

TEACHER EVALUATIONS

A Document Analysis of Teacher Evaluation Systems Specific to Physical Education

*Jason M. Norris, Hans van der Mars, Pamela Kulinna,
Jayoun Kwon, Audrey Amrein-Beardsley*

Abstract

Purpose: *The purpose of this document analysis study was to examine current teacher evaluation systems, understand current practices, and determine whether the instrumentation is a valid measure of teaching quality as reflected in teacher behavior and effectiveness specific to physical education (PE).* **Method:** *An interpretive document analysis study was conducted on three teacher evaluation systems with documents collected from 10 school districts from the western United States. The three teacher evaluation systems were (a) the Danielson Framework for Teaching (FFT), (b) the Teacher Advancement Program (TAP), and (c) Marzano's Teacher Evaluation Model. The National Association of Sport and Physical Education (NASPE) PE teacher evaluation instrument was used in this study as a comparative tool to determine whether the targeted knowledge, skills, and behaviors within PE settings were present within the three teacher evaluation systems.* **Findings:** *Two systems had a high percentage of the NASPE key items present (TAP 87.5%, FFT 82.5%). The Marzano model had the least number of NASPE key items present within its system (62.5%).*

Jason M. Norris is an assistant professor in kinesiology, Department of Curriculum and Pedagogy, Texas A&M International University. Hans van der Mars is a professor in physical education, Mary Lou Fulton Teachers College, Arizona State University. Pamela Kulinna is a professor in physical education, Mary Lou Fulton Teachers College, Arizona State University. Jayoun Kwon is a doctoral candidate, Mary Lou Fulton Teachers College, Arizona State University. Audrey Amrein-Beardsley is a professor in education, Mary Lou Fulton Teachers College, Arizona State University. Please send author correspondence to jason.norris@tamiu.edu

Conclusion: *The TAP and FFT tools appear to be appropriate for use in evaluating PE teachers, contingent on at least two criteria: (a) school administrators having completed proper training on the use of the evaluation tool and (b) the evaluator having adequate PE-specific content knowledge and pedagogical content knowledge.*

Teacher evaluation serves two distinct purposes: (1) personnel decisions (e.g., tenure, termination) and (2) improvement of practice (i.e., professional development; Scriven, 1981). With the current push for more accountability of teachers in the classroom (No Child Left Behind Act [NCLB], 2002; U.S. Department of Education, 2009), it has become a priority to show evidence of student outcomes and teacher effectiveness. Thus, districts have adopted various teacher evaluation systems to help them make high-stakes decisions such as job retention and termination. Furthermore, it is important that schools use high-quality teacher evaluation systems to provide teachers with the tools they need to tailor instruction, enhance practice, and advance student learning continuously (National Education Association, 2014). Therefore, the purpose of this study was to evaluate three generic teacher evaluation systems and determine their effectiveness in evaluating physical education (PE) teachers. The literature addressing the purpose and importance of high-quality teacher evaluation systems served as a conceptual framework for this document analysis (Boyd, 1989; Henneman, Milanowski, Kimball, & Odden, 2006; Koppich, 2008; Scriven, 1981).

Three commonly used teacher evaluation systems were examined in this study: (a) the Teacher Advancement Program (TAP; National Institute for Excellence in Teaching, 2013), (b) the Framework for Teaching (FFT; Danielson, 1996, 2007); and (c) the Marzano Teacher Evaluation Model (Marzano, 2003; Marzano Research Laboratory, 2013). All of these systems target important areas related to quality teaching and emphasize student achievement.

Teacher Evaluation

Darling-Hammond (2000) determined that quality teachers have a great effect on student learning within the classroom. However, determining quality teaching and identifying teacher pedagogical skills may be a difficult endeavor based upon the extreme variability of student aptitude and prior instruction (Zimmerman

& Deckert-Pelton, 2003). Furthermore, to evaluate the quality of a teacher effectively, it is important for the evaluator to know what to measure and how to measure it accurately (Kennedy, 2008). The most common method of teacher observation and evaluation is performed by the school principal or administrator. Typically, district evaluation documents provide little guidance on what to observe across subject areas and how to analyze the outcomes and provide feedback to the teachers (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007). Historically, teacher assessment has been used to weed out underperforming teachers (Halverson, Kelley, & Kimball, 2004). Thus, current evaluation practices lack consistency in measuring teaching effectiveness, which has led to a system in which 90% of teachers are labeled as top performers, with only a small number labeled unsatisfactory (Little, 2009). Moreover, with such a high percentage of teachers deemed as top performers, teachers lacking in pedagogical content knowledge and teaching effectiveness may slip through the cracks (National Education Association, 2014).

