

## From Hurdles to Standards of Quality in Teacher Testing

Asa G. Hilliard III, Fuller E. Callaway Professor of Urban Education, College of Education, Georgia State University

In 1983 Florida's state legislators proposed adoption of a merit-pay plan for its public-school teachers. One requirement in the plan that met with strong opposition, and was later dropped, was that applicants for merit pay must possess a master's degree in the subject they teach. The ramifications of this requirement are exemplified in the following excerpt from a 1984 *Education Week* article.

Only 36 percent of the teachers of the year are eligible to apply for merit pay. . . . The obstacle for most teachers of the year is the 1983 legislature's decision that only teachers with certain master's degrees—those specifically pertaining to the subject they teach—would be eligible for merit pay.

Twenty-one teachers of the year have no master's degrees. Seven have a master's degree that does not meet merit-pay guidelines. Two do not know where they stand because some of those guidelines need further interpretation.

Anna Wollard was chosen the best of the 1,000 teachers in Clay County. She also was named "competency reviewer" for the state to evaluate teachers considered below par to see if they could be helped. "Here your county says you are one of the best teachers and the state says you are an expert, but you can't get merit pay," Ms. Wollard said. "It's deflating."

Bernice McSpadden, Bay County's teacher of the year, is also a trainer of the evaluators who will help determine which teachers deserve merit pay. Because she has no master's degree, she herself is out of the running before the evaluation process even gains. . . .

Robert Bossong, Dade County's teacher of the year, is noted for his success with disruptive youths that other schools have given up on, but he cannot earn a master's degree in the vocational area he teaches because such a degree does not exist. . . .<sup>1</sup>

The Florida experience with competency-assessment practices

<sup>1</sup>Patti Breckenridge, "Florida Merit-Pay Plan to Exclude Many 'Teachers of the Year,'" *Education Week*, March 28, 1984, p. 11.

and merit-pay stipulations for public-school teachers provides insight into real-world problems that result from current attempts to set standards of quality for in-service teachers. Note that teachers who had performed well in the classrooms were not eligible for merit pay because they did not possess a prerequisite that was thought to be necessary for successful job performance. It was not. Therefore, serious questions must be raised about any "competency testing" movement or presumed prerequisite for successful job performance.

By any standard, the quality of public education in general leaves much to be desired, as recent national reports have shown. Since virtually all children in the nation attend public schools, this is a cause for alarm. It is an even greater cause for alarm when one ponders the fact that it is public school or nothing for education of the majority of African American children. Since African American families generally do not have much discretionary surplus income that would enable them to opt for private schooling, almost all members attend public schools.<sup>2</sup> Undereducated or uneducated Americans will become a burden on the society. Just as important, they will be unable to lead full, satisfying, and rewarding lives. Therefore, everyone should have a stake in the provision of quality education for all, not merely for a small elite.

#### AN APPROPRIATE FOCUS ON TEACHER SKILLS IS CORRECT

Teachers must be at the center of any effort to provide quality education for all children. It should be beyond debate that we need public-school teachers who are firmly grounded in academic content and valid professional skills. Both of these presuppose the possession by the teacher of what some have come to call "basic skills." Having content to teach and knowing how to teach it are teacher-education-outcome standards that are appropriate for all teachers, regardless of race or social class. Yet, these "basic skills" are not the ultimate criteria that tell us who is or is not a good teacher. Ultimately, good teachers are known by the quality of achievement of the children they teach.

Another point must be made. The quality of a student's achievement is a function both of the work of competent teachers and the presence or absence of a superior, responsive school environment. This is to say, schools and teachers *alone* can and do produce high quality academic achievement in children, regardless of the race or

<sup>2</sup>National Alliance of Black School Educators, *Saving the African American Child* (Washington, D.C.: NABSE, 1984).

socioeconomic background of the children.<sup>3</sup> It is true that some children also receive a tremendous boost from the support of their parents and community organizations which enables them to achieve in spite of poor school conditions. Further, it is true that some teachers overcome overwhelming odds and succeed with their students, even when the school's support environment is poor. Yet, neither the parents nor the poorly supported good teacher should be abandoned by those who are responsible for providing equal educational opportunity for all. School leaders are responsible for providing high-quality services to all children.

