DOCUMENT RESUME

ED 291 781 TM 011 066

AUTHOR Marso, Ronald N.; Pigge, Fred L.
TITLE Teacher-Made Tests and Testing: Classroom Resources, Guidelines, and Practices.
PUB DATE Oct 87
NOTE 21p.; Paper presented at the Annual Meeting of the Midwestern Educational Research Association (Chicago, IL, October 15-17, 1987).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Assisted Testing; Curriculum Guides; *Educational Resources; Elementary Secondary Education; Grading; Item Analysis; Multiple Choice Tests; Public Schools; *Public School Teachers; Reprography; *Teacher Made Tests; Testing; Test Items; Test Reliability

ABSTRACT

A total of 326 public school teachers of kindergarten through grade 12 responded to a survey about teacher-made tests, testing practices, and the availability of testing resources and guidelines. Testing practices and preferences were found to vary significantly by grade level, subject area, and length of teaching experience. School testing resources and guidelines varied by school setting and grade level. Most teachers (75 percent) reported high accessibility to student records, availability of curricular guides with objectives, instructor manuals with test items, administration of 50 or more tests annually, and preference for multiple-choice items. Fewer teachers (about 50 percent) reported regular access to reproduction services and the availability of school guidelines for assigning marks. Very few teachers (25 percent or fewer) reported access to computer support services. About 80 percent of the teachers reported rarely or never calculating test means or standard deviations. Over 50 percent of the teachers reported never having estimated test reliability or completing item analyses. Tabulations of teacher responses are appended. (Author/TJH)

Teacher-Made Tests and Testing: Classroom Resources, Guidelines, and Practices

Ronald N. Marso and Fred L. Pigge

College of Education and Allied Professions

Bowling Green State University

Bowling Green, Ohio 43403

A paper presented at the annual meeting of the

Midwestern Educational Research Association

Chicago, Illinois

October 15-17, 1987

Running head: TEACHER-MADE TESTS

Re: Marso & Pigge

Abstract

A total of 326 public school teachers (K-12) responded to a survey about teacher-made tests, testing practices, and the availability of testing resources and guidelines. Testing practices and preferences were found to vary significantly by grade level, subject area, and length of teaching experience; also school testing resources and guidelines varied by school setting and grade level. Most teachers (75%) reported high accessibility to pupil records, availability of curricular guides with objectives, instructor manuals with test items, giving of 50 or more tests annually, and preference for multiple-choice items; fewer teachers (about 50%) reported regular access to reproduction services and the availability of school guidelines for assigning marks; and very few teachers (25% or fewer) reported access to computer support services. About 80% of the teachers reported rarely or never calculating test means or standard deviations, and over one-half reported never estimating test reliability or completing item analyses.

Teacher-Made Tests and Testing: Classroom Resources, Guidelines, and Practices

Even though both teachers and students in the public schools expend substantially more time in the preparation for and the taking of teacher-made as compared to standardized tests, standardized tests have received much more public and research attention (Fleming & Chambers, 1983; Kuhs et al., 1985). Perhaps because of this lack of attention, Dwyer (1982) states the advice given to inservice and preservice teachers about classroom test development and test use represents a consensus of professional judgement rather than knowledge derived from research. Also Gullickson (1986) indicates that there is a discrepancy between teacher-made testing needs as identified by teachers and the content emphasis of university tests and measurement courses. Additionally it would appear that we in higher education have a limited understanding of the nature of assessment practices and resources in the nation's K-12 classrooms (Stiggins & Bridgeford, 1985).

A few survey-type studies have been conducted in recent years in an effort to investigate various aspects of the nature of classroom testing practices. Gullickson (1984) surveyed teacher attitudes toward teacher-made testing relative to characteristics such as teacher and student acceptance of these tests, uses of the test results, constraints on their use and preparation, quality of information derived from, and teacher knowledge of tests and test development. He concluded that teachers are frequent users of and are supportive of frequent testing, but they desire more assistance in better meeting their testing responsibilities as they do recognize limitations of their tests. Stiggins and Bridgeford (1985) surveyed teachers about the types of assessments used in their classrooms, frequencies of the use of various assessment types, and teachers' concerns about classroom assessment. They found that assessment procedures varied somewhat by grade level and subject area. For example, teacher-made tests in contrast to published tests (textbook and standardized tests) were more frequently used in the upper grades as compared to the lower grades. Although only about one percent of their sample of teachers expressed concern about their lack of knowledge about testing, approximately 75% of their sample expressed concerns both about improving their teacher-made objective tests and about the lack of sufficient time to do so. Teachers in this sample at all grade levels and in all subject areas reported more use of and more comfort with nonstructured and structured performance assessments (observations and ratings) than with either the published or teacher-made objective tests.

