to be more successful in maximizing or fostering intended meanings as we use existing tasks and as we change to performance tasks that have assessment components and scoring based on models of cognitive processes, expert performance, and desired educational outcomes. To the already existing

complexity of the classroom setting, we are adding
information based on theoretical models with which many
teachers (and students) have little familiarity. Yet,
understanding these theories or epistemologies is critical
to supporting the intended effects of assessments for
teachers and students.

I have drawn my examples of research primarily from
educational classroom settings, viewing teachers as
professionals. Similar examples could be developed for
counselors, school psychologists, and subject matter
supervisors (e.g., mathematics coordinators and
supervisors). The discussion and many of the general
principles I suggest here for assessment research in the
context of classroom practice will generalize to other
professionals using psychological assessments, to clinical
psychologists and industrial organizational psychologists,
as well, to foster understandings that will maximize the
meaningfulness of measurement.

References

Bevan, W. (1991). Contemporary psychology: A tour inside the onion. American Psychologist, 46, 475-483.

Campbell, D.T. (1977). Keeping the data honest in the experimenting society. In H.W.Melton & D.J.H. Watson (Eds.). Interdisciplinary dimensions of accounting for social goals and social organizations. Columbus, Ohio: Grid.

Cohen, D.K. (1990) A revolution in one classroom. Educational Evaluation and Policy Analysis, 12, 311-329.

Ellwein, M.C., Glass, G.V., & Smith, M.L. (1988). Standards of competence: Propositions in the nature of testing reforms. Educational Researcher, 17, 4-9.

Ericsson, K.A., & Simon, H.A.(1984). Protocol analysis. Cambridge, MIT Press/Bradford.

Franke,M.L., Fennema, E., Carpenter, T.P., & Ansell, E. (1992, April). The process of teacher change in cognitively guided instruction. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Goslin, D. (1967). Teachers and testing. NY: Russell Sage.

Haladyna, T.M., Nolen,S.B., & Hass, N.S.(1991). Raising standardized achievement test scores and the origins of test score pollution. Educational Researcher, 20(5), 2-7.

Hall, G.E., & Loucks, S.F. (1977). A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 14, 263-276.

Hecht, D.H., & Tittle, C.K..(1990, June). Technical Report, Mathematics Assessment Questionnaire, Research Edition. New York, NU: Center for Advanced Study in Education, Graduate School, CUNY. (TM 015 811).

Hecht, D., & Tittle, C.K. (1992, August). Teacher use and understanding of computer-based assessments of student beliefs. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991) Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

Mislevy, R.J. (1989). Foundations of a new test theory (RR-89-52ONR). Princeton, NJ: Educational Testing Service

Peterson, P.L. (1990). The California study of elementary mathematics. Educational Evaluation and Policy Analysis, 12, 241-245.

Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. Applied Measurement in Education, 4, 319-331.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening. Phi Delta Kappan, 62, 631-634.

Schifter, D., & Simon, M.A. (in press). Assessing teachers development of a constructivist view of mathematics learning. (ms.)

Shepard, L. (1991). Psychometricians' beliefs about learning. Educational Researcher, 20(6), 2-16.

Smith, M.L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20 (5), 8-11.

Standards for Educational and Psychological Testing. (1985). Washington, DC: American Psychological Association.

Tittle, C.K. (1989). Validity: Whose construction is it in the teaching and learning context? Educational Measurement: Issues & Practice, 8, 5-13, 24.

Tittle, C.K. (1990). Changing models of student and teacher assessment. Educational Psychologist, 26(2), 157-165.

Tittle, C.K. (1991). Assessment and the construction of meaning for instruction and learning: The meaning and use of assessments by teachers. Paper presented at the annual meeting of the American Eductional Research Asociation, San Francisco, CA, April.

Tittle, C.K., & Forehand, G.F. (1992) Assessment and the construction of meaning for learning and instruction. NY: Graduate School, CUNY, ms.

Tittle, C.K., & Hecht, D. (1992). A technology-based innovation to facilitate teacher access to and use of assessments of JHS students' self-regulatory, affective and motivational beliefs about mathematics. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April.

Valencia, S.W., & Calfee, R. (1991). The development and use
of literacy portfolios for students, classes, and teachers.
Applied Measurement in Education, 4, 333-345.

Ward, J.G.(1980). Teachers and testing: A survey of
knowledge and attitudes toward testing. In L. Rudner (Ed.)
Testing in our schools. Washington, DC.: National Institute
of Education, pp.15-24.

Table 1. <u>Teacher Levels of Understanding of</u>
<u>Student Responses</u>

_____

1. Acquiring conceptual structures and procedural skills

    i. acquiring facility with the computer program
       and the general structure of the assessment
       information

    ii. understanding the psychological constructs that
        comprise the assessment information (definitions)

2. Contextualizing student responses in the psychological
   domain

    i. accessing other, relevant information about the
       student

    ii. interpreting the student or class response
        conditional upon this other information

3. Using the contextualized information to select or
   develop specific instructional strategies

4. Internalizing and transforming the assessment
   information about the psychological domain
   into other instructional settings and practices.

_____

Figure 1

Flow Chart of TMAQ