School boards, state education agencies, and the Federal government, therefore, have *two* responsibilities to discharge in the education of children. First, they must provide a quality school environment, i.e., physical plant, equipment, supplies and materials, support services, and school leadership. Second, they must provide teachers who meet appropriate academic and professional standards. As simple and easy as these responsibilities may sound, school leaders are well aware that providing quality education is a complex matter that requires a wealth of financial support and courageous professional and political decision making. When the public becomes broadly dissatisfied with the schools and begins to make strident criticisms of them, many school leaders and public-policy makers take the easy way out and commit themselves to simple-minded, "quick fix" solutions. For example, the major common element in most of the educational reform activity in the states is the so-called teacher-competency-testing focus. Some of the main reasons the competency tests are used are: (1) The tests are cheap to produce (usually paper-and-pencil and multiple-choice type so that they can be machine scored). (2) The minimum score requirements can be shifted up or down with ease. (3) The tests have "face validity," which is the least rigorous type of validity. (4) Because the tests are controversial and generate confused discussions, fundamental scientific problems can become obscure.

While the political climate is such that teacher competency testing has been forced upon the schools as a kind of panacea for quality educational problems, it would be sad indeed if professional opinion followed the political crowd. The real danger in acquiescing is that professionals who do so will lose the opportunity to call attention to fundamental problems with public-education policy, fiscal

<sup>3</sup>See the large volume of literature on "effective schools" that emerged in the late 1970s to mid-80s, particularly the summer 1985 issue of the *Journal of Negro Education* on successful schooling policies, practices, and programs.

support, and, above all, with the nationwide lockstep approach to what I believe to be invalid competency testing itself.

The matter of teacher competency testing is situated in a complex system of education. It cannot be evaluated by itself outside that system. The idea of requiring that professional educators demonstrate an acceptable level of expertise is entirely appropriate. However, the method of that demonstration must be rational, meaningful, and valid. For example, there still is no universal or common agreement on the elements that go to make up the content of traditional academic subjects in high school. An acceptable try at delineating such elements has been made by the College Board<sup>4</sup>—with excellent results. However, the fact that such comprehensive effort is so late in coming (long after the widespread adoption of the teacher-competency-testing practice), and that even now it is neither universally known nor accepted, is evidence enough to conclude that it is absurd to believe that currently used tests are "measuring with precision" the academic knowledge that a teacher needs in order to be licensed. The public-school curriculum, the college curriculum, and the content of an academic competency test must overlap to an acceptable degree. If they do not, the test is not a valid measure of academic competency. Tests that do not have this overlapping element must be regarded as hurdles, not as standards of quality—no matter how esoteric or face-valid they may appear.

The situation is even more absurd in the case of tests of "professional knowledge." A common knowledge base in professional education has yet to be identified or supported by the majority of professional educators. While important strides have been made in educational research and development, the essential skills have not been specified in such a way that teacher educators offer common professional experiences to teachers in college preprofessional programs. How can precision measures be made of an undefined body of professional knowledge? More absurdity!

A part of the total system of education includes student outcomes. Some of the student outcomes desired can be assessed through valid paper-and-pencil tests. However, important student outcomes such as work habits, attitudes, and creativity cannot be so assessed. By allowing the paper-and-pencil teacher-competency test to become the critical filter, determining who will be allowed to teach, public policy does not guarantee standards of excellence;

<sup>4</sup>College Board, *Academic Preparation for College: What Students Need to Know and Be Able to Do* (New York: College Entrance Examination Board, 1983).

rather, it merely requires the measure of a level of performance on the lowest form of assessment. It is on this low form of assessment of extremely questionable validity that teachers must meet minimum "standards." This then is the real problem with teacher competency testing. The problem is not merely that tests may discriminate or are biased against minorities; any normative nationally standardized test will be discriminatory and biased. The problem is one of reliability and validity. The important question is, Do we learn enough from the tests to choose reliably and validly the teachers who can get students to meet our achievement goals? The opposite important question also is not about bias against minorities; rather, it is, Do we learn enough from the tests to choose reliably and validly those teachers who *cannot* get students to meet our achievement goals?

