DOCUMENT RESUME

ED 275 704 TM 860 612

AUTHOR	Tittle, Carol Kehr
TITLE	Next Steps in Competency Testing Programs: Assessment in Context and the Meaningfulness of Results.
PUB DATE	24 Mar 86
NOTE	17p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Content Validity; Elementary School Curriculum; Elementary Secondary Education; *Measurement Objectives; *Minimum Competency Testing; Scores; Secondary School Curriculum; State Programs; Testing Programs; *Test Use; *Test Validity
IDENTIFIERS	*Curricular Validity

ABSTRACT

Statewide minimum competency testing programs have emphasized basic skills in reading, mathematics, and writing. However, continuing concerns are expressed in national reports about the level of achievement, particularly in mathematics and science, and increased testing has been suggested as a means of encouraging curriculum change and evaluating student progress. Increased testing in current competency testing programs is unlikely to meet these broader goals. Two frameworks have implications for evaluating competency testing: (1) the context of assessment--testing of cognitive knowledge, learner characteristics which have an impact on instruction, and curricular validity; and (2) the meaningfulness of test scores and reports to students for self-direction, and to teachers for instructional planning. These frameworks can further increase the links between instruction and the assessments of individual students. A three-page list of references concludes the document. (Author/GDC)

Next steps in competency testing programs: Assessment in context and the meaningfulness of results

Carol Kehr Tittle
Queens College of the
City University of New York

Statewide competency testing programs have emphasized basic skills in reading, mathematics, and writing. However, continuing concerns are expressed in national reports about the level of achievement, particularly in mathematics and science, and increased testing has been suggested as a means of encouraging curriculum change and evaluating student progress. Increased testing in current competency testing programs is unlikely to meet these broader goals. Two frameworks are examined for their implications for evaluating competency testing: 1) the context of assessment (for competency scores), and 2) the meaningfulness of test scores/reports to students for self-direction and to teachers for instructional planning. These frameworks can further increase the links between instruction and the assessments of individual students.

3/24/86

Next steps in competency testing programs: Assessment in context and the meaningfulness of results

Carol Kehr Tittle
Queens College of the
City University of New York[1]

The rapid and continuing growth of large scale testing and assessment programs has historical roots in standardized testing programs. Although many different purposes are stated for competency testing programs, e. g., certification and setting standards for students, accountability and educational policy, and curriculum and educational program evaluation, they are not new purposes. In many aspects current programs provide little change in the types of tests and skills emphasized in past standardized testing programs. Yet there are changes in our views of the learner and the teaching process that will lead to changes in competency testing programs. Part of these changing views are based on a social cognitive perspective, to use Bandura's (1986) terms, on human motivation, thought and action.

The directions of these changes are indicated in a special issue of Educational Measurement (Summer, 1985). In several states, such as Connecticut, with NAEP item administrations, proficiency testing in grade 9, mastery testing of higher order academic skills (grades 4,6, & 8), and the use of fall testing programs, testing has been more concerned with the curricular and instructional impact of the testing program. Testing

[1]Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1986.

is also seen as more than basic skills assessment (and see the extension to other subject areas proposed in Florida).

In the same issue, however, Guthrie and Lissitz point out the importance of distinctions among the use of tests, and Elliott and Hall call attention to "new" purposes of tests and measures. These purposes include: 1)whether the broader assessment programs include measures known (or likely) to be related to student performance; 2) whether measures include those that educators or policy makers can change; and 3) whether the measures are understandable. These are three useful criteria and I will apply them from a different perspective for the state assessment programs that test all students. This perspective is the concern with classroom instruction.

