



Assessing Student Achievement *in* **PHYSICAL EDUCATION** *for* Teacher Evaluation

KEVIN MERCIER
SARAH DOOLITTLE

What evaluation methods are appropriate for measuring teacher effectiveness?

WHILE MANY TEACHERS CONTINUE TO IGNORE the practice of assessing student achievement in physical education, the “trend” of assessment has failed to go away. Recent federal pressure to include student assessment data in teacher evaluation systems (e.g., Race to the Top; U.S. Department of Education, 2009) is yet another indicator that assessment of student outcomes is here to stay—for classroom teachers, and for all other teachers in school. Though there is a strong tradition of assessing teacher practice in physical education, standardized measures of student achievement in physical education are relatively new. The requirement to show data on student learning in physical education as evidence for deciding teacher quality is even more unfamiliar. This article reviews issues about using student achievement data to evaluate physical education teachers. It also presents examples of assessments that could be used to document student achievement for the purpose of teacher evaluation.

Required state tests, together with locally determined assessments, are the usual source of data on student performance for classroom teachers. For subjects without required state assessment, like physical education, states may begin to require a combination of state-approved or locally determined measures of student achievement.

The need to collect student achievement data for teacher evaluation in physical education should not be a surprise. The National Association for Sport and Physical Education (NASPE), the National Council for Accreditation of Teacher Education (NCATE), the National Board for Professional Teaching Standards (NBPTS), and the Centers for Disease Control and Prevention’s (CDC) Physical Education Curriculum Assessment Tool (PECAT) all call for regular assessment to guide instruction and to align programs with mandated standards. Furthermore, assessment has been the third leg of conventional physical education pedagogy (curriculum,

instruction, assessment) for the last decade—at least in theory, if not in practice. What has changed recently, as highlighted in Race to the Top, is the increased emphasis on school and teacher accountability. This accountability must be shown through the use of student assessment data to provide convincing evidence that the resources allocated for public schools, including teachers, are not being wasted and that they produce outcomes that citizens value.

Concerns About Connecting Student Achievement with Teacher Evaluation

While recognizing the need to ensure that teachers are effective, many educational leaders have spoken against directly connecting student test scores to teacher evaluations (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012), especially in subjects where standardized tests are not currently used (Fuhrman, 2010). Student test scores are affected by many factors, including some that are outside the teacher’s control (e.g., private tutoring, a student’s health, access to school-provided services). A complaint often voiced in regard to the use of standardized tests is that it is a snapshot of a child’s performance on a single day, decontextualized from the circumstances of the student’s life. For this reason, test scores are usually only one of a number of factors considered when making important decisions (e.g., school promotion, graduation, college admission). The American Educational Research Association (AERA) released a position statement on high-stakes testing supporting the notion that high-stakes decisions should not be based on a single test (AERA, 2000). Furthermore, high-stakes testing ignores many

Kevin Mercier (kmercier@adelphia.edu) is an assistant professor, and Sarah Doolittle (doolittle@adelphia.edu) is an associate professor, in the Department of Exercise Science, Health Studies, Physical Education, and Sport Management at Adelphi University, in Garden City, NY.

of the goals that schools and teachers set when deciding what to teach. Test scores may not be valid measures of teaching, programs, or schooling.

With federal dollars on the line, the incentive increases for teachers and administrators to cheat in order to gain or maintain needed funding. In addition, the increased focus on test scores and on the resources to improve test scores often results in a narrowing of the curriculum, by decreasing the amount of time available for arts and physical education, which are not tested (Ravitch, 2010). Physical education teachers often cite the lack of time to administer assessments, the inability to maintain a fun environment, and the lack of agreement between physical education goals and established assessments as reasons for not assessing students in physical education (Collier, 2011).

Arguments for Using Student Data to Make Educational Decisions

Though concerns about the validity and unintended consequences of using student achievement scores to measure teacher effectiveness still exist, legislators and educational researchers continue to cite the potential benefits of this type of educational reform (Hargreaves & Fullan, 2012). The standards-based reform movement began with *A Nation at Risk* (National Commission on Excellence in Education, 1983) and influenced the development of national content standards for all subject areas, including physical education (NASPE, 1995, 2004). These standards guide districts, administrators, and teachers in the development of appropriate curricula and instruction for outcomes that are widely valued in our society. Thus the national standards and several states' standards have served to strengthen the role of physical education as an important school subject by articulating clearly what a student should know and be able to do as a result of participating in a quality physical education program. Evaluating programs and the effectiveness of instruction are the next logical steps in a standards-based environment.