One can only conclude from present practices in standardized paper-and-pencil teacher competency testing that educators and policy makers have no serious interest in the answers to the two questions raised above. Yet, these questions cover the very goal that policy makers say has motivated them to institute competency testing in the first place. Therefore, at present the teacher-competency-testing movement fails on two important grounds. First, it fails on scientific grounds, even before sophisticated statistical validation techniques are applied, since professional practice and sound content are not defined. No test-construction convention can correct for the absurdities mentioned above or for many others not mentioned here. The movement fails also on even more important grounds. It fails because policy makers appear incapable of correcting their reliance on a hopelessly flawed technology. If this is done in ignorance, it is bad enough; but if the issues are understood, and if policy makers persist in supporting invalid assessment, then they should openly and honestly declare their commitment to *hurdles* and abandon all pretense at a commitment to standards of excellence. Such behavior is confusing at best, and is an impediment to true progress toward attainment of high standards.

Perhaps the worst potential effect of the minimum-competency-testing movement on African American people in particular, and on others in general, comes from the fact that the movement focuses attention on the wrong aspects of teaching. Teaching involves many skills, attitudes, and understandings. There is some evidence that supports the assertion that minimum-competency tests do not tap the most important teaching behaviors—e.g., establishing and maintaining rapport with students, helping students to become motivated, facilitating communication groups, and stimulating critical consciousness. For example, the Marcus Garvey Elementary

5. There must be substantial *empirically demonstrable criterion-related validity for each component of the assessment process*. This means that student achievement must be utilized in criterion-related validity studies.
6. A *valid and comprehensive set of measures for child outcomes*—academic and others that are of interest—must exist. In other words, more than standardized paper-and-pencil multiple-choice test questions are needed to measure these child-behavior criteria.

It should take little effort to determine that few and probably no presently used teacher-competency tests can meet even one of the above criteria. Yet it is possible for the criteria to be met. To do so would involve a simple matter of cost. Quality assessment is not cheap! For example, an empirical check of the true curricula offerings would require the services of trained observers and interviewers, an examination of syllabi, and so on. No valid judgment about content validity can be made by "expert judges" who have not seen empirical data about course offerings and have not examined them systematically.

Children's lives and teachers' careers are at stake. This calls for the most rigorous application of the criteria to the field of competency assessment.

Many criticisms have been made alleging cultural or racial bias in standardized teacher-competency tests, especially bias against African Americans. It cannot be emphasized too strongly that, while bias can be shown to exist in any standardized test, there is a more fundamental problem with teacher-competency tests. The fundamental problem is one of content or predictive validity for any teacher, regardless of race. Content validity is virtually impossible to demonstrate at present in light of the almost universal failure of educators to meet the criteria for valid competency assessment mentioned above. Moreover, predictive validity cannot be achieved for the same reason. This places a heavy burden of proof on test developers and advocates for the use of tests that have disparate impact on African-American or other populations. Developers and advocates of tests must be able to prove that tests are not arbitrary and capricious. At present this cannot be done.

#### WORKING CONDITIONS FOR THE COMPETENT TEACHER

It is often in the low-income, urban, public schools with heavy concentrations of cultural minority groups that student achievement is low. It is such school environments that prompt concern about educational quality, and about the quality of teacher competencies. Yet, a close look at some urban teaching environments will reveal major non-teacher causes for low student achievement.

For example, the problem of overcrowded schools causes serious interruptions in classwork, lack of physical facilities to accommodate support services, teachers' inability to provide for individual needs, among others.<sup>5</sup> It seems that many policy makers and educational leaders who are under severe pressure find it easier to require higher performance of teachers on tests than to change the terrible conditions under which many instructors are forced to work.

Some of the conditions that produced low student achievement had nothing to do with teacher competency or the lack of it. It is important that these contributing factors be exposed, for they are the ones that must be changed if policy makers are serious about their stated goals for student achievement. Some of the more serious conditions that must be addressed are:

1. A confused curriculum that does not require students to enroll in courses that are prerequisites to later tests.
2. The failure to provide sufficient sections of required courses for all students who need and/or request them.
3. Patterns of teacher assignments that expose some children to teachers who are working outside their areas of academic and/or professional preparation.
4. School facilities that do not provide for adequate and appropriate learning space.
5. A shortage of essential school supplies and equipment.
6. Failure to provide the appropriate work load for competent teachers, e.g., overloading teachers who are expected to teach writing skills.
7. Failure to provide competent building-level leaderships, i.e., principals.
8. Failure to protect competent teachers from the imposition of a disorganized set of central-office programs, some of which are at cross-purposes with each other or with the teachers' own well-organized lessons, e.g., programs that allow for some children in a given class to use materials that are withheld from other students in the same class who could use the materials profitably.
9. Support of student assessment practices that misdiagnose and misplace students into special classes where little is required of them academically.
10. Failure to provide at the school site or within the school district

<sup>5</sup>For examples of problem of overcrowdedness and its effects, see L. Rother, "Crowded Public School Holding Classes in Halls," *New York Times*, November 18, 1985, p. 19; and "Intrusions Abound for Young People," *ibid.*

opportunities for professional discussion to take place among peers on the solution of teaching and learning problems.

