DOCUMENT RESUME

ED 393 807 SP 036 546

AUTHOR Zhang, Zhicheng; Burry-Stock, Judy
TITLE A Multivariate Analysis of Teachers' Perceived Assessment Competency as a Function of Measurement Training and Years of Teaching.
PUB DATE 8 Nov 95
NOTE 25p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 8-10, 1995).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; *Educational Assessment; Elementary School Teachers; Elementary Secondary Education; Higher Education; School Statistics; Secondary School Teachers; Standardized Tests; *Teacher Competencies; Teacher Competency Testing; Teacher Education; *Teaching Experience; Test Interpretation

ABSTRACT

This study investigated inservice teachers' assessment competency as a function of measurement training and years of teaching. Data were collected from 311 teachers on a 67-item Assessment Practices Inventory (API). Seven composite scores were formed based on the underlying dimensions from a principal factor analysis. A 2x3 MANOVA was conducted to examine the effects of measurement training and years of teaching on teachers' perceived competency in the seven assessment categories as reflected in the composite scores. Multivariate interaction effects between measurement training and years of teaching were significant. Subsequent examination revealed significant multivariate simple effects of measurement training at four or more years of teaching in two factor analyzed assessment categories. Follow up comparisons between means indicated that among teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training on standardized test results interpretation, classroom statistics, and using assessment results in decision making. This group also scored significantly higher on performance assessment and information observation. Appendixes contain tables of data and description of seven standards for teacher competence of educational assessment of students. (Contains 3 tables and 40 references.) (Author/JB)

A Multivariate Analysis of Teachers' Perceived Assessment Competency as a Function of Measurement Training and Years of Teaching

Zhicheng Zhang and Judy Burry-Stock
The University of Alabma

Paper presented at the annual meeting of Mid-South
Educational Research Association
Biloxi, Mississippi

November 8, 1995

Abstract

The study investigated inservice teachers' assessment competency as a function of measurement training and years of teaching. Data were collected from 311 teachers on a 67-items Assessment Practices Inventory (API). The reliability of the API was supported by a high Cronbach alpha of .97. Construct validity of the API was examined using Rasch model procedure. Seven composite scores were formed based on the underlying dimensions from a principal factor analysis. A 2x3 MANOVA was conducted to examine the effects of measurement training and years of teaching on teachers' perceived competency in the seven assessment categories as reflected in the composite scores. Multivariate interaction effects between measurement training and years of teaching were significant ($p < .05$). Subsequent examination revealed significant multivariate simple effects of measurement training at four or more years of teaching in two factor analyzed assessment categories ($p < .01$). Follow up comparisons between means indicated that among teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training on standardized test results interpretation/classroom statistics and using assessment results in decision making ($p < .001$). Among teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training on performance assessment and informal observation ($p < .05$).

Introduction

For decades educational outcomes were measured and controlled through standardized paper pencil tests. The focus of assessment research was on large-scale testing in which measurement specialists played a major role. Paradoxically, large-scale testing constitutes only 1% of all the assessment events; whereas the other 99% of assessments are conducted by classroom teachers (Stiggins, 1991c). The situation was challenged by educational reform as we moved into the 1990s. As part of the educational reform movement cognitive psychologists and educators began to emphasize the importance of problem solving skills and higher order thinking skills which are not measured by simple recall items in paper-pencil tests. In addition, the need to measure students' ability to apply knowledge in real life situations, to cooperate with others, and to do hands-on activities gave rise to a new emphasis on alternative assessment methods. These changes also called for the shift of importance to classroom assessment.

The changes in the focus of assessment in the 1990s support the educational reform movement of teaching for learning. Central to these changes is the growing role of classroom assessment. To keep up with the new changes, classroom assessment techniques ought to receive more emphasis in preservice and inservice measurement training (Stiggins, 1991a, 1991b). There needs to be a better understanding of the nature and characteristics of classroom assessment (Stiggins, 1992a). Teachers' assessment competency should be evaluated in relation to measurement training and teaching experience (Zhang, 1995).

Theoretical Framework of the Study

Research indicates that teachers spend up to 50% of their classroom time in assessment related activities (Stiggins, 1991a). This points to the importance of the need for a high level competency of assessment skills for classroom teachers. A few large-scale studies have been conducted in recent years to investigate teacher competence of educational assessment. These studies have invariably come to the conclusion that teachers' preparation

for testing and measurement is inadequate (Jett & Schafer, 1992; Newman & Stallings, 1982; Marso & Pigge, 1989; Plake, 1993; Plake, Impara, & Fager, 1993; Wise, Lukin, & Roos, 1991). The solutions proposed in these studies included strengthening measurement training for preservice and inservice teachers and requiring an assessment component in teacher certification.