Henneman et al. (2006) suggested that credible evaluation systems need to have five crucial elements: (a) evidence-based teaching standards, (b) valid instrumentation, (c) thorough training and recalibrating of raters, (d) multiple evaluators, and (e) established process for providing feedback and targeting support. In addition, teacher evaluation systems need to facilitate increased communication and be transparent so that teachers know what is expected of them (Henneman et al., 2006). Current evaluation systems are structured so scoring rubrics target generic teaching components, that is, teaching skills that are deemed to apply to “all” subject areas (Danielson, 1996, 2007; Little, 2009; Marzano Research Laboratory, 2011; Maricopa County Education Service Agency, 2013). Consequently, generic instrumentations may be invalid in measuring teacher quality and performance in PE, simply because dynamics differ from the general classroom to the physical activity settings.

Value-added models. Value-added models (VAMs) represent assessments used to link academic growth with a particular teacher. These models are associated with tested subjects and are used to hold teachers accountable; they use complex formulas to predict the amount of academic growth achieved by a student in a given year. The validity and reliability of scores produced by VAMs have been questioned because of the lack of consideration of multiple variables

associated with students' academic growth in and outside of the classroom (Amrein-Beardsley & Collins, 2012; Berliner, 2014; Hill, Kapitula, & Umland, 2011).

Commonly Used Teacher Evaluation Systems

TAP. TAP was created to assist schools in restructuring the evaluation and rewards system to recruit and retain talented teachers (Little, 2009). This system links accountability with compensation by focusing on the following key elements: (a) multiple career paths, (b) ongoing applied professional growth, (c) instructionally focused accountability, and (d) performance-based compensation. School districts adopt this system and after 1 year adopt monetary consequences that affect teachers (National Institute for Excellence in Teaching, 2013). Many studies have been conducted that examine TAP and its effects on the professional growth of teachers when it is implemented in schools. In addition, studies have indicated that teachers in TAP schools consistently outperform teachers in control schools across many states (Schacter & Thum, 2004; Solomon, White, Cohen, & Woo, 2007). However, Glazerman and Seifullah (2010) stated that much of the existing evidence on the effects of TAP has been found by the program's developers. Consequently, because of the convenience of self-selected samples, the possibility of program bias is great, which calls for a need of more evidence.

Danielson FFT. Danielson created her Framework for Teaching (FFT) in 1996 to assist in the improvement of teacher instruction (Danielson, 1996). The system has four domains that target different areas of teaching: (a) planning and preparation, (b) classroom environment, (c) instruction, and (d) professional responsibility. Each of these domains has 22 specific performance components, which also include 76 specific task elements that are measured using a rating scale: (a) unsatisfactory, (b) basic, (c) proficient, and (d) distinguished. Over the years, many studies have shown positive correlations with teacher evaluation scores and student achievement (Kane, Taylor, Tyler, & Wooten, 2010; Sartain, Stoelinga, & Brown, 2009). In a study spanning four states, researchers found positive correlations between FFT scores and student achievement (Henneman et al., 2006). Although many studies have validated the effectiveness of the Danielson FFT system, the majority of them are not published in peer-reviewed journals.

Marzano Teacher Evaluation Model. The Marzano Teacher Evaluation Model comprises items based upon previous works associated with teaching areas and behaviors (Marzano Research Laboratory, 2013). Marzano's model includes four domains that target different areas of teaching performance: (a) classroom strategies and behaviors, (b) preparing and planning, (c) reflecting on teaching, and (d) collegiality and professionalism. Moreover, the domains include 60 elements that target various behaviors related to teaching.

The Marzano system maps out a route for administrators and teachers to follow throughout the school year. Thus, specific meetings, self-reflections, observations, goal setting, and feedback sessions are required for administrators and teachers to stay on track. According to the Marzano system's website (<http://www.marzano-evaluation.com>), the system is research based and many studies have shown positive correlations between use of the model with teachers and student achievement (Haystead, 2010; Marzano & Haystead, 2010, 2011; Marzano Research Laboratory, 2010). Most of the studies examining the Marzano system have been conducted and published by Marzano's research group. Thus, further independent and peer-reviewed studies may be needed to determine the effectiveness of this model.