Assessments that are aligned with established learning standards and that demonstrate student achievement have been and continue to be developed by physical education scholars at the state and national levels. In theory, data from valid and reliable assessments can be used for many purposes, including assessing teacher effectiveness (Zhu, Rink, Placek, Graber, Fox et al., 2011). Although the *Race to the Top* program does not specifically address physical education, it would be irresponsible to assume it and other measures will not affect those teaching physical education. If physical educators want to be viewed as qualified teachers of a relevant subject, it seems unwise to excuse them from this mandate. The change toward increased accountability for schools and teachers, as exemplified by the *No Child Left Behind* law and *Race to the Top*, strongly suggests that measuring student achievement will continue to be linked to school and teacher evaluation.

How Physical Education Can Respond

The responsibility to evaluate physical education teachers through approved measures that include student achievement lies with individual schools and districts. A concern is that policy makers will turn to the only standardized tests they know of in physical education—fitness tests—as a way of collecting data to measure student

The responsibility to evaluate physical education teachers through approved measures that include student achievement lies with individual schools and districts.

achievement. For many districts, body mass index (BMI) and fitness test scores are the only student achievement data that are routinely collected in physical education. In addition, while citing the obesity epidemic, decision makers may assume that improving fitness is the primary goal of physical education. It would be inappropriate, however, to use fitness test scores or BMI measures to determine teacher effectiveness in physical education.

For several decades NASPE and most other state education departments have defined physical education as a means for educating all students in the knowledge, skills, and values for a health-enhancing, physically active lifestyle. In fact, most physical education authorities caution against training students to perform well on fitness tests. Instead, there is consensus that physical education should promote enjoyable physical activity, help develop motor skills, and provide opportunities to engage in a wide range of physical activities, both now and in the future (CDC, 2005; NASPE, 2004). If fitness scores were to be used for teacher evaluation, decision makers should be prepared to see physical education move away from life-long physical activities toward mere physical training—similar to the physical-training programs used in the military. If, instead, the purpose of physical education is to extend beyond physical training, to higher-order cognitive goals, skill acquisition, and personal and social development for lifelong physical activity, then fitness tests are not valid assessments of student achievement. Alternative assessments should be considered.

In addition, although training students to do well on fitness tests may improve fitness test scores, it may also contribute to students developing negative feelings about participating in physical activity. Fitness testing has previously been shown to decrease positive attitudes toward physical education (Mercier & Silverman, 2012),



© iStockphoto/Neil8

and fitness testing is the most common negative memory that adults have of physical education (Hopple & Graham, 1995). With this in mind, it is important that physical educators be aware of the long-term effect that fitness testing may have on their students. It appears that students' experiences with fitness testing could have a profound effect on whether or not physical educators meet their goal of promoting lifelong physical activity.

Fitness testing should be part of a quality physical education program that includes instruction on fitness education. A concern with using fitness tests to evaluate student achievement (and by extension, teacher quality) is that they may not serve as an accurate assessment tool because students' scores can easily be affected by factors such as genetics, effort, motivation, the testing environment, and maturation (Fox & Biddle, 1988; Pangrazi & Corbin, 1993; Silverman, Keating, & Phillips, 2008). Since many factors unrelated to teaching and programs could contribute to the score on a fitness test item, placing great significance on the score is inappropriate when evaluating teacher effectiveness. The score on a fitness test is simply a snapshot of the student's health or effort on that one day. What is more important is to use fitness scores as a part of fitness education to teach students the knowledge, dispositions, and skills needed to be active throughout life.