What assessment skills should be included in a testing and measurement course to ensure that preservice and inservice teachers are adequately trained? The answer to this question will delineate the content domain of classroom assessment skills. Drawn from measurement textbooks and published literature, Schafer (1991) specified eight content areas in which teachers need to develop assessment skills:

1. Basic concepts and terminology of assessment. Teachers should be well acquainted with assessment concepts and terms such as objective tests, alternative assessments, formative versus summative assessment, criterion-referenced versus norm-referenced testing and grading, validity, and reliability (Airasian, 1994; Carey, 1994; Shafer, 1991).

2. Uses of assessment. Assessment can be used for planning and evaluating instruction, diagnosing problem areas, monitoring progress, grouping students, assigning grades, developing curriculum, and evaluating school improvement (Burry-Stock, 1995; Stiggins, 1987).

3. Assessment planning and development. To plan assessment, the teacher needs to consider students' ability level, specify assessment targets, and choose appropriate assessment methods (Airasian, 1994; Carey, 1994; Stiggins, 1992b). Writing objectives and designing test specifications are two important devices that can be used at assessment planning stage to ensure validity and reliability of assessment (Burry-Stock, 1995; Zhang & Iran-Nejad, 1993). Three major assessment methods are paper-pencil tests, performance measures, and informal

assessment (Stiggins, 1992b). The first two methods are considered formal and the last one informal (Airasian, 1994).

4. Interpretation of assessments. To interpret properly standardized test results, teachers need to have a good understanding of the meaning and limitations of standard scores such as percentile rank, grade equivalent score, normal curve equivalent, and stanine. They should understand the concept of standard error of measurement and be able to use that concept in explaining a percentile band-of-error. In addition, teachers should know how to judge the adequacy of test norms and use norms to evaluate individuals' performance (Airasian, 1994; Carey, 1994; Shafer, 1991).

5. Description of assessment results. This involves the application of basic statistical analysis to classroom assessment data.

6. Evaluation and improvement of assessments. The teacher should be able to use assessment data to analyze the quality of a test (validity and reliability) and test items (item analysis) (Carey, 1994; Gregory, 1992).

7. Feedback and grading. When assigning grades, teachers need to make three decisions: the grading model to be used (criterion-referenced versus norm-referenced), performance components to be included in grades (achievement-related versus non-achievement-related), and the weight each component should receive (Airasian, 1994). Criterion-referenced grades indicate mastery level of teaching objectives whereas norm-referenced grades suggest an individual's relative standing in a class. To produce valid grades, teachers should only include achievement-related components in the calculation of final grades. Assessment components that reflect effort, ability, attitude, and motivation should not be included in subject matter grades because they are hard to define and measure (Stiggins, Frisbie, & Griswold, 1989). Assessment components that provide more information about students' academic achievement and learning (e.g., final exam) should receive more weights (Airasian, 1994; Carey, 1994). The methods that can be used to

integrate different assessment components to determine the final grades include percentage method, weighted method, and maximum point method (Carey, 1994). When called upon, teachers should explain to students, parents, and other educational personnel about the meaning, implication, and limitation of their grading system used to determine grades.

8. Ethics of assessment. Teachers should guard against overuse and misuse of assessment results. Students' right to fair testing and confidentiality should be protected (Plake, 1993; Schafer, 1991).

These assessment skills summarized the expectations of the assessment community for classroom teachers. In 1990, the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), and National Education Association (NEA) issued the seven Standards for Teacher Competence in Educational Assessment of Students (Appendix A). The seven Standards incorporated the various assessment skills addressed in measurement textbooks and assessment literature. According to the seven Standards, teachers should be skilled in

1. choosing assessment methods,
2. developing assessment methods,
3. administering, scoring, and interpreting assessment results,
4. using assessment results in decision making,
5. grading,
6. communicating assessment results, and
7. ethics in assessment.

The seven standards were intended to guide teacher evaluation and measurement training. The assessment skills addressed in the literature and the seven Standards formed the theoretical foundation for the present study.

Problems in Classroom Assessments

Research suggests that teachers are not well prepared to meet classroom assessment demands (Stiggins, 1991a) due to inadequate training in measurement (Goslin, 1967; Hills, 1991; Jett & Schafer, 1992; O'Sullivan & Chalnick, 1991; Roeder, 1972; Wolmut, 1988). Problems can be found in various aspects of classroom assessment.

Stiggins (1992a) conducted qualitative research with a group of teachers teaching math, science, speaking, and writing in second, fifth, eighth, and eleventh grades. Seventy-eight percent of the teachers reported using performance assessment in their classroom. However, one third of these teachers did not define levels of performance or plan scoring procedures in advance, nor did they inform students of performance evaluation criteria. About 50% of the teachers did not record their scoring during assessment. These practices were not in line with recommended performance assessment methods.

