COORDINATORS OF ACCOUNTABILITY VIEW TEACHERS' MEASUREMENT COMPETENCE

John R. Hills/Florida State University

In late November, 1976, at the suggestion of several of Florida's Coordinators of Accountability, I conducted a survey of their views of the competence of teachers in the measurement aspects of accountability. This is a report of the results of that survey.

First, I will report for each item the average response or the percent making each response. Then on one item in which you were asked to describe the worst mistake you could remember, I will report the kinds of mistakes mentioned, the number of each, and an example of each. There will be a few footnotes to clarify details. Finally, there will be a discussion of the results and some implications.

Results

Question 1. What percent of your district's teachers would you estimate show evidence in their work of having had thorough and effective instruction in educational tests and measurements, i.e., construction of classroom tests, administration and interpretation of standardized tests, grading, and using tests as part of their instruction?[1]

High School Level Teachers (Circle one)

			3.9		
All	75%	50%	25%	None	Don't work with high school teachers.

Junior High Teachers (Circle one)

			3.9		
All	75%	50%	25%	None	Don't work with junior high teachers.

Elementary Teachers (Circle one)

			3.9		
All	75%	50%	25%	None	Don't work with elementary teachers.

2. What percent of your district's teachers would you estimate would be noticeably more effective in their instruction and evaluation activities if they had some *in-service training* in educational measurement and evaluation?

High School Level Teachers (Circle one)

			2.3		
All	75%	50%	25%	None	Don't work with high school teachers.

Junior High Teachers (Circle one)

			2.4		
All	75%	50%	25%	None	Don't work with junior high teachers.

Elementary Teachers (Circle one)

			2.3		
All	75%	50%	25%	None	Don't work with elementary teachers.

3. From your point of view dealing with accountability, would you recommend that state certification require that all teachers have at least one substantial undergraduate course devoted entirely to educational measurement and evaluation? Yes 98% No 2%[2] (Circle one)

4. From your point of view dealing with accountability, would you recommend that recertification require that all applicants show on their transcript at least one substantial course devoted entirely to educational measurement and evaluation? Yes 73% No 27% (Circle one)

5. Have you observed any serious mistakes in use of tests and measurements (including grades) during the last year? Yes 62% No 38% (Circle one)

5a. If you answer "yes," please describe in detail below the worst mistake that you remember, without mentioning names of people or schools.

56%[3]
26 People

6. Twenty percent[4] of my district's teachers could choose the correct answer to the following question and justify it.

A teacher who wishes to compare her successive classes in relation to her own teaching objectives should most probably use

a. A criterion-referenced test.
b. A criterion-related validity coefficient.
c. A norm-referenced test.
d. A standardized commercially published test.

7. Ten percent of my district's teachers could choose the correct answer to the following question and justify it.

For best measurement (highest reliability), the proportion who get each item correct in a multiple-choice test should be

a. About .50.
b. Above .50 for every item.
c. Above .50 for the first items in the test and get more difficult gradually until the last items are answered correctly by much fewer than half of the class.
d. Below .50 for every item.

8. Ten percent of my district's teachers could choose the correct answer to the following question and justify it.

The scores of a large random sample of adults on tests of musical and mechanical aptitude yield a correlation coefficient of −.68. It may be inferred from this that in this sample

a. An individual cannot have high scores in both musical and mechanical aptitudes.
b. Low mechanical scores tend to be associated with high music scores.
c. Low mechanical scores tend to be associated with low music scores.
d. Musicians have about average mechanical aptitude.
e. There is only a very low relationship between musical aptitude and mechanical aptitude in this group.

9. Ten plus percent of my district's teachers could choose the correct answer to the following question and justify it.

The major advantage of norm-referenced tests over criterion-referenced tests is that norm-referenced tests

a. Are designed to maximize differentiation among students so that relatively few of them will have tied scores.
b. Allow comparisons of students in one class with students in other classes and schools by means of norms tables.
c. Have more carefully developed and chosen items.
d. Have built into them standards against which performance is to be judged.