A few researchers have also investigated the compatibility between university tests and measurements courses and the reported testing needs and practices in the public schools. Stiggins and Bridgeford (1985) noted a considerable difference between topics in commonly used tests and measurements textbooks and the reported assessment practices of teachers. One particularly large discrepancy was noted between the slight attention given to nontest assessment in the textbooks in contrast to the heavy emphasis on this as reported by

classroom teachers. In a similar investigation Gullickson (1986) surveyed both university professors and classroom teachers on what topics should be presented in preservice tests and measurements courses. He found several discrepancies between the two groups including a major difference on nontest assessments with teachers placing much more emphasis on informal observations than did the professors; this type of discrepancy between formal and informal assessments was also reported by Stiggins and Bridgeford (1985) and Salmon-Cox (1981). A second area of major discrepancy noted between the two groups was the greater emphasis placed on statistics by the professors as compared to the teachers. In contrast the two groups appeared to equally value topics such as preparing exams, administering and scoring tests, and general assessment information. In another related study Gullickson and Ellwein (1985) surveyed teachers regarding their use of statistical procedures following the administration of their teacher-made tests. They found that contrary to the heavy emphasis professors place on post-hoc statistical test analyses in tests and measurement courses, very few teachers reported frequent use of these procedures. Only about 1% of the teachers in their sample reported "verifiable" use (only teachers reporting both reliability estimates and calculation of means and standard deviations) of reliability estimates, 9% reported calculating standard deviations, 13% means, and 12% medians while 31% reported calculating item difficulty and 40% reported calculating the range of test scores. On the basis of these findings these researchers concluded that measurement instructors appear to be unsuccessful in convincing teachers of the importance of post-hoc statistical test analyses.

In one of the very few investigations involving the direct assessment of actual teacher-made tests, Fleming and Chambers (1983) reported the analysis of 342 teacher-made tests relative to the cognitive level demands of the 8,800 test items, the presence of item construction and test format errors, and teacher use of various item types. They found that test directions are commonly omitted from tests, test copies are frequently illegible, and that grammatical type writing errors are common. For their sample, short-answer response (including fill-in-the-blank) type items were used most frequently followed by matching items, few true-false items were used, and essay type items were least used. They also found that the preponderance of the 8,800 items measured only at the knowledge level with 94% of the items written by junior high teachers, 69% of the items written by the senior high teacher, and 69% of the items written by the elementary school teachers measuring at this level. In summation, these researchers concluded that their teachers could benefit from inservice training on tests and measurements skills and from an increase in the availability of support services in the preparation of tests.

Purpose

The general purposes of this study were to investigate the a) availability of school resources designed to support teacher-made testing activities and b) various teacher testing practices and preferences. The specific objectives for this investigation were:

1. To determine the extent that specific resources, guidelines, and services (e.g., clerical assistance, grading guidelines, computer test scoring, etc.) exist in the public schools to support teacher-made testing and evaluation activities.

2. To determine the testing and evaluation preferences and practices of classroom teachers relative to: a) use of post-hoc statistical test analyses, b) frequency of tests, c) proportion of test items self-constructed, and d) preference for item types.

3. To determine whether or not testing resources and testing practices vary by a) school setting, b) teaching grade level assignment, c) years of teaching experience, and d) teachers subject-area specializations.

Methods and Procedures

A questionnaire was mailed in the spring of 1986 to a sample of approximately 600 K-12 public school classroom teachers who had graduated from Bowling Green State University during the 1975-1985 period and who were in 1986 employed as full-time teachers in Ohio. The names and addresses of the subjects were obtained from matching the list of full-time teachers certified in Ohio for the 1985-86 school year with those names and social security numbers also appearing on the University's graduation listings during the 1975-1985 time period. A total of 326 (54%) of these teachers returned usable survey forms.

The respondents consisted of 122 elementary, 191 secondary, and 13 specialized area (certified K-12) teachers. One hundred and thirty-four (134) described their employing schools to be rural, 150 suburban, and 42 urban. When asked to report the number of years of full-time teaching experience, 68 of the teachers reported one to three years, 85 reported four to six years, and 173 of the teachers reported having seven or more years of teaching experience. The subject area classifications of the secondary teachers were: 45 business education, 33 science, 41 mathematics, 30 English, 32 social studies, and 10 other areas of specialization. Each of these teachers had completed an undergraduate tests and measurements course taught by one of approximately ten professors providing instruction for the course during the 1975-1985 time period.

The survey instrument for this study consisted of a total of 19 items. Twelve of these items appeared under a section titled "availability of school resources to support testing and evaluation" to which for each item the teachers responded yes, no, or in some subjects or sometimes. Items in this section were comprised of questions about the availability of resources, services, or guidelines such as clerical or computer services, grading guidelines, curriculum guides with objectives, and instructors' manuals containing test questions. The remaining seven items appeared under a second section titled "testing preferences and practices" and was comprised of questions related to the use of various statistical procedures, the types of test items preferred, the sources of teachers' test items,

and how frequently teachers scheduled formal tests. The response format for these items varied from a five-point continuum scored from never (1) to always (5), requests for percentage of use, to a request for the number of formal exams scheduled in a typical school year. In addition to the 19 survey items, each respondent was also requested to provide information about himself or his employing school as follows: a) school setting (rural, suburban, or urban), b) teaching grade level assignment (elementary or senior high school), c) if high school teachers, their subject area specialization, and d) years of teaching experience (1-3, 4-6. and 7 or more years). The 19 items appearing in these two survey sections with a summary of teacher responses are reported on Table 1.