From their survey with 130 preservice teachers and 119 inservice teachers, Vanleirsburg and Johns (1991) concluded that approximately 40% of the teachers studied did not know that standardized tests and teacher-made tests should differ in terms of administration. Twenty percent of the inservice teachers and 26% of the preservice teachers disagreed that it was vital to read directly from the test manual when administering a standardized test. Hall and Kleine (1992) defined nonstandard testing practices as teaching test items, increasing time limits, giving hints, and changing students' answers. Fifty-five percent of the teachers in their study reported engaging in these inappropriate practices. The pressure for teachers to raise standardized test scores through means other than instructional improvement was also widely reported by the teachers in another study by Nolen, Haladyna and Haas (1992). In addition, most teachers had trouble understanding and interpreting standard scores such as percentile rank and grade equivalent score (Hills, 1991). Impara, Divine, Bruce, Liverman, and Gay (1991) reported, based on their experimental study, that even with the help of interpretive information, most teachers were weak in interpreting a

percentile band performance profile. Marso and Pigge (1988) concluded from their investigation that teachers were not well trained to use standardized tests in their classroom.

Problems with grading are just as abundant. Wiggins (1988) pointed out that grading criteria used by teachers were arbitrary and mysteries, often a function of teacher taste rather than a representation of inherent and tangible standards. Based on the results of a case study of 15 high school teachers' grading practices, Stiggins, Frisbie, and Griswold (1989) discovered that 80% of the teachers felt that effort, attitude, and motivation should be considered in grading, 50% of the teachers studied incorporated ability into grades, 50% of the teachers studied did not distinguish between formative and summative data and thus did not use weights to reflect the relative importance of assessment components, they gave little attention to assessment errors, and all teachers used subjective nonachievement factors to make decisions about borderline cases.

The common practice of incorporating effort and attitude into grades was also criticized in other studies (Hills, 1991; Jongsma, 1991). Griswold (1993) studied 326 teachers' grading decisions for two borderline cases. He discovered that the vast majority of the participants made effort-based judgments about grades. Griswold suggested that teachers incorporated social-cognitive learning and attribution theories into their grading belief system and practice when they graded on nonachievement factors such as efforts and attitude. It was not clear, however, how teachers operationally define and assess aptitude and other intrinsic dispositions such as effort and attitude.

In sum, previous assessment research has been conducted along two lines: the standards for teacher assessment competency and problems in classroom assessment. Since classroom assessment is a broad area, most studies concentrated on one aspect of classroom assessment. Few researchers have examined teachers' assessment competency against the seven Standards. The present study differs from most other studies in that it is based on the theoretical framework of assessment literature and the seven Standards by the AFT, NCME,

and NEA (1990). The purpose of the study is to investigate the effects of measurement training and teaching experience on teachers' perceived assessment competencies in the entire arena of classroom assessment. In particular, the researchers are interested in finding out if inservice teachers would respond differently to the API due to difference in measurement training and years of training.

Methodology

Sample

The sample was made up of 311 inservice teachers. Two hundred and sixty inservice teachers came from two local school districts in Alabama: Tuscaloosa County School and Tuscaloosa City School. The numbers of elementary, middle school/junior high, and high schools participating in the study were 6, 4, and 6, respectively. A vocational school was also included in the data collection. This was done to ensure a balanced representation of teachers from different grade levels. Another 51 inservice subjects were students enrolled in the graduate courses of Assessment of Classroom Learning (BER 550) and Research Methods in Education (BER 500) at the University of Alabama, and Elementary Science Teaching Methods (CEE 515) at the University of Alabama-Gasden Center during the spring 1995 semester. The inservice subjects were also primarily white (89%) and female (77.4%). The percentages of elementary, junior high/middle school, and high school teachers were 34, 23, and 30, respectively. The remaining 13% were for comprehensive and other types of schools. Forty percent of these teachers obtained a bachelor's degree, another 56% had a Master's degree. About 28% of the inservice teachers had had one measurement course, 45% of the inservice subjects had taken 2-3 measurement courses. The average number of years of teaching was 10.9.

Data Collection

Data from inservice teachers were collected in March, 1995. The instrument, together with a cover letter and computer scanable answer sheet, was distributed to the

teachers by their school principal at a faculty meeting. Since directions were clearly provided in written form, participants responded to the instrument on their own. Those who voluntarily responded to the instrument returned the completed answer sheets to the school secretary. The return rate was approximately 30%.

