

## **Teacher Testing and Implications for Physical Education**

**Mary O'Sullivan and Deborah Tannehill**  
The Ohio State University

There has been a dramatic increase in teacher testing in the last decade. State and national attention to this issue is evidenced by the amount of literature devoted to assessment in general and teacher testing in particular. An invitational conference was held last fall by the Educational Testing Service (ETS) to provide a forum for clarifying issues regarding uses of standardized tests in American education in general and teacher testing in particular. Flippo (1986) outlined how all but six states in the nation are presently involved in planning, designing, or implementing some form of teacher testing. Darling-Hammond (1986) predicts that almost all states will have teacher testing by the early 1990s. Haney and Madaus (1989) reported that the volume of standardized testing of teachers and students increased 10 to 20% in the last 40 years. Organizations supporting some type of teacher testing include the National Education Association, the American Federation of Teachers, the Holmes Group, the Carnegie Foundation, and the National Council for the Accreditation of Teacher Education.

The major purposes of this article are to outline the history of teacher testing in this county, describe current trends and alternatives to traditional forms of teacher assessment, and draw some implications for physical education teacher education.

### **History of Teacher Testing**

Kinney (1964) suggested that "the first official endorsement of teacher testing occurred as early as 1686 when the General Assembly of Virginia requested the governor to appoint persons to examine schoolmasters. The exams were oral in nature and screened candidates not only for their knowledge of subject matter but also assessed their morality and "muscle and courage" so they might keep order among their charges.

The growth in public education in the 1800s witnessed the transition of control for teacher evaluation from local to county and ultimately to state governance.

---

The authors are with the School of Health, Physical Education and Recreation, The Ohio State University, Columbus, OH 43210.

This was part of a larger trend to greater state involvement in education. The belief was that more objective testing would eliminate incompetent teachers. Haney, Madaus, and Kreitzer (1987) concluded of the process and format of teacher testing at the time that,

systems for testing teachers were closely related to the evolving centralization of school governance. Second, the primary topics of teacher examinations and the main basis for teacher selection were knowledge of what was to be taught and the moral and social suitability of the teacher candidates. Although pedagogy and teaching methods gradually came to be included as examination topics, preparation in these subjects was, at the turn of the century, by no means universally required. (pp. 175-176)

The main developments in teacher testing in the early part of the 20th century saw increased centralization of teacher testing and certification at the state level, a movement toward written rather than oral tests, and the introduction of objective tests for teacher selection and certification (Haney et al., 1987). The rise in the school population saw growing state governance of teacher testing, and by 1921 certification in over half of the states was handled by state authorities. In the early part of the 20th century, passing an examination was all that was needed to teach in most states and there was growing criticism of such examinations, considered by many to be so simple that high school students could pass them. By 1930 teacher certification by examination was replaced by an increase in teacher training, with many normal schools extending their programs from 2 to 4 years (Evenden, 1933).

The late 1930s and early 40s saw the growth of mental measurements, particularly Ben Woods' work on subject matter testing in Pennsylvania. Administration of these standardized tests of subject matter competence, for high school students in particular, was administered under Wood through the Cooperative Test Service. Some of the most prominent educational measurement experts were involved in these early efforts, including Thorndike and Tyler. The Cooperative Test Service also developed the first draft of what is now the National Teacher Examination (NTE) initially to help several eastern school superintendents select teachers. Wood then obtained a grant from the Carnegie Corporation and the American Council on Education to build new teacher examinations, and in 1949 the new teacher tests were administered to 3,500 candidates. With the establishment of the Educational Testing Service in 1947, the NTE was transferred to that agency.

The original test took 8 hours and covered "general quality of intellectual performance, knowledge of general culture and contemporary affairs, and professional information" (Haney et al., 1987, p. 179). The nature of the exam suggested that what was desirable in teacher candidates was a broad range of intellectual skills and learning rather than specialized education knowledge. This view of teacher testing is strikingly similar to today's use of teacher tests. The Pre-professional Skills Test and the California Basic Educational Skills Test, which are administered chiefly to those entering teacher education programs, focus on the basic skills of reading, writing, and math.