The teacher responses to each of the 12 resource items (section one on the survey form) were analyzed using chi-square in 3 x k contingency tables to determine if the response patterns differed among the various subgroups. Teacher responses of "yes," "no," or "in some subjects or sometimes" always formed the three rows, and the "k" columns for the various contingency tables were formed by the teacher classification variables, namely school setting (k = 3), grade level assignment (k = 2), subject area specialization (k = 5), and years of teaching experience (k = 3).

The teacher responses to the testing practices and preferences items (section two on the survey form) were analyzed using one-way ANOVA procedures on the "scores" produced by each of the seven items. More specifically the dependent variables for these analyses were the teacher responses on each of the five-point response scales (first five items), the reported number of "major" tests given in a typical school year (the sixth item), and the relative percentage of each identified test item type making up the teachers' cumulative yearly efforts at constructing tests (the last item in this survey section). The classification (independent) variables in these ANOVA analyses, as used in the earlier described chi-square analyses, consisted of teachers' school setting, grade level assignment, subject area specialization, and years of teaching experience.

Results

Resources, guidelines, and services

As data in Table 1 reveal, a high percentage of these classroom teachers reported convenient access to individual student records (91%), availability of curriculum guides with stated objectives for units of instruction (87%), counselor or other school staff assistance in interpreting the results of standardized tests (72%), and instructor manuals containing test questions (71%). Conversely, and rather surprisingly, only 50% of the respondents reported having typing and duplication assistance in preparing tests although another 19% reported that this test preparation support was available "sometimes or in some subjects."

A high percentage of the classroom teachers reported that grading guidelines were not available to them. More specifically, 88%

reported not having schcol or department guidelines on how many A's, B's, or C's to assign to a typical class at the end of a school term, 80% reported not having school or department guidelines on how many scores or tests should be used in deriving a term final grade, and 49% of the respondents indicated that school or departmental guidelines suggesting the weight of the final term examination in deriving the school term final grade were not available to them.

The responses to the questions on availability of resources suggested that approximately 40% of the employing schools had at least some computer services available to support teacher-made testing. Approximately 22% (plus 8% responded in some subjects or sometimes) of the responding teachers indicated that computer test-scoring services were consistently available to them, 16% (plus 11% sometimes or in some subjects) indicated that computer test item analysis was consistently available to them, 26% (plus 17% sometimes or in some subjects) reported that computer grade book record keeping was consistently available to them, and 22% (plus 20% sometimes or in some subjects) reported that a computer program for generating test items for their classes was consistently available to them.

Testing preferences and practices

Most of the teachers reported infrequent use of statistical procedures following the administration of their teacher-made tests: 80% of the responding teachers indicated that they never or rarely calculate test means and standard deviations (5% responded always or nearly always), 60% indicated that they never or rarely estimated the reliability of their tests (15% responded always or nearly always), and 54% of the responding teachers indicated that they never or rarely completed item analyses of their tests (16% nearly always or always).

The teachers did report frequent scheduling of formal tests (excluding quizzes and spelling tests) in a typical class and in a typical school year. The mean of the numbers of tests reported in a typical school year was 54.1 with 31% of the teachers reporting the administration of 60 or more formal tests and 15% of the teachers reporting the administration of 100 or more formal tests in a typical school year. When asked how frequently they scheduled formal tests in a typical class, 20% reported scheduling one or more formal tests each week, 49% reported one every two weeks, 15% one per month, 7% three or four per semester, and only 6% reported scheduling two or fewer formal tests in a typical class during a school semester.

Over one-half of the teachers reported writing three-fourths or nearly all of the items used on their teacher-made tests. More specifically, approximately 37% of the teachers reported writing almost all of their test questions, 20% about three-fourths of their items, 19% about one-half, 8% about one-fourth, and 14% reported writing very few of the test items used in assessing the progress of their students. For all the test items used during an entire school year, the teachers were asked to estimate the proportion of each item type used; the average of their percentage responses were: 23% problems, 19% multiple-choice, 16% completion, 16% essay, 14%

matching, and 12% true-false. These survey items and teacher responses to them are presented on Table 1, section two.

Variations in Resources and Practices by Teacher and School

The third and final objective of this study was to determine if reported school resources, teacher testing practices, and teacher-testing preferences differed in terms of variations in teachers' school setting, grade level assignments, teaching experience, and subject area specializations. When the teacher responses to the school resources items were examined relative to these various classifications, the school setting classification revealed differences on five school resources for testing; the grade level assignment revealed differences on nine school resources for testing; and the subject area specialization and years of teaching experience each revealed differences on only a single school resource for testing.