Currently, there is a disconnect between fitness testing and fitness education. Fitness testing is too often an isolated event, the purpose of which is unclear to students. Often fitness testing merely provides students with a score and does not require students to demonstrate

knowledge of what that score measured. Quality physical education programs should give students the opportunity to learn about the aspects of health-related fitness through fitness testing, data analysis, and exercise planning. Students could demonstrate knowledge of health-related fitness, as well as an understanding of the importance of being and staying fit by developing a personal fitness program. While a student-developed product could be used to assess the knowledge gained as a result of a quality fitness-education program, fitness test scores do not allow students to demonstrate an understanding of the components of fitness. With factors such as heredity, effort, and maturation possibly contributing more to fitness test results than physical education class, and the invalid use of scores to evaluate student knowledge of health-related fitness, it seems unfair to assess teacher effectiveness based on student achievement on fitness tests.

What Could Be Used to Measure Student Achievement?

If not fitness test scores, then what might be appropriate to measure student achievement? As states and districts look for ways to improve evaluations of teacher effectiveness by adding student achievement data, a number of scholars have designed assessments that may be more suitable. National and state assessments of student achievement have also been developed, and, if used judiciously, could provide compelling evidence of program and teaching quality. Table 1 lists examples of traditional physical-education program goals together with suggested assessments at the elementary, middle, and high school levels.

A first step toward choosing appropriate assessments for teacher evaluation is for teachers and administrators to determine the most important goals that should be addressed in the district's program. While national and state physical-education standards provide global goals, each school or district program has locally established goals, as well as constraints that limit what they can actually achieve. A consensus among physical education teachers and administrators on the most important instructional goals (also called student-learning objectives or SLOs) will serve as the foundation for important decisions about which assessments to use for teacher evaluation.

The assessments listed in table 1 were chosen because they are summative or curricular-level assessments—not formative assessments that may be more familiar to teachers. Summative assessments focus on showing levels of student achievement on the unit, semester, or yearly physical education goals set at the beginning of instruction. Formative assessments are used at the lesson level to check how well students have learned the components or the necessary steps toward achieving unit or program goals.

The summative assessments in table 1 are drawn from three main sources: PE Metrics (NASPE, 2010, 2011), state level assessments, and summative assessments designed by curriculum and pedagogy scholars in the field. PE Metrics is an assessment system for elementary and secondary students that is aligned with the national standards for physical education. This assessment system meets most criteria for quality assessments when administered as designed. PE Metrics assessments for national standard 1 are performance assessments, while assessments for standards 2 through 6 are multiple-choice test items that could be used to provide evidence of cognitive achievement at each level of K-12 programs (e.g., see NASPE, 2010, pp. 131–132).

State assessment programs similarly align with specific state standards for physical education. South Carolina, New York, Ohio, and



Table 1.

Sample Summative Assessments for Student and Teacher Evaluation in Physical Education