Teacher Evaluation in PE

Unless evaluators are trained and highly knowledgeable in PE content and reliability issues, bias-scoring outcomes may occur (Brandt et al., 2007). Bias scoring occurs when evaluators do not have adequate training and are unaware of the characteristics and behaviors that the evaluation is designed to measure (Oliva, Mathers, & Laine, 2009).

PE is considered a non-core subject and is not subject to standardized testing procedures (NCLB, 2001). Moreover, current trends in value-added measures of student achievement are either non-existent in PE or used based upon other school-wide measures of achievement (Prince et al., 2008).

With little to no accountability measures enforced in PE, formal teacher evaluations serve as the sole indicator of teacher quality and effectiveness (National Association for Sport and Physical Education [NASPE], 2012). Therefore, the purpose of this document analysis study was to examine three current teacher evaluation

systems, understand current practices, and determine whether the instrumentation is a valid measure of teaching quality as reflected in teacher behavior and effectiveness specific to PE.

Method

Participants and Setting

Ten school districts from the western United States were used in this study. The districts were located within the largest county in their state and consisted of rural, suburban, and urban school districts.

Data Collection

Recruitment letters requesting all documents associated with teacher evaluation systems for PE teachers were sent to the superintendent's office of 56 school districts within the western United States. To increase response rates, the researcher called each district office and spoke directly with a curriculum coordinator or representative responsible for teacher evaluation. Of the 56 school districts, 10 responded and supplied the evaluation system information. From these documents, three common evaluation systems were identified and served as the data sources for this study.

Comparative Evaluation Tool

NASPE (2007) created a teacher evaluation tool to identify the knowledge, skills, and behaviors needed by teachers to provide sound instruction in the K–12 PE classroom. A NASPE-appointed task force of experts in PE teacher education (PETE) constructed the tool. The task force constructed the tool by using the quality PE literature as a conceptual framework (NASPE, 2003, 2004). The tool is a resource for evaluating teacher behaviors and effectiveness in PE settings. The NASPE tool consists of five domains: (1) Instruction, (2) Evidence of Student Learning, (3) Management/Organization, (4) Learning Climate, and (5) Professionalism. Within each domain, multiple elements reflect different teaching behaviors. In total, 67 elements reside within the five domains.

The NASPE PE teacher evaluation instrument was used in this study to determine whether the targeted knowledge, skills, and behaviors within PE settings were present within the three teacher evaluation systems being used in the state. Domain 5 (Professionalism) was

not used in this study because not all teaching behavior targeted in this element were observable.

Data Analysis

The researchers used two rounds of document analyses to ensure an accurate evaluation of the three teacher evaluation systems.

Key items from the NASPE instrument. In the first round of analysis, the researcher and another independent reviewer examined and discussed each of the domains and elements within the NASPE evaluation tool. Based on overlapping elements and nonobservable teaching behaviors (e.g., class planning and preparation), the researchers narrowed down the list of 67 key items (elements) to 55. The researchers discussed the 55 key items from the NASPE tool and determined the items necessary to include and exclude from the final list of key items for evaluation of PE teachers during observations. After discussing and agreeing upon the lists of key items from the NASPE tool, the researchers sent this list to five experts in the field of PE for validation. These experts were chosen based upon the theory triangulation literature. Theory triangulation is the use of multiple perspectives to interpret a single set of data from experts in various status levels within a field or profession (Guion, Diehl, & McDonald, 2011). The experts consisted of two professors of PE, two doctoral candidates of PE, and one K–12 teacher with over 10 years of experience in teaching PE. The experts reviewed the list of key items and decided to agree or disagree if the items were key indicators of teacher evaluation based on observations. When an expert disagreed, the researchers requested a rationale for the decision. Once the researchers received the list with feedback from the experts, they reconvened and discussed the outcomes. There was a consistent agreement among the experts ($\geq 80\%$) on 45 items (81% from the initial decision), 36 to include and nine items not to include. After discussions, the researchers decided to exclude two items from the initial decision and to include one item as recommended by the experts. After feedback, they also decided to combine seven key items into three items based upon overlapping definitions (key items 12 and 54; 39 and 40; and 50, 51, and 52). The final list included 40 key items.