More specifically and as noted on Table 2, the school setting classification comparisons indicated that: counselors are less frequently available to teachers in rural as compared to urban and suburban schools (item 3, $X^2 = 10.25$, $p = .04$); grade frequency assignment guidelines are less likely to be available to teachers in urban as compared to rural and suburban schools (item 6, X^2 10.58, $p = .03$); and that computer support services such as scoring tests (item 9, $X^2 = 13.03$, $p = .01$), item analysis (item 10, $X^2 = 12.97$, $p = .01$), and grade book record keeping (item 11, $X^2 = 13.27$, $p = .01$) are more frequently available in suburban school settings than in the other two settings.

The comparisons of teacher responses by grade level assignment revealed that counselor assistance (item 3, $X^2 = 8.08$, $p = .02$), guidelines on the weight of final term exams (item 7, $X^2 = 67.86$, $p = .001$), guidelines on how many scores are to be available in calculating term grades (item 8, $X^2 = 5.77$, $p = .06$), computer test scoring (item 9, $X^2 = 14.36$, $p = .001$), computer grade book record keeping (item 11, $X^2 = 6.45$, $p = .04$), and computer test item generation programs (item 12, $X^2 = 12.03$, $p = .002$) are less frequently available in the elementary as compared to the secondary level schools. Conversely, the elementary teachers more frequently reported the availability of student school records (item 2, $X^2 = 7.88$, $p = .02$), of curriculum guides (item 4, $X^2 = 9.82$, $p = .01$), and instructors' manuals with test items (item 5, $X^2 = 5.67$, $p = .06$) than did the secondary teachers.

With teacher responses classified by teaching area, it was found that science teachers less frequently reported the availability of test typing and duplication services (item 1, $X^2 = 15.47$, $p = .05$) than did the teachers in the other four specializations, but the responses of the teachers within the five subject area specializations (business, science, math, English, and social studies) did not differ on any of the other 11 resources items. And lastly, when the teachers' responses were classified by years of teaching experience, only one of the 12 resource items revealed a significant difference.

Namely, fewer of the less experienced (1-3 years) than the more experienced teachers reported the availability of instructors' manuals with test items (item 5, $X^2 = 10.09$, $p = .04$). The results of this latter set of chi-square analyses are not presented in table form.

When the teacher responses to the seven testing practice or preference items were classified by teacher and school characteristics, it was found that neither the school setting (rural, urban, and suburban) nor the years of teaching experience (1-3, 4-6, and 7 or more years) classifications revealed differences; whereas, the grade level and subject area classifications of the teacher responses each revealed differences on five of the survey items. As shown on Table 3, the secondary teachers as compared to the elementary teachers indicated that they: more frequently calculated means and standard deviations for their tests (item 1.a, elem. $\bar{X} = 1.58$, sec. $\bar{X} = 1.89$, $F = 8.67$, $p = .01$), more frequently completed item analysis procedures (item 1.c, elem. $\bar{X} = 2.20$, sec. $\bar{X} = 2.46$, $F = 3.84$, $p = .05$), wrote proportionately more of their own test items (item 2, elem. $\bar{X} = 2.66$, sec. $\bar{X} = 4.12$, $F = 96.87$, $p = .001$), and gave more frequent tests during a typical course (item 1.d, elem. $\bar{X} = 2.45$, sec. $\bar{X} = 2.14$, $F = 6.51$, $p = .01$). For two testing practices differences between the elementary and secondary teachers were not statistically significant: frequency of calculating reliability after administering teacher-made tests and the number of formal tests given in a typical school year.

Additionally and as presented in Table 4, the secondary teachers as compared to the elementary teachers reported using proportionately more essay items (elem. $\bar{X} = 7.33$, sec. $\bar{X} = 13.31$, $F = 10.06$, $p = .002$) and more problem type items (elem. $\bar{X} = 13.98$, sec. $\bar{X} = 26.37$, $F = 12.59$, $p = .001$), but somewhat fewer completion (elem. $\bar{X} = 18.97$, sec. $\bar{X} = 15.35$, $F = 3.00$, $p = .08$), and fewer multiple-choice item types (elem. $\bar{X} = 24.48$, sec. $\bar{X} = 16.72$, $F = 11.21$, $p = .001$) during a typical academic year. The elementary and the secondary teachers did not differ in their reported use of matching, true/false, and "other" item types.