With the shortage of teacher candidates in the early 1940s, fewer signed up to take the test. ETS turned for financial support to teacher training institutions rather than urban school superintendents. The financially strapped organization curtailed the test from 8 to 4 hours and, perhaps as a consequence of their new benefactors, increased the weight given to professional education material in the NTE to 40%. Wilson (1984) recounts a rise in the fortunes of the NTE in the 1950s and early 60s due to increased use of the NTE in the Southeast and suggests it was related to efforts to desegregate schools in that region:

In 1954, shortly after the Brown decision by the U.S. Supreme Court, Arthur Benson (then head of the NTE program) pointed out to southern school officials that black and white teachers tended to score differently on the teacher examinations. He suggested that with the use of the exams "the south [could] face its future with confidence." (Wilson, 1984, p. 306)

Three significant changes in the NTE in the early 1970s resulted from outside pressure on the ETS from the National Education Association, who criticized the misuse of the NTE for promotion and tenure decisions in several states, and from minority educators, who criticized the test's bias against minorities (see Table 1). After NTE officials convened a group of minority educators to review the test, efforts were made to include materials to reflect minority experience and culture. The length of the test was curtailed from 4 to a little over 3 hours. Professional knowledge was reduced by combining the three professional educa-

**Table 1**  
**Extent to Which NTE Discriminates by Race\***

Category	Pass		Fail	
	<i>N</i>	%	<i>N</i>	%
<b>NTE communication skills</b>				
White	34,026	93.6	2,326	6.4
Black	2,697	48.1	2,910	51.9
Total	36,723	87.5	5,237	12.5
<b>NTE general knowledge</b>				
White	36,817	92.6	2,942	7.4
Black	2,733	39.1	4,258	60.9
Total	39,550	84.6	7,200	15.4
<b>NTE professional knowledge</b>				
White	36,309	93.9	2,359	6.1
Black	2,951	52.1	2,714	47.9
Total	39,260	84.6	5,073	15.4

\*Adapted from Haney, Madaus, & Kreitzer, 1987, p. 226.

tion tests into one. At the same time, the number of subject matter specialty tests increased from 9 in 1953 to 26 in 1985.

The 1980s saw efforts by ETS officials to professionalize the NTE test by increasing the portion of the examination devoted to professional knowledge. The irony was that several states wanted to screen prospective and competent teachers on the basis of their general knowledge, not their professional knowledge. The ETS obliged state officials by helping states develop customized teacher tests such as the Texas Pre-professional Skills Test.

### Current Trends in Teacher Testing

Teacher testing as currently implemented in most states utilizes traditional paper and pencil formats from one of a variety of standardized tests. Two major developers of teacher tests are the Educational Testing Service and the National Evaluation Systems. The former continues to administer the NTE and the latter has developed several tests including customized standardized teacher tests for Alabama, Georgia, and Oklahoma. Teacher tests are classified into three categories: (a) admission tests to screen prospective teachers applying to teacher education programs, (b) tests to gain initial certification to teach, and (c) tests given to practicing teachers as a condition for recertification. The Texas Pre-professional Skills Test is the most frequently used test to gain admittance to teacher preparation programs, although in some states it is used to obtain a provisional teaching certification and entry to teaching. Nationwide, the NTE is the most commonly used certification test for entry to the teaching profession. Several states have designed customized tests for their geopolitical region that are adaptations of the NTE.

#### *Reasons for Testing*

Arguments for national teacher testing can be summarized as follows. First, there is a widespread need to improve education in America and the assumption is that teacher testing can contribute to improved standards of teaching and student learning. Second, there is a belief that many teachers in our schools are incompetent and that some measure of teacher testing may be useful in screening out the least prepared of these candidates. In a study of teacher incompetence, Bridges (1986) suggested "The most common type of failure is weakness in maintaining discipline . . . [and] is the leading cause for dismissal" (p. 5). Haney et al. argued that there

is no good evidence available to indicate that the proportion of teachers who are incompetent is any greater than the proportion of other professionals such as doctors, who are incompetent. [Evidence has not] been provided to indicate that currently used tests are significantly related to general teacher competence or performance. (1987, p. 216)

A third argument made for teacher testing is that although we cannot measure teacher quality directly, we can and should measure teachers' knowledge of subject matter. This should include attention to basic skills, literacy, and professional and general knowledge to at least ensure that they are minimally compe-

tent. Haney et al. (1987) argue that not only are present teacher tests low predictors of teacher quality but that such testing is highly inefficient in terms of the relatively high number of false acceptances and false rejections for a test that at least 90% of the candidates are expected to pass.

Finally, since we cannot measure general teacher quality, we must instead rely on validation in terms of the content of teacher education programs. If academic preparation is the yardstick, then what is actually being tested? Is preparation in content knowledge, pedagogical knowledge, essential knowledge, or content knowledge necessary to teach? A response to any of these questions places the purpose of assessment in education as improving instruction rather than the outcome of instruction, student learning.