10. Fifty percent of my district's teachers could choose the correct answer to the following question and justify it.

A student received a grade-equivalent score of 10.2. This score indicates that

a. He ranks in his class at the equivalent of a rank of 10.2 for the grade 10 students of the normative group.
b. He should be placed in the tenth grade in instruction in this subject.
c. His raw score is the same as the median score earned by all students in the norm group who were 10.2 years old at the time of testing.
d. His raw score on this test is the same as the approximate median of scores made by pupils in the second month of the tenth grade.

11. Twenty plus percent of my district's teachers could choose the correct answer to the following question and justify it.

One important advantage of grading on the basis of standards is that this method

a. Assigns a fixed proportion of the students in each class to each of the possible grade levels.
b. Insures that there will always be a wide spread in the distribution of grades awarded.
c. Permits the grades awarded to reflect the quality of instruction.
d. Ties the grades awarded to the pretest scores for the students.
e. None of the above.

12. Ten percent of my district's teachers could choose the correct answer to the following question and justify it.

Which of the following is *not* a basis on which the courts have agreed to hear a case concerning grade or graduation disputes between a student and his school?

When the appeal claims that

a. Grades were awarded on the basis of unreliable data.
b. Grades were influenced by nonacademic criteria.
c. The school's grading standards were unreasonable.
d. The student's instruction was ineffective.

The responses to the request for descriptions of the worst mistake that the respondent could

remember were varied and interesting. After listing each response and considering the entire group, I organized them into seven different categories and a small miscellaneous group. The categories, the number of responses in each category, and an example to give the flavor of the category follow in order of frequency of occurrence:

Category 1: Misinterpretation of scores
N = 20
Example: Teacher changed standard scores to grade-equivalent scores by inserting a decimal point between digits.

Category 2: Misuse of test results
N = 11
Example: Failure to recognize effect of item difficulty on "Observed student achievement."

Category 3: Poor teacher-made tests
N = 6
Example: Failure to understand differences in construction of norm-referenced and criterion-referenced tests.

Category 4: Failure to use tests
N = 6
Example: Teachers don't use the results.

Category 5: Poor test administration
N = 5
Example: Allowed only one half of the appropriate time for a standardized test.

Category 6: Dishonesty
N = 4
Example: Teacher marks the tests booklets for her elementary class.

Category 7: Errors in grading
N = 3
Example: Grades were reduced for punitive purposes.

Category 8: Miscellaneous
N = 2
Example: Teacher averaged scores of tests with different number of items.

Discussion

Just to clarify while the information is in your mind, you should note that while we only received reports of 57 mistakes; that does not imply that there were only 57 mistakes made in Florida last year! We asked for only the *worst* mistake observed by each respondent, so we have here a compilation of the kinds of errors that are most glaring to accountability coordinators. Perhaps some credit can be given to the idea that the kinds reported most often are most widespread, and certainly the reports of these kinds of errors suggest that atrociously ignorant procedures are being inflicted on students in Florida's counties, but this catalog should not be interpreted as a study of the nature and frequency of errors in measurement activities in Florida schools.

Clearly, this modest survey indicates that most of Florida's school teachers, at all levels from elementary through high school, fail to reflect in their school activities evidence of having learned the rudiments of tests and measurements. Only 25 percent were judged by coordinators of accountability to be displaying evidence of sound training in this area, and typically only 10 to 20 percent of the teachers were judged to have the knowledge to answer correctly questions about relatively simple and highly salient and practical aspects of classroom measurement. Perhaps most of Florida's teachers never had any training in this area; a course in it is not required for certification in Florida. If they did have training in it, the training apparently was ineffective. Or at least that is the way it appears to Florida's coordinators of accountability. In fact, nearly every coordinator who responded would recommend that a course in testing be required for certification.

The coordinators also report that at all levels of education from elementary through high school, most of Florida's teachers would be more effective in evaluation and in instruction if they had in-service training in measurement and evaluation. Since requiring a course in testing for certification would not reach those who are already certified but still incompetent in this important aspect of the profession, it is not surprising that three-fourths of the coordinators would recommend requirement of a course in testing before teachers could be recertified. The following three relatively simple steps might, then, make a dramatic reduction in the number of mistakes that are made:

1. Require a substantial course in testing for certification at all levels in all programs in the state.
2. Provide in-service training for teachers now in the schools who have not had such a course.
3. Require such a course for any teacher who applies for recertification.