The subject area classification of teacher responses to the testing practice or preference items revealed (see section two, Table 3) that social studies teachers reported less frequent calculation of test means or standard deviations than did the science teachers, but neither of the means of these two groups of teachers differed from the means of the other three teacher specialization groups (science $\bar{X} = 2.39$, social studies $\bar{X} = 1.46$, English $\bar{X} = 1.73$, math $\bar{X} = 1.95$, business $\bar{X} = 1.86$, $F = 4.18$, $p = .01$) with the post-hoc mean pair comparisons set at the .10 level of significance. Similarly, the social studies teachers as compared to the business teachers reported less frequent use of item analysis techniques (business $\bar{X} = 2.84$, math $\bar{X} = 2.58$, science $\bar{X} = 2.39$, English $\bar{X} = 2.17$, social studies $\bar{X} = 2.09$, $F = 2.99$, $p = .05$), but wrote more of their own test items (social studies $\bar{X} = 4.50$, science $\bar{X} = 4.33$, English $\bar{X} = 4.13$, math $\bar{X} = 4.05$, business $\bar{X} = 3.67$, $F = 3.18$, $p = .02$) than did the business teachers; whereas neither the social studies nor business groups differed significantly from the three other groups of teachers

on these two items. Additionally, the English teachers reported using fewer formal tests during a typical course than did any of the other groups of teachers (English $\bar{X}$ = 2.77, math $\bar{X}$ = 2.10, business $\bar{X}$ = 2.09, science $\bar{X}$ = 2.10, social studies $\bar{X}$ = 1.71, F = 6.58, p = .001); no other pair-wise mean differences were significant ("scores" for this survey item were: 1 = one or more each week through 5 = two or fewer per semester).

When the proportionate use of each item type relative to total number of test items used in preparing tests over a school year was examined relative to the teachers' subject area classification, significant differences among the five specializations were noted on each of the test item types as reported on Table 4. Social studies teachers reported using more completion type items than math teachers with neither of these means being significantly different from the means of the other three groups (social studies $\bar{X}$ = 21.97, science $\bar{X}$ = 16.58, business $\bar{X}$ = 15.44, English $\bar{X}$ = 13.90, math $\bar{X}$ = 7.66, F = 3.54, p = .008). For matching exercises the math teachers reported less use of this item type than all four other groups of teachers (science $\bar{X}$ = 20.33, social studies $\bar{X}$ = 19.54, English $\bar{X}$ = 15.57, business $\bar{X}$ = 14.38, math $\bar{X}$ = 3.41, F = 11.29, p = .001). For true-false items the math teachers reported less use of this item type than did the social studies and business education teachers, the English teachers reported less use of true-false items than did the business education and social studies teachers, and the science teachers reported less use of the true-false items than did the business education teachers (business $\bar{X}$ = 14.69, social studies $\bar{X}$ = 8.52, English $\bar{X}$ = 7.20, math $\bar{X}$ = 3.44, F = 12.50, p = .001). Likewise for multiple-choice type items the math teachers reported less use of these items than any of the other four groups with none of the other group means differing significantly from one another (English $\bar{X}$ = 23.63, science $\bar{X}$ = 26.36, social studies $\bar{X}$ = 18.75, business $\bar{X}$ = 17.00, math $\bar{X}$ = 3.17, F = 13.10, p = .001). The social studies and English teachers did not report a difference for their use of essay type items, but both of these groups used this item type to a greater extent than did each of the other three groups of teachers (English $\bar{X}$ = 29.87, social studies $\bar{X}$ = 21.06, business $\bar{X}$ = 7.44, science $\bar{X}$ = 9.85, math $\bar{X}$ = .32, F = 21.93, p = .001). Last and as expected, the math teachers reported more use of problem type items than did the other four groups of teachers; whereas the business education and the science teachers did not differ in their reported use of problem type items but both reported a greater use of these items than did the English and social studies teachers (math $\bar{X}$ = 78.76, business $\bar{X}$ = 26.47, science $\bar{X}$ = 15.48, social studies $\bar{X}$ = 1.25, English $\bar{X}$ = 1.25, F = 106.55, p = .001).

Summary, Discussion, and Implications

The teacher responses to the survey indicated that individual student records, counselor or other staff assistance in interpreting standardized test results, curriculum guides including unit objectives, and instructor manuals with test questions are available to most teachers; but clerical assistance in the preparation of teacher-made tests appears to be regularly available to only about

one-half of the teachers. Guidelines for assigning grades and for the number of tests or other scores that should be available in calculating term grades are not available to most teachers, but guidelines for the weighting of term final examinations in calculating term grades are available to approximately one-half the teachers. Approximately 40% of the teachers reported the availability of some computer support in meeting their testing responsibilities, but only one in four or fewer teachers reported the availability of specific computer services such as item analysis, test scoring, grade record keeping, or item generation data pools in their schools.

Most teachers (at least 54%) indicated that they never or rarely calculate means or standard deviations, complete item analyses procedures, or estimate the reliability of their teacher-made tests. On the other hand these teachers reported that they frequently prepared and gave many formal teacher-made tests during a typical school year. They reported extensive use of problem, multiple-choice, completion, and matching item types but less use of essay and true-false items. As Gullickson (1984) also reported, most teachers reported scheduling at least one formal test about every two weeks in a typical class. The average number of formal tests scheduled by this sample of classroom teachers in a typical school year was 54.1. Of the total number of items used in a school year, the teachers reported that approximately one of each four items was of the problem type, one in five was either a multiple-choice or completion type item, and only about one in ten items was either a true-false or matching type item.