What, then, is the purpose of teacher testing? Is the purpose political or educational? Does the growing number of states using standardized tests reflect an effort to appease the public, to provide measures of accountability, or to improve educational outcomes? Are the tests used appropriately, and who should be responsible to monitor their use? ETS will not allow the NTE to be used as a test for inservice teachers. Yet the Texas Pre-professional Skills Test is used in some states as an entrance test to a professional preparation program while in other states it is used as an entry test to the teaching profession.

What becomes of teacher preparation programs if the curricula are driven by the certification tests? Are not teacher tests really instruments of educational reform guiding our ideology by signaling changes in the curriculum? Lauren Resnick (1989) suggests we develop tests that can be taught to without apology. They must become educational tools rather than curriculum alignment measures. They must not become old wine in new bottles.

### *Alternative Displays of Knowledge*

As the quantity of standardized paper and pencil tests has become so great, so also have the complaints about standardized testing. The Carnegie Foundation financed Lee Shulman and his colleagues at Stanford to examine alternative strategies to assessing the effectiveness of teachers. The National Certification Board is also working on a national volunteer teaching license that will include a multifaceted assessment procedure of teaching effectiveness.

Shulman (1988) argues that since educational reform efforts place the locus of control for decision making at the school and local levels, standards for teacher competence must necessarily rise and become more explicit:

Twenty-first-century conceptions of school reform and the professionalization of teaching cannot co-exist with early twentieth-century models of teacher testing and evaluation, especially when they yield unacceptably simplistic definitions of teaching. [Educators] need to become more proactive in asserting the purposes and conditions for teacher assessment. (p. 39)

One of the main features of this work is the movement from paper and pencil tests to alternative displays of teachers' knowledge that are closer to the reality of life in the classroom. The purpose of assessment from Shulman's perspective is to improve the teaching profession by making teachers and practicing educators more reflective about what they do (Brandt, 1988). Shulman does not sup-

port teacher assessment as a selection tool but as a tool "to provide formative feedback to improve the quality of those engaged in learning to be teachers" (Brandt, 1988, p. 43). Assessment procedures should also improve teacher education. He sees assessment as a process that

unfolds and extends over time, in which written tests of knowledge, systematic documentation of accomplishments, formal attestations by colleagues and supervisors, and analyses of performance in assessment centers and in the workplace are combined and integrated in a variety of ways to achieve a representation of candidate's pedagogical capacities. (1988, p. 38)

Shulman (1988) argued that while each approach is flawed, together the flaws of individual approaches are offset by the virtues of their fellows, in what he termed the "union of insufficiencies" (p. 36). His ideal assessment combines the virtues of portfolios and direct observation as sources of information that would be more faithful to the practice of teaching than are current approaches. Teacher assessment would conclude with the completion of simulation exercises at an assessment center that would follow up on the contents of a portfolio created by the teacher as well as other simulation exercises. All of these would be supplemented with "a new generation of written examinations less fragmented and discontinuous than the current crop" (Shulman, 1988, p. 40). The final section of this paper describes these assessment procedures in more detail and discusses the application of a similar assessment model for physical education.

### **Teacher Testing and Implications for Physical Education**

The type of test now used to test prospective teachers of physical education is narrow in focus and fails to address much of what is known about teaching physical education from over a decade of research on physical education teaching. The specs for the revised specialty test in physical education for the state of Ohio look like a repackaging of the old ones. A total of 150 items on the test cover six general content areas, with a specific weighting assigned to each area: history and philosophy of physical education (8%), scientific foundations (25%), curriculum development and planning (30%), organization and administration of total physical education program (22%), professional responsibilities (knowledge of organizations, resources, ethics and professional research) (5%), and evaluation (5%). According to IOX Assessment Associates (authors of the test), "the test questions include assessment of higher level cognitive skills, such as comprehension, application and analysis, and synthesis and evaluation" (personal correspondence, November 9, 1989).

A brief perusal at the specifications shows the absence of even one test item that assesses prospective teachers' specific knowledge of their subject matter, if defined as the technical aspects of motor skills, strategic approaches to sport, training implications for improved performance, and developmental considerations, norms, values, and traditions of various sports and/or how they relate to culture. Written communication with the authors of the specifications sought a rationale for this omission. The response was that the physical educators who developed the test felt "that the rules, field demarcations and so on could be easily

ascertained from a reference book." It was clear that, at best, the authors of the specifications did not understand the concept of physical education content knowledge or, at worst, saw little need for content knowledge in the assessment of prospective physical educators. If we are going to assess physical educators on information that is essential to them as effective teachers, what content knowledge fits this category?