Florida is not unique in having teachers in the field who are inadequately trained in measurement and evaluation. While it has been recognized for years that a large proportion of the serious problems in the work of teachers involves measurement, and you only have to listen to teachers around grading time to appreciate it, surveys over the years have revealed how little teachers know and are required to learn about testing. For example, Noll, in 1955, surveyed the requirements in various states and found that only 14% of 80 selected teacher-training institutions required an introductory course in measurement for graduation. Only 10% of the states specified a course in measurement for certification. Allen, in a similar study in 1956, found similar results. Mayo, in 1964, gave a test of measurement competencies to graduating seniors in 86 teacher-training institutions, and then retested them two years later while they were practicing teachers. He found that not only was their performance weak and showed but little improvement with practice, but less than one-half of them had taken as much as one full course in tests and measurements. Mayo concluded that there is a strong implication that a measurement course should be made compulsory for every teacher. Goslin, in a study for the Russell Sage Foundation reported in 1967, collected data from 1,450 teachers selected according to quota sampling procedures to represent the universe of more than 21,000 public secondary schools in the United States. He found that less than 40% of all teachers had more than minimal exposure (one course) to tests and measurement, and a sizable proportion not only never had a course but never even attended anything like a clinic in which testing was discussed. Elementary teachers reported a particularly striking lack of exposure to instruction in this area, while at the same time nearly three-fourths of the elementary teachers reported that they were routinely responsible for administering standardized tests to their pupils each year. Goslin concluded that explicit consideration should be given to the problem of teacher training in the field of measurement. Here we are in Florida, ten years later, having made no progress and continuing to have the same problems.

One of the most intriguing studies reported of the failure of teacher training to include testing and measurement, while including and requiring a multitude of other kinds of training, many of which are clearly of less significance, is that of Harold Roeder, a faculty member at the time (1973) at the State University College of Fredonia, New York. As a joke, Roeder bet a colleague that most elementary education majors are better prepared to conduct impromptu art and music lessons than they are to evaluate pupil performance. To win his bet, he conducted a nationwide survey of teacher preparatory institutions to see how many colleges and universities throughout the United States required prospective elementary teachers to complete a course in evaluation. He found that while less than half of the institutions reported that they required prospective elementary teachers to complete an evaluation course, nearly all of them required courses in art methods, music methods, and physical education.

About one-third of Roeder's colleges did require an evaluation course, but only half of these required a full three semester hours, and less than five percent required more than three semester hours. (A full three semester hours or four quarter hours is no more than enough time for motivated, volunteer students to master the fundamentals in the author's nonrequired course at Florida State University. If all students were required to take the course, many would find that it was an unusually heavy four-quarter-hour course.)

What makes Roeder's study especially intriguing is that he found that while only twelve institutions in his study required teachers to complete more than three semester hours in measurement and evaluation, three hundred and five institutions required more than three semester hours of religion! As Roeder states, this is clearly a ludicrous state of affairs in teacher preparation, but it is not something at which we should laugh. Remember, with less training in measurement than in religion, elementary teachers are expected to construct, select, administer, score, interpret, and implement the findings of standardized and informal evaluation instruments regularly throughout their careers as teachers.

Possible Action

It seems clear from this survey that the Coordinators of Accountability in Florida are much less than satisfied with the competence of Florida's teachers in the measurement and evaluation skills necessary for effective instruction and effective operation of a system of accountability. What can be done to correct this situation?

First, there is in Florida a Council on Teacher Education (COTE). It is a statutory advisory

council appointed by the State Board of Education to advise the Commissioner of Education on all matters dealing with teacher education and certification. It has made "essential competency studies" in order to develop and make recommendations for desirable standards relating to programs and policies for the development, certification, improvement, and maintenance of competencies of educational personnel. Thus far it has developed a set of generic essential competencies, i.e., the minimum essential competencies to be specified in state policy for certification of all educational personnel. These were approved by the Commissioner of Education in 1976 as competencies to be included in state policy for approval of teacher education programs.

The 23 approved Essential Generic Competencies include the following that can be expected to be developed in a course in tests and measurements:

7. Using diagnostic tests, teacher observations, and student records, diagnose the entry knowledge and/or skill of students for a given set of instructional objectives.