In accord with the findings of Gullickson and Ellwein (1985) and Gullickson (1986), comparatively very few of these teachers reported regular use of post-hoc statistical procedures (e.g., computing reliability, means, standard deviations, etc.) on the results of their teacher-made tests. Further, and as Gullickson and Ellwein found, teacher responses to the items dealing with statistical procedures appeared to be somewhat inconsistent as many teachers in both studies reported completing estimates of test reliability but calculating means and standard deviations to a much lesser extent; most of us would assume the latter would typically be necessary before performing the former.

It would appear that years of teaching experience factor is not truly related to the availability of school resources for testing activities or to teachers' testing practices and preferences. The only significant finding noted in using this classification in assessing teacher responses to the survey was that the less experienced teachers (1 to 3 years) were less likely to have instructor manuals to assist them in their testing responsibilities. Similarly, testing practices and preferences do not appear to change with additional years of teaching experience or indirectly perhaps that inservice training experiences do not truly affect testing practices or preferences.

The school setting classification (rural, urban, suburban) appeared not to be related to teachers' testing practices and preferences; however, school setting was found to be related to the

availability of five types of testing resources. In this latter regard, rural school teachers reported less frequent assistance from counselors than did the urban and suburban teachers; urban teachers less frequently reported the availability of grade frequency assignment guidelines; and suburban teachers were most likely to report the availability of various computer support services for testing. These differences likely are best explained by differences in human and financial resources available to schools in the various school settings.

The grade level classification (elementary and secondary) was found to be related in several instances to teachers' testing practices and the availability of testing resources. Differences between elementary and secondary teacher responses were noted for nine of the 12 resource items and for five of the seven items devoted to testing practices. The secondary teachers reported greater availability of counselor assistance, grading guidelines, and computer type resources in meeting their testing and evaluation activities than did the elementary teachers; whereas the elementary teachers were more likely to have access to official student records, curriculum guides and instructor manuals. In terms of testing practices (perhaps due to the differences in classroom demands on their time), the secondary teachers appeared to spend more time on their testing as suggested by their reports of more frequent calculation of means and standard deviations, more frequent tests per course, more frequent use of item analysis procedures, and the personal construction of a larger proportion of the items used on their tests. Relative to the use of various item types, the secondary teachers reported relatively more use of essay and problem type items (which are often considered more appropriate for older students) and less frequent use of completion and multiple-choice items than did the elementary level teachers; whereas the reported relative use of matching and true-false items did not differ between the elementary and secondary teachers.

The comparisons among the five subject area classifications of the teachers resulted in response differences on only one resource survey item; this similarity of resources among subject was expected because school testing resources tend not to be allocated by subject area. Teachers' subject area classification did result in the identification of teacher response differences for five of the seven items concerning testing practices. Many of these differences were caused by or associated with the math teachers (perhaps due to the relative uniqueness of the content in this subject area). The math teachers reported more use of problem type test items and less use of all other item types as compared to one or more of the other four groups of teachers, namely, business, science, English and social studies. A second but smaller group of differences was associated with the social studies teachers. These teachers reported less frequent calculation of means and standard deviations and less frequent use of item analysis procedures, but they reported more frequent writing of their own test items and more frequent use of completion and essay item types than did one or more of the other teacher groups. The business education and the science teachers, like the math teachers, reported more frequent use of problem-type items

than did the English and social studies teachers; the English teachers reported the most frequent use of essay-type items; and the business education and social studies teachers reported relatively more frequent use of true-false items than did the other teachers.

In terms of possible implications or recommendations from the data gathered and analyzed from this sample of teachers the following are offered: a) Teachers are not convinced of the value of statistical procedures in improving and using their teacher-made tests to the extent that measurement textbooks and professors emphasize these procedures. b) Teachers and their students expend considerable effort and time in testing. c) It would seem that most school systems need to increase the support available to assist teachers in meeting their testing and related responsibilities. This is particularly true in regard to clerical and computer support services; these would certainly appear to be essential to the improvement of teacher testing in light of the limited amount of time a typical classroom teacher has available to meet these responsibilities. Relative to the possible need for clerical support, Fleming and Chambers' (1983) examination of a sample of teacher-made tests revealed that a large proportion of these tests tended to be illegible and many were handwritten. d) Supervisors and principals should routinely see that beginning teachers acquire instructor manuals as less experienced teachers were found to least likely have available this important test development resource. e) Inservice training should be provided periodically for teachers as it was found that teachers' testing practices and preferences did not change with additional years of teaching experience. f) Supervisors and principals need to be prepared to assist teachers in interpreting the results of standardized tests and to secure various computerized testing support services. This would appear to be particularly necessary for rural and urban schools and in elementary schools where counselors frequently are not available to provide this type of assistance. And g) in light of the frequently reported difficulties and inconsistencies in grading practices, schools need to make a conserted effort to make grading guidelines available to teachers. Very few of this sample of teachers reported having access to such guidelines.