Instrumentation

The instrument used in the study was the Assessment Practices Inventory (API) designed by Zhang and Burry-Stock (1994). After three pilot studies and numerous revisions, the current version of the API has 67 items each of which describes a classroom assessment practice. The API was developed according to the seven Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, & NEA, 1990). The statistical analysis of the data from 311 inservice teachers yielded a Cronbach alpha coefficient of .97, indicating the instrument had a high measure of reliability. All item-to-total correlations were above .37 with the highest one being .69. The standard error of measurement was 7.7.

The content validity of the instrument was built into the construction process by developing the items according to the seven Standards for Teacher Competence in Educational Assessment of Students. The construct validity of the API was examined using Rasch model (Rasch, 1960) computer program BIGSTEPS (Linacre & Wright, 1994). The item calibration was estimated for each of the 67 items and then located along a continuum. The distribution of item logits from -.89 to 1.31 (Appendix B) indicated that the items, to a degree, defined the theoretical construct "perceived skill level" of classroom assessment for inservice teachers (Wright & Stone, 1979).

Data Analysis

Principal Factor Analysis. To identify the underlying dimensions of the API, a principal factor analysis was conducted with principal axis method of extraction (squared multiple correlations were used on the diagonals of the correlation matrix) and a varimax

orthogonal rotation. Based on eigenvalues greater than one criterion (Kaiser, 1960), seven factors were retained. The seven factors accounted for 51.91% of the variance. The seven factors and their percent of variance were:

1. Develop and administer paper-pencil tests, choose/revise tests for classroom use (12.24),
2. Interpret standardized test results/classroom statistics, use assessment results in decision making (10.1),
3. Develop and use performance assessment, informal assessment (8.31),
4. Communicate test results (6.84),
5. Non-Achievement based grading (5.91),
6. Ethics in assessment (4.43), and
7. Grading (4.07).

Appendix C shows rank ordered factor loadings of individual items for a seven factor solution. For each factor, only high loadings (greater than .31) selected by the computer program are presented. The final communality estimates for each item, the sum of squared factor loadings for each factor, and the percent of variance explained by each factor are also presented.

2x3 MANOVA. Based on the factor analysis, the 67 items were classified into seven assessment categories each of which was represented by a factor. A composite score was calculated for each person by summing up all the numbers of the scaled responses to the items loading high on that factor. The seven composite scores served as the dependent variables. The two independent variables were measurement training (no training, at least one measurement course) and years of teaching (one or less year of teaching, two-three years of teaching, four or more years of teaching). A 2x3 MANOVA was then conducted on the API data.

Using GLM in SAS, the results of the 2x3 MANOVA were checked in the following way. First, multivariate interaction effects were checked; if multivariate interaction effects were significant, multivariate simple effects were checked; if multivariate simple effects were significant, univariate simple effects were checked (Littell, Freund, & [illegible] ctor, 1991).

Results

Table 1 shows significant overall cell effects (F= 1.61, p=.0146), suggesting there were significant multivariate interaction effects between measurement training and years of teaching. Subsequent examinations of the multivariate simple effects indicated significant multivariate simple effects of training at four or more years of teaching (F=3.3, p=.0022). The multivariate simple effects are presented in Table 2.

Table 1

A 2x3 MANOVA on the API Data: Overall Cell Effects of Measurement Training by Years of Teaching on Assessment Practices in Seven Categories N=311

	Value	F Value	p Value
Wilks' Lambda	.82	1.61	.0146*

* significant at alpha=.05

Table 2

A 2x3 MANOVA on the API Data: Multivariate Simple Effects of Measurement Training at Four or More Years of Teaching on Assessment Practices in Seven Categories N=311

	Value	F Value	p Value
Wilks' Lambda	.92	3.3	.0022**

** significant at alpha=.01

Examinations of the univariate simple effects of measurement training at four or more years of teaching in the seven assessment categories revealed significant univariate simple effects in category two (interpret standardized test results/classroom statistics, use assessment results in decision making) (F=12.74, p=.0004) and category three (develop and use performance assessment, informal assessment) (F=4.26, p=.0399). The results are presented in Table 3.

Table 3

A 2x3 MANOVA of the API Data: Univariate Simple Effects of Measurement Training at Four or More Years of Teaching on Assessment Practices in Seven Categories N=311

Assessment Category	F Value	p Value
1. Develop and administer paper-pencil tests, choose/revise tests for classroom u.	1.52	.2187
2. Interpret standardized test results/classroom statistics, use assessment results in decision making.	12.74	.0004 **
3. Develop and use performance assessment, informal assessment.	4.26	.0399 *
4. Communicate test results.	2.76	.0977
5. Non-Achievement based grading.	.19	.6672
6. Ethics in assessment.	2.58	.1091
7. Grading.	.22	.6391

* significant at alpha=.05 ** significant at alpha=.01

Follow up comparison between means indicated that among teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training in category two (46.93 versus 39.46, F=12.74, p=.0004). Among teachers who had taught four or more years, those with measurement training scored significantly higher than those without measurement training in category three (40.86 versus 37.90, F=4.26, p=.0399).