Evaluation of three teacher evaluation instruments using NASPE key items. In the second round, the researchers determined

the presence of the 40 NASPE key items in the teacher evaluation systems. First, the researchers independently determined whether each of the 40 NASPE key items were present within the teacher evaluation systems. After both researchers evaluated the instruments for the 40 key items, they held multiple debriefing sessions to discuss their decisions, rationales, and overall findings for each evaluation instrument. After multiple discussions, the researchers agreed which NASPE key items were present in each teacher evaluation system. Second, the researchers validated their findings by sending these to three of the five aforementioned experts. The experts consisted of one professor of PE, one doctoral candidate of PE, and one K–12 teacher with over 10 years of experience in teaching PE. The experts reviewed the findings and either agreed or disagreed with the researchers (e.g., see Table 1). If they agreed, the experts checked agreed, and if they disagreed, the experts checked disagree and explained in a short rationale. Once experts' data were received, the researchers discussed the agreement levels. The researchers held multiple debriefing sessions to analyze the data and finalize outcomes.

Trustworthiness of data. The researchers used theory triangulation methods in analyses in the first and second rounds of this study. Theory triangulation is the use of multiple perspectives from experts in various status levels within a field or profession to interpret a single set of data (Guion et al., 2011). The researchers used constant peer debriefing to determine credibility and dependability of the data and to establish trustworthiness. Peer debriefing sessions are “a process of exposing oneself to a disinterested peer in a manner paralleling an analytical session and for the purpose of exploring aspects of the inquiry that might otherwise remain only implicit within the inquirers mind” (Lincoln & Guba, 1985, p. 308). The researchers also used member checks to determine if experts agreed that the final items on each teacher evaluation instrument represented the key items on the NASPE instrument. Lincoln and Guba (1985) defined member checking as a technique for establishing the validity of an account (p. 298).

Results

This section displays and discusses data including (a) the percentage of common parts and parts that are not present within each

Table 1
Sample Researcher Findings for Expert Validation

NASPE			Teacher evaluation system				
Domain	Key item	Description	Domain	Item	Description	Agree/ disagree	If disagree please explain
Instruction	Lesson introduction	Teacher provides an introduction that is appropriate for the lesson	NA	NA			
Instruction	Learning expectations/ objectives/ instructional goals	Teacher states the skill or concept the students are to learn	Content	Content accessibility	Discusses plan for making content accessible.		
Instruction	Content is accurate and current	Skills and content are taught accurately	Content	Conceptual understanding	Guides all Students to create...		

Note. NA signifies NASPE key item not found within teacher evaluation system of inquiry.

teacher evaluation system in comparison to the items in the NASPE teacher evaluation tool, (b) key items from NASPE that are present in 100% of the teacher evaluation systems, and (c) NASPE key items that are missing in $\geq 75\%$ of the teacher evaluation systems.

Danielson FFT

The document analysis of the FFT indicates that seven items (17.5%) from the NASPE tool were not present within the FFT and 33 (82.5%) were present (see Table 2).

Table 2

NASPE PE Teacher Evaluation Tool Versus Danielson FFT

NASPE		Danielson	
Domain	Key items	Domain	Item
Instruction	Lesson introduction is appropriate	NA	NA
Instruction	Learning expectations/objectives/instructional goals are clearly communicated to students	Instruction	Communicating with students
		Planning and Preparation	Setting instructional outcomes
Instruction	Content is accurate and current	Planning and Preparation	Demonstrating knowledge of content and pedagogy
Instruction	Content and tasks are developmentally appropriate and properly sequenced	Planning and Preparation	Designing coherent instruction
		Instruction	Engaging students in learning
Instruction	Content and tasks are presented concisely and clearly, emphasizing key elements	Instruction	Communicating with students
Instruction	Engages students in learning by enabling all learners to participate through multiple modalities (opportunities to practice the skill)	Instruction	Engaging students in learning

Table 2 (cont.)