11. The tendency to allow large units to swell to enormous proportions—large districts or schools—causing the focus of leaders to shift from instructional priorities to power struggles.
12. Failure to reward excellent teaching in tangible and meaningful ways.

The above certainly is not an exhaustive list. However, the purpose for presenting such a list here is to provide a set of points of reference for use in evaluating the will and motivation of educational leaders and policy makers to move toward true quality in education. If the basic matters listed herein and on similar lists are not addressed with the same level of attention that is presently being received by teacher competency testing, then a good case can be made that policy makers are creating hurdles for teachers, not standards of quality.

#### DATA NEEDS FROM THE TEACHER-COMPETENCY-TESTING MOVEMENT

After all is said and done, there are those who still might choose to argue that a "quick and dirty" paper-and-pencil, multiple-choice test may yield enough meaningful information that is related to teaching performance to justify its use. Because of the possibility that such an argument will be made, it is essential that data be collected which will permit an evaluation of the impact of the use of such teacher-competency tests.

An appropriate question to be raised is, What does a cost benefit analysis show about the use of existing teacher-competency tests, particularly the multiple-choice, paper-and-pencil tests that are so widely used? It is the faddish character of this national movement that calls for rigorous professional evaluation. Almost overnight, we have witnessed the spread of the teacher-competency-testing practice to the vast majority of our states. In some states, the rush to develop teacher-competency tests has produced embarrassing activities. For example, uniform state-level requirements in curriculum in teacher education have followed rather than preceded the development of the examination intended to measure competencies!

When we speak of "teacher-competency" tests, we may mean any one of at least three types of tests: (1) the "basic academic skills" type, usually including reading comprehension, basic arithmetic computation, and sometimes written English expression; (2) the professional knowledge type; and (3) the academic knowledge type. Any one or all three types of tests may be used for several purposes,

among which are (1) certification of prerequisite skills for entry into or exit from teacher-education programs, (2) certification of professional competence, (3) employment and/or retention/tenure decisions, and (4) selection of merit-pay recipients.

Complicating matters further is the fact that each of the fifty states makes unilateral decisions, usually different ones, on which types of tests to use, the purpose for which they are to be used, and the particular test that will satisfy what are perceived to be unique needs. Given three or more types of tests, three or more purposes, and an unlimited number of tests under each type, it should be clear to even the most casual observer that there is no meaningful general definition of "teacher competency." Therefore, we may say that what most professionals and public-policy makers refer to when they use the term "teacher competency" is something that is arbitrary and situation-specific, not generalizable beyond certain school districts and states. In the mind of the general public, teacher-competency judgments are not arbitrary. They are seen as being scientific and rational, especially when standardized tests are used. Therefore, without having access to the data on the actual arbitrariness and irrationality of the present form of paper-and-pencil standardized tests, almost any standardized test has face validity in the eyes of the general public. This should highlight the need for systematic data collection. Rational judgments about the validity of tests must be based on scientific data about the real world.

#### SUMMARY AND CONCLUSION

Educational leaders and public-policy makers must make a strong commitment to the assessment of quality in teaching. However, quality in teaching is manifested most strongly in the quality of achievement of the students who are taught. This means that appropriate valid assessment of the achievements of students is an indispensable prerequisite to the valid assessment of teacher competencies.

The state of the art in assessment technology is sufficiently developed to permit the creation of valid assessment of both teacher and student competency. Yet, valid assessment is not cheap assessment. Further, valid assessment requires more than paper-and-pencil, multiple-choice standardized tests.

Appropriate concern for the quality of teaching includes a concern for the quality of the school environment. No matter how competent the teacher and the student, they can be defeated in the struggle to reach school objectives when they are placed in school environments that are not conducive to achievement. Therefore, if

educators and policy makers are serious about improving the quality of education for all students, they must not expect the teacher-competency tests to be the major tool for educational reform.