Discussion and Conclusion

The analysis of the data suggested that among teachers who had taught for four or more years, those who had received measurement training were more skilled in interpreting standardized test results, conducting classroom statistics, and using assessment results in decision making (assessment category II) than those who had not received any measurement training. It was also suggested that among teachers who had taught for four or more years, those who had received measurement training were more skilled in using performance assessment and informal observation (assessment category III) than those who had not received any measurement training.

Zhang's research using Rasch model analysis (1995) indicated that interpreting standardized test results, conducting classroom assessment, using assessment results in decision making were perceived by inservice teachers to be the hardest assessment category. According to Impara at el. (1991), teachers had trouble interpreting a percentile band performance profile even with the help of interpretive information. The present research findings suggested that it was in this assessment category that measurement training and teaching experience worked together most effectively to enhance teachers' assessment competency.

Zhang's study (1995) also suggested that performance assessment and informal observation were considered by inservice teachers to be harder than paper-pencil tests. Teachers in general were less proficient in performance assessment and they often failed to

follow the recommended practices in instrument construction, assessment recording, and assessment criteria communication (Stiggins, 1992a). Yet, it is in this category that teachers with measurement training and teaching experience (four or more years) were more skilled than those without measurement training and with less teaching experience. These two findings testified the value of measurement training and teaching experience.

In conclusion, the research findings of the present study provided evidence concerning the effects of measurement training and teaching experience in two assessment categories that were normally considered to be difficult by inservice teachers. One possible explanation of the interaction effects is that measurement training provides the principles and techniques needed for effective classroom assessment and teaching/testing in the classroom gives the teacher an opportunity to practice and use the book knowledge in a hands-on approach. The length of time suggested here (four or more years of teaching) is consistent with the findings from expert teaching literature that teachers develop their expertise through a series of stages as they move from novice to experts (Berliner, 1987). Future study should focus on exploring the assessment areas in which experienced and novice teachers differ and use that information to guide preservice and inservice measurement training.

Since the self-report inventory was used only with 311 inservice teachers mainly from two local school districts, the present research findings should be interpreted with caution. The replication of the study with a larger sample is desired to confirm the present research findings.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. Washington, DC: National Council on Measurement in Education.

Airasian, P. W. (1994). Classroom assessment. McGraw-Hill, Inc.

Berliner, D. C. (1987). In pursuit of expert pedagogue. Educational Researcher, 15(7), 5-13.

Burry-Stock, J. A. (Ed.). (1995). BER 450 Handbook. Tuscaloosa, AL: The University of Alabama.

Carey, L. M. (1994). Measuring and evaluating school learning. Allyn and Bacon.

Gregory, R. J. (1992). Psychological testing history, principles, and application. Boston, MA: Allyn & Bacon.

Goslin, D. A. (1967). Teachers and testing. New York: Russell Sage Foundation

Griswold, P. A. (1993). Beliefs and inferences about grading elicited from students performance sketches. Educational Assessment, 1(4), 311-328.

Hall, J. L. & Kleine, P. F. (1992). Educators' perceptions of NRT misuse. Educational Measurement: Issues and Practices, 11(2), 18-22.

Hills, J. R. (1991). Apathy concerning grading and testing. Phi Delta Kappa. 72(7), 540-545.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? Educational Measurement Issues and Practice, 10(4), 16-18.

Jett, D. L. & Schafer, W. D. (1992). Classroom teacher move to center stage in the assessment area--ready or not. Paper Presented at American Educational Research Association. ED 346144.

Jongsma, K. S. (1991). Rethinking grading practices. The Reading Teacher, 45 (4), 319-320

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20(1), 141-151.

Linacre, J. M. & Wright, B. D. (1994). A user's guide to BIGSTEPS. MESA Press: Chicago

Littell, R. C., Freund, R. J., & Spector, P. C. (1991). SAS series in statistical applications. SAS Institute Inc. Cary, NC.

Marso, R. N. & Pigge, F. L. (1989). Staff development implications from a state-wide assessment of classroom teachers' testing skills and practices. Paper presented at Mid-Western Educational Research Association. ED 31122309

Newman, D. C. & Stallings, W. M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at NCME. ED 220491.

Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and abuses of achievement test scores. Educational Measurement: Issues and Practice.

O'Sullivan, R. G. & Chalnick, M. K. (1991). Measurement related coursework requirements for teacher certification and recertification. Educational Measurement: Issues and Practice, 10(1), 17-19, 23.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. Mid-Western educational Researcher. 6(1), 21-27.