NASPE		Danielson	
Domain	Key items	Domain	Item
Instruction	Opportunities for teachable moments are recognized and utilized	Instruction	Demonstrating flexibility and responsiveness
Instruction	Instruction is differentiated for all learners (accommodations and modifications are made for students with disabilities or varied learning styles)	Instruction	Demonstrating flexibility and responsiveness
Instruction	Specific, meaningful, and timely feedback is provided to students (e.g., performance, efforts, and positive contributions)	Instruction	Using assessment in instruction
Instruction	Content is linked to and promotes the transfer of learning within physical education units and among other subject content areas	Planning and Preparation	Setting instructional outcomes
Instruction	Student performance is continually assessed to guide instruction	Instruction	Using assessment in instruction
Instruction	Independent learning is promoted, encouraged, and reinforced through daily assessments	Instruction	Using assessment in instruction
Instruction	Lesson pace is appropriate	Instruction	Engaging students in learning
Instruction	Appropriate closure is provided	NA	NA

Table 2 (cont.)

NASPE		Danielson	
Domain	Key items	Domain	Item
Evidence of Student Learning	Assessment is based on mastery of learning expectations that are aligned with local, state, and national standards	NA	NA
Evidence of Student Learning	There is ongoing formal and informal assessment	Instruction	Using assessment in instruction
Evidence of Student Learning	Assessment criteria is communicated to students	Instruction	Using assessment in instruction
Evidence of Student Learning	Multiple assessment strategies and tools are used (formative and summative) to monitor student learning	Instruction	Using assessment in instruction
Evidence of Student Learning	Students are able to articulate relevance and transfer of learning	NA	NA
Evidence of Student Learning	Student progress is documented in a retrievable record-keeping system	Professional responsibilities	Maintaining accurate records
Evidence of Student Learning	Student progress and achievement is communicated regularly to relevant stakeholders	Professional responsibilities	Communicating with families
Management/Organization	Lesson plans and curriculum are aligned with current local, state, and national standards	NA	NA
Management/Organization	Instructional area is safe, orderly, and supports learning activities	The classroom environment	Organizing physical space
Management/Organization	Adequate and developmentally appropriate equipment is accessible and utilized	The classroom environment	Organizing physical space

Table 2 (cont.)

NASPE		Danielson	
Domain	Key items	Domain	Item
Management/ Organization	Instructional support materials are utilized to enhance the lesson	Planning and Preparation	Demonstrating knowledge of resources
		The classroom environment	Organizing physical space
Management/ Organization	Students understand and adhere to class rules, routines, and behavioral expectations	The classroom environment	Managing student behavior
Management/ Organization	Class routines maximize instructional time	The classroom environment	Managing classroom procedures
Management/ Organization	There is a behavior management plan that is fair, firm, and equitable	The classroom environment	Managing student behavior
Management/ Organization	Appropriate behaviors are reinforced consistently	The classroom environment	Managing student behavior
Management/ Organization	Students are actively monitored and closely supervised using effective management strategies	The classroom environment	Managing student behavior
Management/ Organization	Students are appropriately grouped	The classroom environment	Managing classroom procedures
Management/ Organization	Effective and smooth transitions are apparent	The classroom environment	Managing classroom procedures
Management/ Organization	Allocated time is used effectively and efficiently, allowing students to remain focused on the lesson and task expectations	The classroom environment	Managing classroom procedures

Table 2 (cont.)

NASPE		Danielson	
Domain	Key items	Domain	Item
Management/ Organization	Students are engaged in relevant, meaningful physical activity a minimum of 50–60% of the instructional time	NA	NA
Management/ Organization	Accurate records are maintained	Professional responsibilities	Maintaining accurate records
Learning Climate	Lifelong physical activity and skillful movement are promoted	NA	NA
Learning Climate	There is a safe, secure learning environment that promotes success, appropriate risk taking, positive self-expression, and enjoyment	The classroom environment	Organizing physical space
Learning Climate	High expectations for learning and behavior are evident	The classroom environment	Establishing a culture for learning
		The classroom environment	Managing student behavior
Learning Climate	Climate of courtesy and respect is established	The classroom environment	Creating an environment of respect and rapport
Learning Climate	Students support the learning of others	The classroom environment	Creating an environment of respect and rapport

Note. NA = not present.

The evidence indicates that the FFT and NASPE tool are very common across the instructional domains. Both systems support maximum student engagement and the teacher's ability to demonstrate pedagogical content knowledge in the classroom. One area of disconnect was related to the structure of the lessons; the NASPE

tool targets an appropriate lesson introduction and closure, whereas these key items are not present within the FFT system.