Because competency testing has in many instances demonstrated its effectiveness or leverage on schools, teachers, and students, there is now a need to re-examine the general perspectives or assumptions on which we develop and evaluate statewide assessment programs. There is a need to go beyond defining objectives and setting standards as these programs expand the subject areas and grades tested. For this re-examination we need a framework that permits an understanding of test results on the basis of the context in which they are embedded. We also need a framework that permits an understanding of how meaningful test scores and reports are to students and other decision makers--teachers, parents, curriculum planners, and policy makers. In the next sections two tasks are undertaken:

1. The examination of the context of scores and the implications of this view for assessment programs; and
2. The examination of the meaningfulness of test scores/reports, to students and instructional planners, and the implications of this view for state-wide testing programs.

The focus on context is to enhance instructional planning, building from a research base to identify modifiable variables that relate to the achievement performance of students. The focus on meaningfulness of test scores/reports is to enhance instructional planning and student self-direction, building on studies of student uses of evaluative information and studies of teacher decision processes and uses of test results.

Assessment in Context

Messick (1984) has summarized a broad perspective for the measurement process. The framework implies, "...a strategy of comprehensive assessment in context that focuses on the processes and structures involved in subject matter competence as moderated in performance by personal and environmental influences' (p. 215). The three categories of the framework are: 1. subject matter competence; 2. intraindividual (student) characteristics; and 3. situational contexts. To understand assessment results in context means that

1) the acquisition of knowledge and cognitive skills are seen in developmental terms. One aspect of this developmental perspective is that the constructs emphasized in educational measures may differ according to whether students are at the beginning, intermediate or advanced levels of learning of the subject area. Eventually we may be able to construct items and tasks that relate the sources of student difficulties to the cognitive process and structures that appear at different developmental levels. Attempts to measure problem solving skills appear to incorporate some aspects of this perspective, and draw on research on expert and novice approaches to problem solving.

I am less optimistic than Messick that the use of independent assessments of cognitive abilities (independent of achievement or knowledge-based problems of achievement) will be useful in instructional planning. Such measures seem an extension of the psychological trait approach and instruction remains embedded in subject matter; Burstall (1986) may provide an alternative approach, as do the approaches of Fuerstein and others.

Present research, however, provides support for more immediate use of the other two components of assessment in context. These two arise initially from the need to rule out alternative interpretations of performances, particularly poor performance. More directly, these two components have valuable contributions to make to instructional planning.

2) student characteristics may suggest facilitating or interfering influences on achievement. Particularly of concern here are affective and motivational characteristics that may provide direction to instructional planning. (In Messick's (1984) discussion, drawing on the special education area in which the framework was developed, there is emphasis on screening for biomedical factors and emotional or behavioral deviance that would be destructive of classroom instruction and test performance.)

3) situational contexts also influence test performance. Of particular concern here is whether the student has had an opportunity to learn the knowledges and skills assessed on the test. While other aspects of the student's sociocultural environment may be of interest, such a modifiable variable is directly related to instructional planning.

The implications of the last two components for competency testing programs are supported by studies in the research and evaluation literature. For the situational context, Messick, for example, raises the question whether a student needs remediation, a different mode of instruction, or

better teaching within the current mode. Some evidence related to these instructional planning decisions can be developed by examining the match of curriculum (texts, materials), instruction, and test items for individual students. Studies which have examined aspects of opportunity to learn suggest that as states move to testing other curriculum areas it may be important (at least for samples of classrooms) to have teachers indicate which items in tests students in fact have had an opportunity to learn in texts and in instruction. (See, for example, Porter et al, no date; Cooley & Leinhardt, 1980; Leinhardt & Seewald, 1981; Hanson, McMorris, & Bailey, 1986.)

For the student characteristics component (2), an implication is to include measures of student affect or motivation that appear to be associated with cognitive learning (e.g., Messick, 1979; 1985). As an illustration, consider the current status of research on gender and mathematics. There have been a series of studies over the past 15 years, including those by Fennema and her colleagues (Fennema & Sherman, 1977; Fennema, Wolleat, Pedro & Becker, 1981; Fennema & Peterson, 1985), by Eccles and her colleagues (1983, 1984, 1985), as well as many other researchers (summarized in Chipman, Brush & Wilson, 1985; Dweck, 1986; and Tittle, 1986). This research has yielded findings to date that have implications for the assessment in mathematics of characteristics of both boys and girls. Among these findings are:

- the importance of student perceptions as influences in mathematics achievement and course taking (participation in further and advanced mathematics), particularly as reflected in measures of
  - interests, especially career-related interests, and the perceived need for mathematics and its usefulness

- attitudes toward mathematics, towards oneself as a learner in mathematics and one's expectations for and attributions for success and failure, in perceptions of task difficulties
- positive support from others such as teachers and counselors.