Plake, B. S., Impara, J. C. & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. Educational Measurement: Issues and Practice. Winter, 10-12.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

Roeder, H. H. (1972). Are today's teachers prepared to use tests? Peabody Journal of Education. 59, 239-240.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. Educational Measurement: Issues and Practice, 10(1), 3-6.

Stiggins, R. J. (1987). Design and development of performance assessments. Educational Measurement: Issues and Practice, 6(3), 33-42.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. Educational Measurement: Issues and Practices, 8(2), 5-14.

Stiggins, R. J. (1991a). Relevant classroom assessment training for teachers. Educational Measurement: Issues and Practice, 10(1), 7-12.

Stiggins, R. J. (1991b). Assessment Literacy. Phi Delta Kappa, 72(7), 534-539.

Stiggins, R. J. (1991c). Facing the challenges of a new era of educational assessment. Applied Measurement in Education, 4(4), 263-272.

Stiggins, R. J. (1992a). In teachers' hands: Investigating the practices of classroom assessment. Albany: State University of New York Press.

Stiggins, R. J. (1992b). High quality classroom assessment: What does it really mean? Educational Measurement: Issues and Practice, 11(2), 35-39.

Vanleirsburg, R. & Jones, J. L. (1991, December). Assessment literacy: Perceptions of preservice and inservice teachers regarding ethical considerations of standardized testing procedures. Literacy Research Report No. 12. ED 341666

Wiggins, G.(1988). Rational numbers: Toward grading and scoring that help rather than harm learning. American Educator, 12, 20-25, 45-48.

Wise, S. L., Lukin, L. E. & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education. 42(1), 37-42.

Wolmut, P. (1988). On the matter of testing misinformation. Paper presented at the SRA, Inc, Invitational conference, Phoenix Az.

Wright, B. D. & Stone, M. H. (1979). Best test design. Chicago: MESA Press

Zhang, Z. & Iran-Nejad, A. (1993, November). A thematic approach to teaching tests and measurement . Paper presented at the annul meeting of Mid-South Educational Research Association. New Orleans, LA.

Zhang, Z. & Burry-Stock, J. A. (1994). Assessment practices inventory. Tuscaloosa, AL: The University of Alabama.

Zhang, Z. (1995). Investigating Teachers' Perceived Assessment Practices and Assessment Competencies on Assessment Practices Inventory (API). Unpublished doctoral dissertation. Tuscaloosa, AL: The University of Alabama.

Appendix A

The Seven Standards for Teacher Competence of Educational Assessment of Students

1 Teachers should be skilled in *choosing* assessment methods appropriate for instructional decisions.

Skills in choosing appropriate, useful, administratively convenient, technically adequate, and fair assessment methods are prerequisite to good use of information to support instructional decisions. Teachers need to be well-acquainted with the kinds of information provided by a broad range of assessment alternatives and their strengths and weaknesses. In particular, they should be familiar with criteria for evaluating and selecting assessment methods in light of instructional plans.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use the concepts of assessment error and validity when developing or selecting their approaches to classroom assessment of students. They will understand how valid assessment data can support instructional activities such as providing appropriate feedback to students, diagnosing group and individual learning needs, planning for individualized educational programs, motivating students, and evaluating instructional procedures. They will understand how invalid information can affect instructional decisions about students. They will also be able to use and evaluate assessment options available to them, considering among other things, the cultural, social, economic, and language backgrounds of students. They will be aware that different assessment approaches can be incompatible with certain instructional goals and may impact quite differently on their teaching.

Teachers will know, for each assessment approach they use, its appropriateness for making decisions about their pupils. Moreover, teachers will know of where to find information about and/or reviews of various assessment methods. Assessment options are diverse and include text- and curriculum-embedded questions and tests, standardized criterion-referenced and norm-referenced tests, oral questioning, spontaneous and structured performance assessments, portfolios, exhibitions, demonstrations, rating scales, writing samples, paper-and-pencil tests, seatwork and homework, peer- and self-assessments, student records, observations, questionnaires, interviews, projects, products, and others' opinions.

2 Teachers should be skilled in *developing* assessment methods appropriate for instructional decisions.

While teachers often use published or other external assessment tools, the bulk of the assessment information they use for decision-making comes from approaches they create and implement. Indeed, the assessment demands of the classroom go well beyond readily available instruments.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will be skilled in planning the collection of information that facilitates the decisions they will make. They will know and follow appropriate principles for developing and using assessment methods in their teaching, avoiding common pitfalls in student assessment. Such techniques may include several of the options listed at the end of the first standard. The teacher will select the techniques which are appropriate to the intent of the teacher's instruction.

Teachers meeting this standard will also be skilled in using student data to analyze the quality of each assessment technique they use. Since most teachers do not have access to assessment specialists, they must be prepared to do these analyses themselves.