7/2

References

Dwyer, C. A. (1982). Achievement testing. In H. E. Mitzel (Ed.), Encyclopedia of Educational Research (4th ed., Vol. 1, pp. 13-22). New York: The Free Press.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. New Directions for Testing and Measurement, 19, 29-38.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Gullickson, A. R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23, 347-354.

Gullickson, A. R., & Ellwein, M. C. (1985). Post-hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, Spring, 15-18.

Kuhs, T., Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1985). Differences among teachers in their use of curriculum embedded tests. The Elementary School Journal, 86, 141-153.

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

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

7/2

Table 1

Teacher Responses to Availability of Testing Resources and to Testing Practices and Preferences Survey Items

Availability of School Resources to Support Testing and Evaluation

	% Yes	% No	% In some subjects or sometimes
1. Typing and duplication assistance in preparing tests.	50	30	19
2. Convenient access to individual student records, tests, etc.	91	3	7
3. Counselor or other school staff to assist in interpreting class or individual standardized test results.	72	12	15
4. Curriculum guides with stated objectives for units of instruction.	87	4	9
5. Instructor manuals which provide you with questions for tests.	71	9	20
6. School or department guidelines on how many A's, B's, C's etc. to assign to a typical class at the end of the term.	7	88	5
7. School or department guidelines on relative weighting of the final term test or other scores in deriving final term grades.	45	49	6
8. School or department guidelines on how many scores or tests are required in deriving a term final grade.	14	80	6
9. Computer test scoring service for teacher-made tests.	22	71	8
10. Computer analysis of student responses to test questions.	16	72	11
11. Computer grade book record keeping for your classes.	26	57	17
12. Computer programs for generating tests for your classes.	22	57	20

Testing Preferences and Practices

	% Never	% Rarely	% Occasionally	% Nearly Always	% Always
1. How often do you:					
a) Calculate test means and standard deviations?	49	31	14	4	1
b) Estimate reliability for your tests?	35	25	23	12	3
c) Complete item analysis of your tests (check item difficulty, etc.) to determine which questions "worked"?	29	25	28	13	3
d) Give formal (major) tests (exclude spelling or other quizzes) in a typical subject or class?	% One or more Each Week 20	% One Every Two Weeks 49	% One Per Month 15	% 3 or 4 Per Semester 7	% 2 or Fewer Per Semester 6

(Table Continues)

Table 1 (Continued)

2. What proportion of the questions used on your formal tests in a typical school year have you written yourself?

% Very Few	% About 1/4	% About 1/2	% About 3/4	% Almost All
14	8	19	20	37

3. Including all classes or subjects taught, approximately how many formal tests (exclude spelling and other quizzes) do you give during a typical school year? (Hint: Check your grade book.) Number = _____.

$\bar{X}$ = 54.1;

Range of number of tests per year	
10 or more 92%	50 or more 42%
20 or more 75%	60 or more 31%
30 or more 58%	100 or more 15%

4. Of all test questions you use in a typical school year what approximate percentage of the total are of the following types? (Your percents should add to 100.)

Problems	22%	Essay	11%
Multiple Choice	20%	True/False	10%
Completion	17%	Other	4%
Matching	15%		

Table 2

Contingency Tables: Resources and Guidelines, Frequencies and Percentages by School Setting and by Grade Level

School Setting

Resource Item	Teacher Group	Yes f	Yes %	No f	No %	Sometimes f	Sometimes %	X^2	p
#3 Counselor Assistance	Rural	92	(69)	19	(14)	23	(17)	10.25	.04
	Suburban	115	(77)	11	(7)	23	(15)		
	Urban	27	(66)	10	(20)	4	(10)		
#6 Guidelines Frequency A's, B's, & C's	Rural	10	(8)	112	(84)	12	(9)	10.58	.03
	Suburban	13	(9)	134	(89)	3	(2)		
	Urban	0	(0)	39	(95)	2	(5)		
#9 Computer Test Scoring	Rural	19	(14)	105	(78)	10	(8)	13.03	.01
	Suburban	44	(29)	97	(65)	9	(6)		
	Urban	7	(17)	28	(68)	6	(15)		
#10 Computer Item Analysis	Rural	16	(12)	109	(82)	8	(6)	12.97	.01
	Suburban	31	(21)	94	(64)	23	(16)		
	Urban	5	(12)	31	(76)	5	(12)		
#11 Computer Grade Book	Rural	23	(65)	87	(65)	24	(18)	13.27	.01
	Suburban	50	(33)	73	(49)	27	(18)		
	Urban	12	(29)	26	(63)	3	(7)		