. the general decrease in positive attitudes towards mathematics as students move from the elementary education level to the secondary level.

Student attitudes, interests, expectancies and values are linked with educational achievement, grades and continuing participation in mathematics.

Results of classroom observation studies suggest that teachers typically do not include statements or discussions that are designed to influence these perceptions. (See the studies reported in Wilkinson & Marrett, 1985, and Stodolsky's descriptions of elementary school classrooms, Telling math: Origins of math aversion and anxiety, 1985.) Thus an extension of student measures to the affective area offers opportunities to link such assessments to reports of achievement measures. This may increase the usefulness of test results to students and to teachers in planning instruction. Teachers' statements and activities that might influence student motivation (causal attributions, confidence, perceived usefulness of mathematics) are now identified at a level of specificity that is likely to add to the meaningfulness of reports to teachers and students. Research is continuing to examine constructs such as autonomous learning behaviors (Grieb & Easley, 1984; Fennema & Peterson, 1985) and adaptive motivational patterns (Dweck, 1986).

The implications of this research appear to be translatable into assessment measures that can accompany current competency achievement measures. Further implications are found in the re-examination of how test scores are used and how results are reported to students, teachers and parents.

Meaningfulness of Test Scores and Results

There is a small and growing body of research on teacher use of test results in instructional planning and reports that teachers might prefer for instructional planning purposes (Rudman et al, 1980; Shavelson & Stern, 1981; DeCasper, 1985), as well as surveys of teacher attitudes toward minimum competency testing (described in Tittle, 1982). A related area of research examines the meaning that individuals (such as students) derive from test scores or other evaluative information. The importance of this research rests on the view of the learner, mentioned earlier. The self-regulation of motivation and one's actions operate partly through internal standards and evaluative reactions to one's own performances (Bandura, 1986). In Bandura's view,

> Among the types of thoughts that affect action, none is more central or pervasive than people's judgments of their capabilities to deal effectively with different realities. It is partly on the basis of self percepts of efficacy that they choose what to do, how much effort to invest in activities, how long to persevere in the face of disappointing results, and whether tasks are approached anxiously or self-assuredly..."(p. 21)

Drawing on these research areas, several criteria may be approriate for examining the meaningfulness of test results to students and to teachers. These are identified separately for students and teachers.

For students:

1) The (literal) comprehension of test scores and interpretive reports. What is the accuracy with which terms such as percent, percentile, standard scores, above/below average, are understood by students?

2) The perception of test information

Students interpret events (test results) and organize the information derived from them into beliefs about what leads to what. In Bandura's terms, test scores are environmental cues, affecting the likelihood of particular actions, such as deciding to stay in school, deciding to work harder, through their judged predictive function.

At least two areas are pertinent here. The first area is subject matter learning.

> What does the student understand to be the meaning of the test information for further academic progress? For learning activities to be undertaken? For predicting teacher actions? For predicting parent actions?

The second area is that of self-direction (self-efficacy, self-regulation).

> What does the student take as the meaning of the information for the future? For example, for expectations for continued levels of performance or persistence in the subject matter (values, attitudes, interests, occupational goals)? If the student decides to increase (or decrease) learning activities, what outcomes are these actions likely to produce?

Thus the criteria for the meaningfulness of test scores and reports to students are concerned with the accuracy of communication and with the socially constructed interpretations or meanings that result when students receive test scores. The importance of the criterion of the student's perceptions and use of the information to construct expectations is supported by recent work in motivational theory cited earlier (Bandura, 1986; Dweck, 1986).