#### SELECTED REFERENCES AND BIBLIOGRAPHY IN STANDARDIZED TESTING

- Algera, J. A., Paul G. W. Jansen, Robert A. Roe, and Pieter Vijn. 1974. Validity Generalization: Some Critical Remarks on the Schmidt-Hunter Procedure. *Journal of Occupational Psychology* (Great Britain) 57:197-210.
- American Psychological Association. 1981. Ethical Principles of Psychologists. *American Psychologist* 36:633-638.
- . 1966. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association.
- Atlanta Constitution. 1984. Progress by Black Grad Students Noted. (Reports on the Research of Dr. Richard Scott and Dr. Marvin Shaw, University of Florida.), p. 3A.
- Bullock, R. J., and Daniel J. Svyantek. 1985. Analyzing Meta-Analysis: Potential Problems and Unsuccessful Replication, and Evaluation Criteria. *Journal of Applied Psychology* 70:108-115.
- Chase, Allen. 1977. *The Legacy of Malthus: The Social Cost of Scientific Racism*. New York: Alfred A. Knopf.
- College Board. 1983. *Academic Preparation for College: What Students Need to Know and Be Able to Do*. New York: College Entrance Examination Board.
- Education Week* (various articles on the testing programs).
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. 1978. Adoption by Four Agencies of Uniform Guidelines on Employee Selection Procedures. *Federal Register* 43(166):38290-38315. (Plus questions and answers to clarify and provide a common interpretation of the UGESP. *Federal Register* 44(43):11996-12009 and *ibid.* 45(87):29530-29531.
- Garcia, Peter. 1985. A Study on Teacher Competency Testing and Test Validity with Implications for Minorities and the Results and Implications of the Use of the Pre-Professional Skills Test (PPST) as a Screening Device for Entrance into Teacher Education Programs in Texas. Edinberg, Texas: Pan-American University.
- Gould, S. 1981. *The Mismeasure of Man*. New York: W. W. Norton and Co.
- Heller, Kirby A., Wayne H. Holtzman, and Samuel Messick, eds. 1982. *Placing Children in Special Education: A Strategy for Equity*. Washington, D.C.: National Academy Press.

- Hilliard, Asa G., III. 1984. I.Q. Thinking as the Emperor's New Clothes. In Cecil Reynolds and Robert T. Brown, eds. *Perspectives on Bias in Mental Testing*. New York: Plenum Press, pp. 139-169.
- . 1983. Psychological Factors Associated with Language and the Education of the African American Child. *Journal of Negro Education* 52:24-34.
- . 1982. The Sociopolitical Implications of Minimum Competency Testing. In John H. Neel and Shirley W. Goldwasser, eds. *Minimum Competency Education: Issues, Methodology and Policy for Local School Systems*. Atlanta: Georgia State University, pp. 86-106.
- Kamin, Leon. 1974. *The Science and Politics of I.Q.* New York: John Wiley & Sons.
- Medley, D. M. and T. J. Quirk. 1974. The Applications of a Factorial Design to the Study of Cultural Bias in General Culture Items on the National Teacher Examinations (NTE). *Journal of Educational Measurement* 11:4.
- Messick, S. 1980. *The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the '50s to the FTC*. Princeton, N.J.: Educational Testing Service.
- Nairn, Allen and Associates. 1980. *The Reign of ETS: The Corporation That Makes Up Minds*. (The Ralph Nader Report on the Educational Testing Service.) Washington, D.C.: Learning Research Project.
- National Alliance of Black School Educators. 1984. *Saving the African American Child*. Washington, D.C.: NABSE.
- New York Times*. 1985. Intrusions Abound for Young People. November 18, p. 19.
- Orwin, Robert G. and David S. Corday. 1985. Effects of Deficient Reporting on Meta-Analysis: A Conceptual Framework and Reanalysis. *Psychological Bulletin* 97:134-147.
- Owen, David. 1985. *None of the Above: Behind the Myth of Scholastic Aptitude*. Boston: Houghton-Mifflin.
- Rother, L. 1985. Crowded Public School Holding Classes in Halls. *New York Times*, November 18, p. 19.
- Slack, Warner V. and D. Porter. 1980. The Scholastic Aptitude Test: A Critical Appraisal. *Harvard Educational Review* 50:154-175.
- White, David. 1981. *Towards a Diversified Legal Profession: An Inquiry into the Law School Admissions Test, Grade Inflation, and Current Admissions Policies*. San Francisco: Julian Richardson Associates.