3 The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.

It is not enough that teachers are able to select and develop good assessment methods; they must also be able to apply them properly. Teachers should be skilled in administering, scoring, and interpreting results from diverse assessment methods.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be skilled in interpreting informal and formal teacher-produced assessment results, including pupils' performances in class and on homework assignments. Teachers will be able to use guides for scoring essay questions and projects, stencils for scoring response-choice questions, and scales for rating performance assessments. They will be able to use these in ways that produce consistent results.

Teachers will be able to administer standardized achievement tests and be able to interpret the commonly reported scores: percentile ranks, percentile band scores, standard scores, and grade equivalents. They will have a conceptual understanding of the summary indexes commonly reported with assessment results: measures of central tendency, dispersion, relationships, reliability, and errors of measurement.

Teachers will be able to apply these concepts of score and summary indices in ways that enhance their use of the assessments that they develop. They will be able to analyze assessment results to identify pupils' strengths and errors. If they get inconsistent results, they will seek other explanations for the discrepancy or other data to attempt to resolve the uncertainty before arriving at a decision. They will be able to use assessment methods in ways that encourage students' educational development and that do not inappropriately increase students' anxiety levels.

4 Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.

Assessment results are used to make educational decisions at several levels: in the classroom about students, in the community about a school and a school district, and in society, generally, about the purposes and outcomes of the educational enterprise. Teachers play a vital role when participating in decision-making at each of these levels and must be able to use assessment results effectively.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use accumulated assessment information to organize a sound instructional plan for facilitating students' educational development. When using assessment results to plan and/or evaluate instruction and curriculum, teachers will interpret the results correctly and avoid common misinterpretations, such as basing decisions on scores that lack curriculum validity. They will be informed about the results of local, regional, state, and national assessments and about their appropriate use for pupil, classroom, school, district, state, and national educational improvement.

5 Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.

Grading students is an important part of professional practice for teachers. Grading is defined as indicating both a student's level of performance and a teacher's valuing of that performance. The principles for using assessments to obtain valid grades are known and teachers should employ them.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to devise, implement, and explain a procedure for developing grades composed of marks from various assignments, projects, in-class activities, quizzes, tests, and/or other assessments that they may use. Teachers will understand and be able to articulate why the grades they assign are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading procedures in order to improve the validity of the interpretations made from them about students' attainments.

6 Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.

Teachers must routinely report assessment results to students and to parents or guardians. In addition, they are frequently asked to report or to discuss assessment results with other educators and with diverse lay audiences. If the results are not communicated effectively, they may be misused or not used. To communicate effectively with others on matters of student assessment, teachers must be able to use assessment terminology appropriately and must be able to articulate the meaning, limitations, and implications of assessment results. Furthermore, teachers will sometimes be in a position that will require them to defend their own assessment procedures and their interpretations of them. At other times, teachers may need to help the public to interpret assessment results appropriately.

Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will understand and be able to give appropriate explanations of how the interpretation of student assessments must be moderated by the student's socio-economic, cultural, language, and other background factors. Teachers will be able to explain that assessment results do not imply that such background factors limit a student's ultimate educational development. They will be able to communicate to students and to their parents or guardians how they may assess the student's educational progress. Teachers will understand and be able to explain the importance of taking measurement errors into account when using assessments to make decisions about individual students. Teachers will be able to explain the limitations of different informal and formal assessment methods. They will be able to explain printed reports of the results of pupil assessments at the classroom, school district, state, and national levels.