Several types of studies are suggested by these criteria. For example, studies may vary the interpretive material in which test scores are embedded for the effect on students with different motivational characteristics and performance levels. Studies may also examine the effect of teacher interpretations designed to affect self efficacy of perceptions of students with different motivational and achievement levels. Some guidance as to the variables to be examined in student perceptions is also given in Bandura (1986, p. 337). As part of the self-monitoring process, test scores can be described as observations that can be examined for accuracy, appropriateness/relevance, and for the standards the individual uses in judging the information (personal standards, referential performances, valuation of activity, and performance attribution--personal or external locus), and self-reaction (positive, negative, tangible, or no self-reaction).

These criteria and perspective encourage the study of test information itself for its meaning to students as self-diagnostic and self-motivating information. Similar criteria would apply to determine the meaningfulness of test scores and results to parents. The focus would be on comprehension and perceptions for predicting their children's academic and motivational actions in the future.

The criteria for meaningfulness need further elaboration for teachers. The criteria of comprehension and perceptions of test information are appropriate for understanding the test results in relation to an individual student. However, there are additional criteria that are related to the instructional process in two areas, instructional planning and teacher self-efficacy.

For teachers:

1) The (literal) comprehension of test scores and interpretive reports.

   What is the accuracy with which terms such as percent, percentile, standard scores, above/below average, are understood by teachers?

2) The perception of test information

   What does the teacher understand to be the meaning of the test information for further academic progress of the student?

   What does the teacher understand as the likely meaning of the information for the student's perception of further academic progress and self-direction (future actions)?

3) The application of test information to instructional planning, its meaningfulness for instruction.

   What does the teacher (curriculum decision maker) plan to do in the instructional process based on the test information?

   Planning will incorporate both areas, student academic (subject matter) learning and self development (motivation and affective) influences on performance.

4) The application of test information for teacher self-efficacy.

   What does a teacher predict will be the outcomes of instructional planning based on tests scores/results?

   What effect does the test report have on a teacher's perceptions of her/his ability to enhance student achievement/motivation?

Test reports are being expanded in content and interpretations are being provided to link test scores to instructional planning. For example, statements of objectives and references to instructional resources are used or are potentially available for reports (DeCasper, 1985).

Because of the important role of the individual's perceptions of information in guiding future actions, it is suggested as another criterion for examining the meaningfulness of test information to teachers.

Studies of teacher criteria can also take a variety of forms, as suggested by the studies for the student criteria. Although not elaborated here, the criteria for evaluting competency assessment programs needs to include school factors that affect the usefulness of results (factors such as the timing and organization for reporting of individual results to teachers, DeCasper, 1985).

Summary

"Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores"(Standards, 1985, p. 9).

In this discussion I have been concerned with aspects of the validity of competency testing programs. I am suggesting an evaluation of the link between testing and instruction, within a view that sees the learner and the teacher as responsible agents in the educational process. I have used two concepts, assessment in context and meaningfulness, that need to be added to our ideas of content in developing validity-related evidence for educational measurements. Assessment in context fits well within the traditional view of construct-related evidence of validity. Thus assessment in context is a useful addition to the validity related eviden for educational measurement.

Validity-related evidence for the meaningfulness and usefulness of specific inferences from test scores or interpretive reports have not been well specified; they are usually noted by their absence, without

criteria to define their presence.

I have suggested that in educational measurement we need validity-related evidence that gives substance to the ideas of appropriateness, meaningfulness and usefulness of inferences based on test scores. In order to do this we must go beyond content-related judgments to strengthen our links with other researchers in education and to draw on new frameworks for integrating teaching, learning, and testing. The resources and importance of the competency testing programs suggest the state education agencies will provide leadership in expanding our frameworks for evaluating assessment programs.

References

Bandura, A. (1986). Social foundations of thought and action. Englewood Cliffs, NJ: Prentice-Hall.