Evidence of student learning was supported in the NASPE and FFT systems. Both systems support the use of ongoing formal assessments, along with the use of multiple assessment strategies to monitor student learning. An area within the Evidence of Student Learning domain that was supported in the NASPE tool but not present within the FFT was assessment of student mastery of learning expectations aligned with national, state, and local standards.

The evidence also indicates that both systems support the management and organization of a healthy learning environment. Moreover, both systems commonly use terms such as *on-task behavior*, *grouping*, and *classroom procedures*. Furthermore, one area that is targeted in the NASPE tool, which is missing from the FFT, is the amount of time students are to be engaged in meaningful physical activity.

TAP

The document analysis of the TAP system indicates that five items (12.5%) from the NASPE tool are not included in the TAP system and 35 items (87.5%) were present (see Table 3).

Table 3

NASPE PE Teacher Evaluation Tool VS. the TAP System

NASPE		TAP	
Domain	Key items	Domain	Item
Instruction	Lesson introduction is appropriate	Instruction	Lesson Structure and Pacing
Instruction	Learning expectations/objectives/instructional goals are clearly communicated to students	Instruction	Standards and Objectives
Instruction	Content is accurate and current	Instruction	Presenting Instructional Content Teacher Content Knowledge

Table 3 (cont.)

NASPE		TAP	
Domain	Key items	Domain	Item
Instruction	Content and tasks are developmentally appropriate and properly sequenced	Instruction	Presenting Instructional Content
Instruction	Content and tasks are presented concisely and clearly, emphasizing key elements	Instruction	Presenting Instructional Content
Instruction	Engages students in learning by enabling all learners to participate through multiple modalities (opportunities to practice the skill)	Instruction	Activities and Materials Teacher Knowledge of Students
Instruction	Opportunities for teachable moments are recognized and utilized	Instruction	Motivating Students
Instruction	Instruction is differentiated for all learners (accommodations and modifications are made for students with disabilities or varied learning styles)	Instruction	Teacher Knowledge of Students
Instruction	Specific, meaningful, and timely feedback is provided to students (e.g., performance, efforts, and positive contributions)	Instruction	Academic Feedback
Instruction	Content is linked to and promotes the transfer of learning within physical education units and among other subject content areas	Designing and Planning Instruction	Student Work
Instruction	Student performance is continually assessed to guide instruction	Instruction	Academic Feedback
Instruction	Independent learning is promoted, encouraged, and reinforced through daily assessments	Instruction	Academic Feedback

Table 3 (cont.)

NASPE		TAP	
Domain	Key items	Domain	Item
Instruction	Lesson pace is appropriate	Instruction	Lesson Structure and Pacing
Instruction	Appropriate closure is provided	Instruction	Lesson Structure and Pacing
Evidence of Student Learning	Assessment is based on mastery of learning expectations that are aligned with local, state, and national standards	Designing and Planning Instruction	Assessment
Evidence of Student Learning	There is ongoing formal and informal assessment	Designing and Planning Instruction	Assessment
Evidence of Student Learning	Assessment criteria is communicated to students	Designing and Planning Instruction	Assessment
Evidence of Student Learning	Multiple assessment strategies and tools are used (formative and summative) to monitor student learning	Designing and Planning Instruction	Assessment
Evidence of Student Learning	Students are able to articulate relevance and transfer of learning	Designing and Planning Instruction	Student Work
Evidence of Student Learning	Student progress is documented in a retrievable record-keeping system	NA	NA
Evidence of Student Learning	Student progress and achievement is communicated regularly to relevant stakeholders	NA	NA
Management/Organization	Lesson plans and curriculum are aligned with current local, state, and national standards	Designing and Planning Instruction	Instructional Plans
Management/Organization	Instructional area is safe, orderly, and supports learning activities	The Learning Environment	Environment

Table 3 (cont.)