Grade Level Assignment

Resource Item	Teacher Group	Yes f	Yes %	No f	No %	Sometimes f	Sometimes %	X^2	p
#2 Student records	Elementary	118	(97)	1	(1)	3	(3)	7.88	.02
	Secondary	167	(87)	6	(3)	18	(9)		
#3 Counselor assistance	Elementary	78	(65)	22	(18)	21	(17)	8.08	.02
	Secondary	148	(78)	16	(8)	27	(14)		
#4 Curriculum Guide	Elementary	114	(93)	0	(0)	8	(7)	9.82	.01
	Secondary	159	(83)	12	(6)	20	(11)		
#5 Instructor Manuals	Elementary	95	(78)	7	(6)	20	(16)	5.67	.06
	Secondary	124	(66)	21	(11)	44	(23)		
#7 Weight Final Exam	Elementary	20	(16)	92	(75)	10	(8)	67.86	.00
	Secondary	122	(64)	61	(32)	8	(4)		
#8 Final Grade Scores Required	Elementary	10	(8)	105	(86)	7	(6)	5.77	.06
	Secondary	34	(18)	146	(77)	10	(5)		

(table continues)

Table 2 (Continued)

Resource Item	Teacher Group	Yes f	Yes %	No f	No %	Sometimes f	Sometimes %	χ^2	p
#9 Test Scoring Service	Elementary	15	(12)	101	(83)	6	(5)	14.36	.00
	Secondary	53	(28)	120	(63)	18	(9)		
#11 Computer Grade Book	Elementary	31	(25)	78	(64)	13	(11)	6.45	.04
	Secondary	51	(27)	100	(52)	40	(21)		
#12 Computer Test Generation	Elementary	24	(20)	82	(68)	14	(12)	12.03	.002
	Secondary	42	(22)	97	(51)	5	(27)		

Table 3

One-Way ANOVA's for Testing Practices by Grade Level and by Subject Area

Grade Level Assignment Means

Practice		Elem.	Sec.	Total	F Value	p
1.a	Calculate $\bar{X}$'s & SD's	1.58	1.89	1.77	8.67	.01
1.b	Calculate Reliability	2.12	2.31	2.23	1.94	.17
1.c	Do Item Analysis	2.20	2.46	2.36	3.84	.05
1.d	No. Tests Course***	2.45	2.14	2.26	6.51	.01
2	Write Own Items	2.66	4.12	3.57	96.87	.00
3	No. Tests Per Year	47.62	53.65	51.43	0.89	.35

Subject Area Specialization Means

Practice*	(1) Bus.	(2) Sci.	(3) Math	(4) Eng.	(5) Soc. St.	Total	F	p	Scheffe**
1.a	1.86	2.39	1.95	1.73	1.46	1.89	4.18	.01	5<2
1.b	2.42	2.51	2.44	2.03	2.00	2.31	1.48	.21	-----
1.c	2.84	2.39	2.58	2.17	2.09	2.46	2.99	.02	5<1
1.d	2.09	2.00	2.10	2.77	1.71	2.12	6.58	.00	4>1,2,3, & 5
2.	3.67	4.33	4.05	4.13	4.50	4.10	3.18	.02	5>1
3.	66.09	47.28	53.68	45.39	48.35	53.29	0.83	.51	-----

* See top section of this table for item descriptions.

** Scheffe' post-hoc pair-wise mean comparisons alpha @ .10.

*** Lower numbers here indicate more frequent test scheduling ('1' = one or more each week to '5' = two or fewer per semester).

Table 4

One-Way ANOVA's for Proportions of Test Item Types Used by Teachers by Grade Level and by Subject Area

Grade Level Assignment Means

Item Type	Elementary	Secondary	Total	F	p
Completion	18.97	15.33	16.75	3.00	.084
Matching	13.79	14.46	14.19	.19	.666
True/False	9.32	9.74	9.58	.13	.724
Multiple Choice	24.48	16.72	19.74	11.21	.001
Essay	7.33	13.31	10.98	10.06	.002
Problems	13.98	26.33	21.51	12.59	.001
Other	4.16	2.91	3.39	.55	.460
(N)	(122)	(191)	(313)		

Subject Area Specialization Means

Item Type	(1) Bus.	(2) Sci.	(3) Math	(4) Eng.	(5) Soc. St.	Total	F	p	Scheffe*
Completion	15.44	16.58	7.66	13.90	21.97	14.78	3.54	.008	5>3
Matching	14.38	20.33	3.41	15.57	19.56	14.09	11.29	.001	3<1,2,4,5
True/False	14.69	8.52	3.44	7.20	14.56	9.75	12.10	.001	3<1,5; 4<1,5; 2<1
Multiple Choice	17.00	26.36	3.17	23.63	18.75	16.98	13.10	.001	3<1,2,4,5
Essay	7.44	9.85	.32	29.87	21.06	12.39	21.93	.001	5>1,2,3; 4>1,2,3
Problem	26.47	15.48	78.76	1.17	1.25	27.66	106.55	.001	3>1,2,4,5; 1>4,5, 2>4,5
Other	2.38	2.12	3.24	5.33	2.65	3.07	.37	.832	
(N)	(45)	(33)	(41)	(30)	(32)	(181)			

* Scheffe' post-hoc pair-wise comparisons alpha @ .10.

7/2