Burstall, C. (1986). Innovative forms of assessment: A United Kingdom perspective. Educational Measurement, 5, 17-22

Chipman, S.F., Brush, L.R., & Wilson, D.M. (1985). Women and mathematics: Balancing the equation. Hillsdale, NJ: Erlbaum.

Cooley, W.W., & Leinhardt, G. The instructional dimensions study. Educational Evaluation and Policy Analysis, 1980, 2, 7-25.

DeCasper, H.S. (1985). A state-wide survey of teacher use of narrative report forms for instructional decision-making. Unpublished doctoral dissertation, University of North Carolina at Greensboro, Greensboro, North Carolina.

Dweck, C.S. (1986). Motivation. Paper prepared for the special issue, Psychology and Education, American Psychologist, in press.

Eccles, J. (1984). Sex differences in mathematics participation. In M.W. Steinkamp & M.L.Maehr (Eds.). Women in science (pp. 93-137). Greenwich, CT: JAI Press.

Eccles, J. et al (1985). Self-perceptions, task perceptions, socializing influences, and the decision to enroll in mathematics. In S.F. Chipman et al (Eds.) Women and mathematics (pp. 95-121). Hillsdale, NJ: Erlbaum.

Eccles (Parsons), J.E., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), Achievement and achievement motives: Psychological and sociological approaches (pp. 78-146). San Francisco, CA: W.H. Freeman.

Fennema, E., & Peterson, P. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L.C.Wilkinson & C.B. Marrett (Eds.), Gender-related differences in classroom interaction (pp. 17-35). New York: Academic Press.

Fennema, E., & Sherman, J.(1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. American Educational Research Journal, 14, 51-71.

Fennema, E., Wolleat, P.L., Pedro, J.D., & Becker, A.D. (1981). Increasing women's participation in mathematics. Journal for Research in Mathematics Education, 12,3-14.

Grieb, A., & Easley, J.(1984). A primary school impediment to mathematical equity: Case studies in rule-dependent socialization. In M.W. Steinkamp & M.L.Maehr (Eds.), Women in science (Vol. 2) (pp. 317-362). Greenwich, CT: JAI Press.

Hanson, R.A., McMorris, R.F., & Bailey, J.D. (1986). Differences in instructional sensitivity between item formats and between achievement test items. Journal of Educational Measurement, 23, 1-12.

Leinhardt, G., & Seewald, A.M. (1981). Overlap: What's tested, what's taught? Journal of Educational Measurement, 18, 85-96.

Messick, S. (1979). Potential uses of noncognitive measures in education. Journal of Educational Psychology, 71, 281-292.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21, 215-237.

Messick, S. (1985) Structural relationships across cognition, personality, and style. In R.E.Snow & M.J.Farr(Eds.) Aptitude, learning, and instruction:Vol 3, Conative and affective process analysis. Hillsdale, NJ. Erlbaum.

Porter, A.C., Kuhs, T.M., Freeman, D.J., Floden, R.E., Knappen, L.B., Schmidt, W.H., & Schwille, J.R. (no date). Is there a core curriculum in elementary school mathematics? A textbook perspective. East Lansing MI: Institute for Research on Teaching, Michigan State University.

Rudman, H., Kelly, J., Wanous, D., Mehrens, W., Clark, C., & Porter, A. (1980). Integrating assessment with instruction: A review (1922-1980). Research Report 75. East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Shavelson, R.J., & Stern, P. (1981) Research on teachers' pedagogical thoughts, judgements, decisions, and behaviors. Review of Educational Research, 51, 455-498.

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

Stodolosky, S.S. (1985). Telling math: Origins of math aversion and anxiety. Educational Psychologist, 20, 125-133.

Tittle, C.K. (1982). Competency testing. In H.E.Mitzel (Ed.) Encyclopedia of Educational Research (5th Ed.) Chicago, IL: Rand McNally.

Tittle, C.K. (1986). Gender research and education. Paper prepared for the special issue, Psychology and Education, American Psychologist, in press.