### *What is the Nature of Physical Education?*

What are the realities of teaching physical education? How can the knowledges, skills, and attitudes needed as an effective physical educator be assessed? Several groups (Haney & Madaus, 1989; Peterson & Mitchell, 1985; Shulman, 1987a) have suggested multiple lines of evidence such as teaching performance, student and parental ratings of teachers, and portfolio documents together with specific exercises that would require the teacher to critique and respond to aspects of his/her own and others' teaching. How might Shulman's model of assessment be applied to physical education?

Written examinations enjoy high reliability in scoring. They permit broad sampling of material in a relatively economical way. Their weakness lies in their remoteness from the complexity of the teaching environment and allows one to display knowledge about teaching only in written form (Brice Heath, 1989). Written displays of knowledge can be used to test physical education content prior to application of knowledge through another assessment format. The standardized tests currently employed (e.g., NTE specialty test) are outdated. Newer forms of essay examination and experimental methods that employ computers to administer and manage objective tests promise further improvement. The seeming ignorance of prominent leaders in physical education about issues of validity and formats of teacher testing (see Tannehill and O'Sullivan, 1990) does not bode well for proactive leadership in this area.

Direct observation of teaching performance is widely employed for a permanent teaching license, particularly in the southern states. "While the full complexity of teaching can in principle be observed, the problem of sampling is staggering" (Shulman, 1988, p. 39). Direct observations fail in many ways to tap some of teaching's critical dimensions, yet direct observation of physical education teachers can play a significant role in the assessment of teachers' effectiveness. Teaching skills can be tested using systematic observation strategies and/or narrative descriptions of the teaching context. We have several qualitative instruments as well as appropriate quantitative methodologies with which to do such observations and analyses (Darst, Zakrajsek, & Mancini, 1989).

Alternate displays of a teacher's knowledge could be gathered through a portfolio of the teacher's work. This could include evidence of the teacher's lesson plans, instructional units, instructional aids, and/or computer programs that display student test scores, fitness levels, motor skill proficiency, and so forth. Shulman proposes that portfolio material be collected over an extended period of time and represent the candidate's "coached performance." In some ways it is not unlike the assistant professor's dossier which is developed over time and submitted as evidence of necessary criteria (competence) for a promotion and tenure decision made by peers. Shulman sees the portfolio development as an occasion for interaction and mentoring among peers and would "require that every

portfolio entry be cosigned or commented upon by a peer" (1988, p. 40). What might be some appropriate artifacts of a physical educator's portfolio?

The fourth approach to teacher testing is simulated exercises that simulate real problems and processes of teaching. Shulman and his colleagues have to date designed at least 20 such exercises for classroom teachers. One exercise asks the teacher to analyze a textbook and plan to adapt it for use in the classroom. An analogy to this in physical education might be to have teachers take a school district curriculum guide and adapt it for use in their school. A second exercise presents examples of students' work in their subject area and asks the teacher to respond to the students' errors and insights. Physical education teachers need to observe motor performance, analyze the performance, decide on major errors that need to be intervened upon first, and develop a strategy to improve that aspect of the performance. A third exercise presents teachers with a videotape of teaching and asks them to observe, critique, and recommend what they might do under similar conditions. All of these assignments call on the teacher to reflect on the process of teaching and learning and the complexity of the task. Shulman believes performance assessment of the type described

can reflect the complexities of teaching more faithfully than do test items. . . . Even though they are far more realistic than tests, they still cut off the teacher from the actual contexts of his or her own teaching—the school, the students, and the history of teaching and learning they share. (Shulman, 1988, p. 38)

The teacher's portfolio materials can be the connection with the "contexts and personal histories that characterize real teaching" (p. 39). These four approaches together can provide a more faithful representation of the realities of teaching. It behooves us to contemplate the nature of such assessment procedures in physical education and to consider the implications of such assessment for teacher education programming.

### *Implications for Teacher Education Programs in Physical Education*

Developing an assessment tool has fundamental implications for the process of teacher education. According to Shulman, it

renders the questions about the definition and operationalization of knowledge in teaching as far more than academic exercises. If teachers are to be certified on the basis of well-grounded judgments, then those standards on which a national board relies must be legitimized by three factors . . . scholarship in the academic discipline . . . foundations of the process of education . . . [and] intuitive credibility (or face validity) in the opinions of the professional community in whose interests they have been designed. (1987b, p. 5)

Shulman argues that we need to radically reconstruct what we do in the preparation of teachers. "Anyone who treats certification [testing] as a magic bullet that will somehow select the best and brightest teachers from an unchanging teacher education process is going to be badly disappointed" (Shulman, 1988, p. 43).