Major Program Goals	Elementary (K-5) Assessments	Middle School (6-8) Assessments	High School (9-12) Assessments
Fitness Development (Physical fitness test with criterion or normative scoring)	<ul style="list-style-type: none"> • FITNESSGRAM (Cooper Institute, 2007) • President's Challenge (2011) 	<ul style="list-style-type: none"> • FITNESSGRAM • President's Challenge 	<ul style="list-style-type: none"> • FITNESSGRAM • President's Challenge
Knowledge About Fitness (Written test)	PE Metrics (NASPE, 2010), standards 3 and 4, multiple-choice test for grades 2 and 5	PE Metrics (NASPE, 2010), standards 3 and 4, multiple-choice test for grade 8	PE Metrics (NASPE, 2011), standards 3 and 4, multiple-choice test for high school
Personal Fitness Planning (Written project)			<ul style="list-style-type: none"> • New York State PE Profile (NYSED, 2007), NYS Standard 1B • Personal physical activity participation log or journal (Lund & Kirk, 2010)
Daily Participation in Moderate-to-Vigorous Physical Activity (Physical activity log with parental signature)	Activitygram (NASPE 2007)	Activitygram Personal physical activity log (Lund & Kirk, 2010)	<ul style="list-style-type: none"> • Activitygram • Physical activity log (Darst, Pangrazi, Sariscsany, & Brusseau, 2012) • Step count record and personal goal setting (Darst et al.)
Motor Skill Performance and/or Complex Game/Physical Activity Performance (Teacher observation with rubric)	PE Metrics (2010), standard 1 for grades K, 2, and 5 performance assessment	PE Metrics (2010), standard 1 for grade 8 performance assessment	<ul style="list-style-type: none"> • PE Metrics (2011), standard 1 for high-school performance assessment • South Carolina Assessment Program (2007), performance indicator 1 performance assessments • New York State PE Profile, standards 1A and 2
Knowledge About Motor Skills, Sports, and Physical Activities (Written test)	<ul style="list-style-type: none"> • PE Metrics (2010), standard 2, multiple-choice test for grades 2 and 5 • See also summative assessments in Schiemer (2000), Hopple (2005), and Graham, Holt-Hale, and Parker (2007) 	<ul style="list-style-type: none"> • PE Metrics (2010), standard 2, multiple-choice test for grade 8 • See also summative assessments in Mohnsen (2008) and Darst, et al. (2012) 	<ul style="list-style-type: none"> • PE Metrics (2011), standard 2, multiple-choice test for high school • See also summative assessments in Lund and Kirk (2010), Darst, et al. (2012) and Chepko and Arnold (2000)
Knowledge About Personal and Social Skills in Activity Settings (Written test or project)	PE Metrics (2010), standards 5 and 6, multiple-choice test for grades 2 and 5	PE Metrics (2010), standards 5 and 6, multiple-choice test for grades 6-12	<ul style="list-style-type: none"> • PE Metrics (2011), standards 5 and 6, multiple-choice test for grades 6-12 • New York State PE Profile, standard 2
Personal and Social Behavior in Activity Settings (each observation with rubrics)	<ul style="list-style-type: none"> • Hellison (2011) • Hichwa (1998) 	<ul style="list-style-type: none"> • Hellison (2011) • Hichwa (1998) 	<ul style="list-style-type: none"> • New York State PE Profile, standards 1A & 2 • Hellison (2011)

other states have developed assessments intended to show how well students are progressing toward the major physical education goals articulated by state education officials. Table 1 also includes examples of summative assessments from South Carolina (South Carolina Physical Education Assessment Program, 2007) and New York (New York State Education Department [NYSED], 2007). Check

your state's education department for similar standards-based assessment tools.

Summative assessments found in the physical education literature provide additional alternatives to the physical and state-level assessments. For example, Lund and Kirk (2010) developed a physical activity log with parental sign-off and reflection questions that

As physical education enters the evolving world of evidence-based teacher evaluation, it is more important than ever to attend to the practice of assessment as a part of the overall instructional process.

Fox, K. R., & Biddle, S. J. H. (1988). The use of fitness tests: Educational and psychological considerations. *Journal of Physical Education, Recreation & Dance*, 59(2), 47-53.

Fuhrman, S. (2010). Tying teacher evaluation to student achievement. *Education Week*, 29(28), 32-33.

Graham, G., Holt-Hale, S., & Parker, M. (2007). *Children moving*. New York: McGraw-Hill.

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York: Teachers College Press.

can show students' understanding of factors that are essential for regular, moderate-to-vigorous, physical activity participation (pp. 208-210). There are other assessment tools in the literature connecting physical activity in school with physical activity outside of school with families and in the community to emphasize the goal of building daily physical-activity habits.

Conversations with administrators and teachers working with assessment programs in physical education have made it clear that in order for an assessment program to be successful, the following aspects must be in place: (1) an environment with supportive colleagues and administrators; (2) resources for teacher training and to record and track data electronically; (3) a method for informing students and parents about assessments in physical education, especially if results are to affect grades or student school records; and (4) teachers, students, and administrators must be convinced of the value of making time for the assessment process.

As physical education enters the evolving world of evidence-based teacher evaluation, it is more important than ever to attend to the practice of assessment as a part of the overall instructional process. Teachers, administrators, and teacher educators can no longer assume that assessment is not appropriate for physical education. Instead, the profession may need to reignite discussion about how to design, select, and use practical and appropriate measures of student achievement to benefit students, teachers, and programs. Researchers and practitioners may need to work together with greater urgency to study the implementation of appropriate methods for evaluating student learning, as well as teacher effectiveness, both of which are fundamental to quality physical education programs.