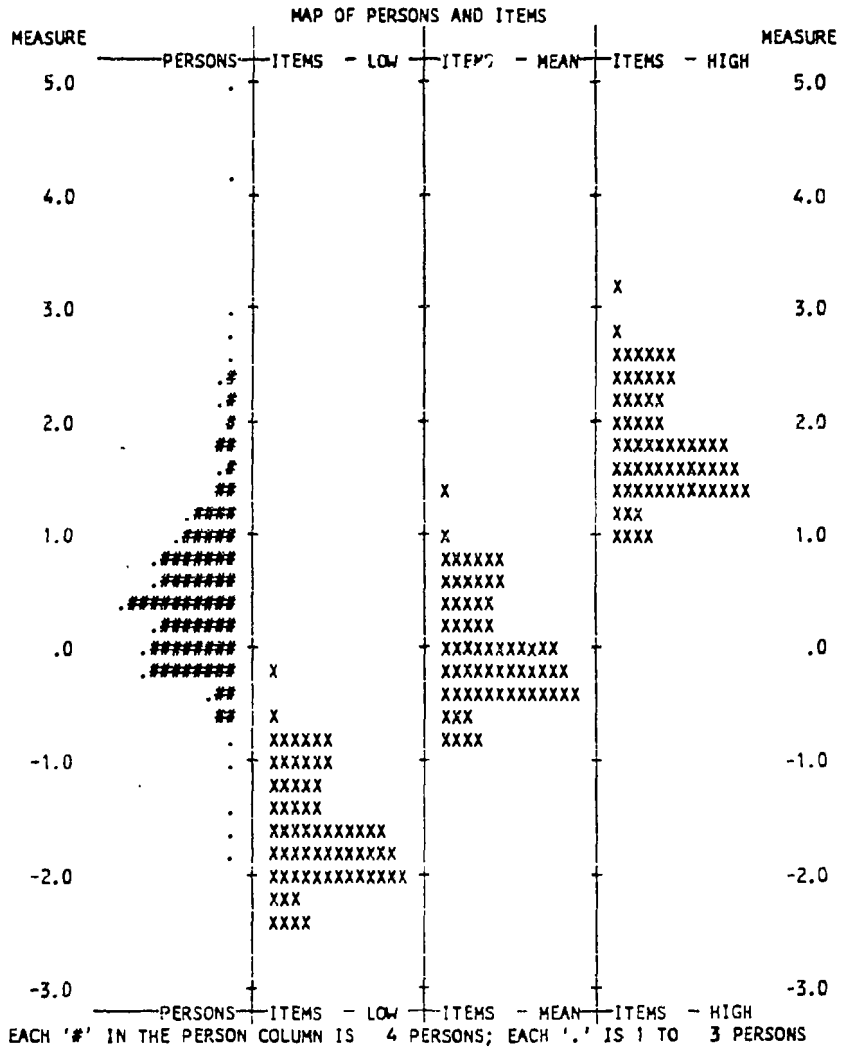
7 Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Fairness, the rights of all concerned, and professional ethical behavior must undergird all student assessment activities, from the initial planning for and gathering of information to the interpretation, use, and communication of the results. Teachers must be well-versed in their own ethical and legal responsibilities in assessment. In addition, they should also attempt to have the inappropriate assessment practices of others discontinued whenever they are encountered. Teachers should also participate with the wider educational community in defining the limits of appropriate professional behavior in assessment.

Teachers who meet this standard will have the conceptual and application skills that follow. They will know those laws and case decisions which affect their classroom, school district, and state assessment practices. Teachers will be aware that various assessment procedures can be misused or overused resulting in harmful consequences such as embarrassing students, violating a student's right to confidentiality, and inappropriately using students' standardized achievement test scores to measure teaching effectiveness.

Appendix B

Rasch Model Analysis of the API Data: Locate Item Calibrations Along a Continuum N=311



Appendix C

Principal Factor Analysis of the API Data: A Seven-Factor Solution With a Varimax Rotation N=311

	Item	Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI	Factor VII	Final Communality Estimates
1	14	.73							.62
2	12	.72							.66
3	15	.72							.59
4	13	.69							.54
5	16	.69							.63
6	4	.65							.56
7	17	.61							.57
8	2	.52							.41
9	5	.52							.38
10	18	.51							.56
11	49	.50							.55
12	32	.50							.49
13	3	.49							.40
14	19	.48							.50
15	10	.43							.55
16	20	.41							.47
17	52	.41							.51
18	11	.41							.43
19	1	.40							.39
20	33		.70						.64
21	34		.69						.56
22	35		.69						.53
23	36		.68						.57
24	38		.65						.58
25	37		.64						.54
26	46		.57						.50
27	39		.54						.60
28	43		.54						.46
29	40		.53						.56
30	9		.48						.40
31	47		.47						.48
32	25		.46						.35
33	8		.46						.31
34	41		.45						.51
35	29			.76					.72
36	28			.74					.70
37	27			.67					.63
38	30			.62					.62
39	26			.56					.60
40	24			.55					.46
41	31			.53					.42

Appendix C Continued

	Item	Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI	Factor VII	Final Communality Estimates
42	7			.48					.46
43	21			.45					.61
44	6			.43					.40
45	22			.43					.48
46	61			.57					.52
47	60				.57				.56
48	62				.52				.52
49	63				.49				.48
50	65				.47				.43
51	42				.44				.53
52	59				.43				.42
53	51				.43				.52
54	64				.40				.36
55	58				.34				.36
56	56					.75			.66
57	54					.73			.62
58	55					.71			.61
59	57					.63			.49
60	53					.63			.58
61	50					.31			.34
62	67						.69		.62
63	66						.66		.58
64	45							.46	.53
65	44							.44	.55
66	23							.39	.47
67	48							.39	.54
Sum of Squared Factor Loadings		8.20	6.77	5.57	4.58	3.96	2.97	2.73	34.78
% Variance		12.24	10.1	8.31	6.84	5.91	4.43	4.07	51.91