8. Construct and sequence related short-term objectives for a given subject area.

11. Select/develop and sequence related learning activities appropriate for a given set of instructional objectives and student learning needs.

14. Construct or assemble a classroom test to measure student performance according to criteria based upon objectives.

18. Identify and/or develop a system for keeping records of class and individual student progress.

Since these competencies are to be expected of all teachers and since they are stated without the associated enabling competencies which must be available in order for these general competencies to function, it might be appropriate for the Coordinators of Accountability to make an effort as a group to elaborate on these competencies in a manner to include that which is necessary for the Accountability Law to be effective. The result could be presented to the Council on Teacher Education. Representatives of the Coordinators of Accountability might want to request a hearing before the COTE to present their recommendations or a formal written report of their analysis and recommendations might be appropriate. appropriate.

A second possibility is that individual Coordinators of Accountability might approach their own legislators and ask them to pass legislation to support the Legislature's Accountability Law. There may be regulations that prevent Coordinators from approaching legislators as representing their schools or districts, but no one can prevent a concerned citizen from trying to get his representative informed and oriented toward sound legislation. If we have lost that right, we are nearly too far gone to rescue! On the other hand, going directly to the Legislature may not be the best approach. Sometimes legislation is beneficial in the short term but detrimental as conditions change, and it is hard to get the law appropriately updated.

A third possibility is that Coordinators of Accountability can approach individual institutions in Florida and try to get them to incorporate course work in tests and measurements in their programs for teachers as those programs are approved. One institution, Florida Atlantic University, now requires a substantial course in tests and measurements for all prospective teachers. If this requirement could become more universal, much of the Coordinators' problems with incompetent teachers and administrators in this area could be reduced.

Footnotes

[1] The value 3.9 for high school level teachers for Question 1 is a mean score value for all respondents on a scale in which All was given a value of 1, 75% a value of 2, 50% a value of 3, 25% a value of 4 and None a value of 5. The 3.9 is placed above 25% to indicate that the mean response was just slightly less than 25%, or 3.9 on the scale. This same procedure was used in Question 2.

[2] The percentages reported for questions 3, 4, and 5 are the percent of those responding to each question who gave each answer. Thus 98% of those of you who responded to this question indicated that state certification should require that all teachers have at least one substantial undergraduate course devoted entirely to educational measurement and evaluation.

[3]The 56% indicates that 56% of the respondents gave incidents of "worst mistakes." These are presented after the results of Question 12.

[4]The percent given is the median of the distribution of answers given to the question by those who responded to it. Distributions to many of these questions were markedly skewed, so medians rather than means are reported. Sometimes it is stated that a given percentage "plus" could respond knowledgeably to the question. The plus means that responses of early returns were compared with responses on late returns, and there was a significant difference between them. In both cases where this occurred, the difference was in the direction that late respondents gave a higher percentage. Since about one-third of the coordinators did not respond, one might conjecture that if they too had responded, they might have responded more like later than earlier respondents, and therefore the percentage might be biased slightly due to nonrespondents.

References

Allen, Margaret E. "Status of Measurement Course for Undergraduates in Teacher-Training Institutions." *13th Yearbook, National Council on Measurement in Education.* New York: NCME, 1956, 69-73.

Byram, Harold M. *Some Problems in the Provision of Professional Education for College Teachers.* Teachers College, Columbia University Contributions to Education, No. 576, New York: Bureau of Publications of Columbia University, 1933.

Davis, Robert A. "The Teaching Problems of 1075 Public School Teachers." *Journal of Experimental Education,* 1940, *9,* 41-60.

Goslin, David A. *Teachers and Testing.* New York: Russell Sage Foundation, 1967.

Mayo, Samuel T. *Pre-Service Preparation of Teachers in Educational Measurement.* Chicago: Loyola University, 1967. (Project No. 5-0807, Contract No. OE 4-10-011, U. S. Dept. of H. E. W.)

Noll, Victor H. "Requirements in Educational Measurement for Prospective Teachers." *School and Society,* 1955, *82,* 88-90.

Roeder, Harold H. "Teacher Education Curricula—Your Final Grade is F." *Journal of Educational Measurement,* 1973, *10,* 141-143.