References

- American Educational Researcher Association. (2000). *AERA position statement on high-stakes testing in Pre-K-12 education*. Retrieved December 11, 2012, from <http://www.aera.net/AboutAERA/AERARulesPolicies/AERAPolicyStatements/PositionStatementonHighStakesTesting/tabid/11083/Default.aspx>.
- Centers for Disease Control and Prevention. (2005). *National health and nutrition examination survey*. Retrieved June 6, 2012, from <http://www.cdc.gov/nchs/nhanes.htm>.
- Chepko, S., & Arnold, R. (2000). *Guidelines for physical education programs*. Boston: Allyn & Bacon.
- Collier, D. (2011). Increasing the value of physical education: The role of assessment. *Journal of Physical Education, Recreation & Dance*, 82(7), 38-41.
- Cooper Institute. (2007). *Fitnessgram/Activitygram test administration manual* (4th ed.). Champaign IL: Human Kinetics.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Darst, P. W., Pangrazi, R. P., Sariscsany, M. J., & Brusseau, T. (2012). *Dynamic physical education for secondary students*. Boston: Benjamin Cummings.
- Hellison, D. (2011). *Teaching for personal and social responsibility through physical activity*. Champaign, IL: Human Kinetics.
- Hichwa, J. (1998). *Right fielders are people too*. Champaign, IL: Human Kinetics.
- Hopple, C. (2005). *Elementary physical education teaching and assessment*. Champaign, IL: Human Kinetics.
- Hopple, C., & Graham, G. (1995). What children think, feel, and know about physical fitness testing. *Journal of Teaching in Physical Education*, 14, 408-17.
- Lund, J., & Kirk, M. (2010). *Performance-based assessment for middle and high school physical education*. Champaign, IL: Human Kinetics.
- Mercier, K., & Silverman, S. (2012). Secondary school students' attitudes toward fitness testing. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, BC.
- Mohsen, B. (2008). *Teaching middle school physical education*. Champaign, IL: Human Kinetics.
- National Association for Sport and Physical Education. (1995). *Moving into the future: National standards for physical education*. Reston, VA: Author.
- National Association for Sport and Physical Education. (2004). *Moving into the future: National standards for physical education* (2nd ed.). Reston, VA: Author.
- National Association for Sport and Physical Education. (2010). *PE Metrics: Assessing national standards 1-6 in elementary school*. Reston, VA: Author.
- National Association for Sport and Physical Education. (2011). *PE metrics: Assessing national standards 1-6 in secondary school*. Reston, VA: Author.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education*. Washington, DC: Author. Available from <http://teachertenture.procon.org/sourcefiles/nation-at-risk-tenure-april-1983.pdf>.
- New York State Education Department. (2007). *New York State physical education profile*. Retrieved December 11, 2012, from <http://www.p12.nysed.gov/ciai/pe/>.
- Pangrazi, R. P., & Corbin, C. B. (1993). Physical fitness: Questions teachers ask. *Journal of Physical Education, Recreation & Dance*, 64(7), 14-19.
- The President's Challenge. (2011). *Physical fitness test*. Retrieved June 6, 2012, from <https://www.presidentschallenge.org/challenge/physical/index.shtml>.
- Ravitch, D. (2010). *Obama's race to the top will not improve education*. Retrieved June 6, 2012, from http://www.huffingtonpost.com/diane-ravitch/obamas-race-to-the-top-wi_b_666598.html.
- Schiemer, S. (2000). *Assessment strategies for elementary physical education*. Champaign, IL: Human Kinetics.
- Silverman, S., Keating, X. D., & Phillips, S. R. (2008). A lasting impression: A pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science*, 12, 146-166.
- South Carolina Physical Education Assessment Program. (2007). *High school notebook*. Retrieved December 11, 2012, from http://www.scah-perd.org/HS_Notebook_1-3-2007.pdf.
- U.S. Department of Education. (2009). *Race to the top executive summary*. Retrieved June 6, 2012, from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- Zhu, W., Rink, J., Placek, J., Graber, K., Fox, C., Fisette, J., Dyson, B., et al. (2011). Development and calibration of an item bank for PE Metrics assessments: Standard 1. *Measurement in Physical Education and Exercise Science*, 15, 119-137.