A recommended change in teacher education is the introduction of the case method similar to what is now done in law and business education. Realistic problems are addressed both from theoretical and practical perspectives. As with the simulation exercises discussed earlier (testing exercises in assessment centers), preservice teachers can engage in reflective and critical discussion of the teaching and learning process. Initial efforts with physical education case literature have shown promising signs as an instructional strategy. In a supervision course taught by one of us, vignettes written by a physical education cooperating teacher about her supervisory experiences generated substantive discussions about how preservice teachers learn to teach and how they might assist in that process. For examples of classroom case literature, see Shulman and Colbert (1987, 1988).

Observation, analysis, and critique of one's own and others' teaching during a teacher education program allow students opportunities to engage in discussions about teaching. A progression of teaching experiences include teaching one on one, peer teaching, microteaching, and large-group teaching for a short time in the schools; finally student teaching can, if correctly designed and well supervised, allow students appropriate opportunities in learning to teach. Students should be held accountable for their teaching performance and encouraged to set their own goals and develop strategies to achieve those goals. How can they be taught to use the university supervisor and the cooperating teacher as resource personnel as they develop self-management of their teaching experiences? The choice of appropriate reflective exercises tied to each teaching assignment are important.

Preservice teachers develop many artifacts as they complete their teacher education program. What if students were asked to collect a sample of their best work over the course of their program? Such artifacts might include lesson plans, units plans, reflective pieces such as an essay on the "teacher I want to be," videotape of their teaching, resumé, and so forth. Discussions with the faculty advisor, cooperating teachers, and university advisors would help the preservice teacher select the most appropriate material for his/her portfolio. An oral presentation/review of this portfolio would serve as partial requirements for teaching certification. Such a review would be a natural progression from the formative review and feedback on learning to teach in earlier physical education program assignments in learning to teach.

In physical education, as with other teacher education programs, teacher educators have a moral obligation to create experiences that correspond to appropriate images of excellent professional preparation and practice. Such practice should ensure that prospective teachers are knowledgeable in their content knowledge and should encourage critical engagement with the subject matter. Improved teacher education programming and assessment should include an emphasis on performance rather than on simple paper and pencil responses, on documentation in the field, and on experiences to develop communication skills and efficacy in the classroom.

## References

- Brandt, R. (1988). On assessment of teaching: A conversation with L. Shulman. *Educational Leadership*, 46(3), 42-46.
- Brice Heath, S. (1989, October). *New approaches to assessment in education*. Paper presented to the College of Education faculty, The Ohio State University.

- Bridges, E. (1986). *Teacher incompetence*. Philadelphia: Falmer.
- Darling-Hammond, L. (1986, Fall). Teaching knowledge: How do we test it? *American Educator*, pp. 18-21, 46-47.
- Darst, P., Zakrajsek, D., & Mancini, V. (1989). *Analyzing physical education and sport*. Champaign, IL: Human Kinetics.
- Evenden, E. (1933). *Summary and interpretations. National survey of the education of teachers* (Vol. VI) (Office of Education, Federal Security Agency, Bulletin, 1933, No. 10). Washington, DC: USGPO.
- Flippo, R. (1986). Teacher certification testing: Perspectives and issues. *Journal of Teacher Education*, 37(2), 2-9.
- Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan*, 70, 683-687.
- Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. *Review of Research in Education*, 14, 169-238.
- Kinney, L.B. (1964). *Certification in education*. Englewood Cliffs, NJ: Prentice-Hall.
- Peterson, K., & Mitchell, A. (1985). Teacher controlled evaluation in a career ladder program. *Educational Leadership*, 43, 44-47.
- Resnick, L. (1989, October). *Tests as standards of achievement in schools*. Paper presented at the 50th ETS Invitational Conference, New York.
- Shulman, J., & Colbert, J. (1987). *The mentor teacher casebook*. Eugene: ERIC Clearinghouse on Educational Management, University of Oregon.
- Shulman, J., & Colbert, J. (1988). *The intern teacher casebook*. Eugene: ERIC Clearinghouse on Educational Management, University of Oregon.
- Shulman, L. (1987a). Assessment for teaching: An initiative for the profession. *Phi Delta Kappan*, 68, 38-44.
- Shulman, L. (1987b). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Shulman, L. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership*, 46(3), 36-41.
- Tannehill, D., & O'Sullivan, M. (1990). Teacher certification testing in physical education. *Journal of Teaching in Physical Education*, 9, 227-239.
- Wilson, A. (1984). *Knowledge for teachers: The National Teacher Examinations program, 1940 to 1970*. Unpublished doctoral dissertation, University of Wisconsin. (University Microfilms International No. 84-14265)

Copyright of Journal of Teaching in Physical Education is the property of Human Kinetics Publishers, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.