NASPE		TAP	
Domain	Key items	Domain	Item
Management/ Organization	Adequate and developmentally appropriate equipment is accessible and utilized	The Learning Environment	Environment
Management/ Organization	Instructional support materials are utilized to enhance the lesson.	The Learning Environment	Environment
Management/ Organization	Students understand and adhere to class rules, routines, and behavioral expectations	The Learning Environment	Managing Student Behavior
Management/ Organization	Class routines maximize instructional time	The Learning Environment	Managing Student Behavior
Management/ Organization	There is a behavior management plan that is fair, firm, and equitable	The Learning Environment	Managing Student Behavior Respectful Culture
Management/ Organization	Appropriate behaviors are reinforced consistently	The Learning Environment	Managing Student Behavior
Management/ Organization	Students are actively monitored and closely supervised using effective management strategies	The Learning Environment	Managing Student Behavior
Management/ Organization	Students are appropriately grouped	Instruction	Grouping Students
Management/ Organization	Effective and smooth transitions are apparent	Instruction	Lesson Structure and Pacing
Management/ Organization	Allocated time is used effectively and efficiently, allowing students to remain focused on the lesson and task expectations	Instruction	Lesson Structure and Pacing Expectations

Table 3 (cont.)

NASPE		TAP	
Domain	Key items	Domain	Item
Management/ Organization	Students are engaged in relevant, meaningful physical activity a minimum of 50–60 % of the instructional time	NA	NA
Management/ Organization	Accurate records are maintained	NA	NA
Learning Climate	Lifelong physical activity and skillful movement are promoted	NA	NA
Learning Climate	There is a safe, secure learning environment that promotes success, appropriate risk taking, positive self-expression, and enjoyment	The Learning Environment	Environment
Learning Climate	High expectations for learning and behavior are evident	The Learning Environment	Expectations
Learning Climate	Climate of courtesy and respect is established	The Learning Environment	Respectful Culture
Learning Climate	Students support the learning of others	The Learning Environment	Respectful Culture

Note. NA = not present.

The evidence indicates that the NASPE tool and TAP system are very similar across all domains. Moreover, the data indicate that the NASPE tool is 100% common with the TAP instrument in the Instruction domain.

There is a disconnect between the two systems in the areas of keeping accurate records of student progression and communicating progression to relevant stakeholders. The NASPE tool emphasizes the documentation of student progress and communication to relevant stakeholders, whereas the TAP system does not. Last, the TAP system does not indicate the amount of time students are to be engaged in meaningful physical activity during class.

Marzano’s Teacher Evaluation Model

The document analysis of Marzano’s Teacher Evaluation Model indicates that 15 items (37.5%) from the NASPE tool are not included and 25 items (62.5%) are present (see Table 4).

Table 4

*NASPE PE Teacher Evaluation Tool Versus
Marzano’s Teacher Evaluation Model*

NASPE		Marzano	
Domain	Key item	Domain	Item
Instruction	Lesson introduction is appropriate	NA	NA
Instruction	Learning expectations/objectives/instructional goals are clearly communicated to students	Classroom Strategies and Behaviors	Providing Clear Learning Goals and Scales (Rubrics)
Instruction	Content is accurate and current	NA	NA
Instruction	Content and tasks are developmentally appropriate and properly sequenced	Classroom Strategies and Behaviors	Previewing New Content Chunking Content Into “Digestible Bites”
Instruction	Content and tasks are presented concisely and clearly, emphasizing key elements	Classroom Strategies and Behaviors	Chunking Content into “Digestible Bites” Practicing Skills, Strategies, and Processes
Instruction	Engages students in learning by enabling all learners to participate through multiple modalities (opportunities to practice the skill)	Classroom Strategies and Behaviors	Practicing Skills, Strategies, and Processes
Instruction	Opportunities for teachable moments are recognized and utilized	Classroom Strategies and Behaviors	Elaborating on New Information

Table 4 (cont.)

NASPE		Marzano	
Domain	Key item	Domain	Item
Instruction	Instruction is differentiated for all learners (accommodations and modifications are made for students with disabilities or varied learning styles)	NA	NA
Instruction	Specific, meaningful, and timely feedback is provided to students (e.g., performance, efforts, and positive contributions)	Classroom Strategies and Behaviors	Celebrating Success Reflecting on Learning
Instruction	Content is linked to and promotes the transfer of learning within physical education units and among other subject content areas	Classroom Strategies and Behaviors	Previewing New Content
Instruction	Student performance is continually assessed to guide instruction	NA	NA
Instruction	Independent learning is promoted, encouraged, and reinforced through daily assessments	NA	NA
Instruction	Lesson pace is appropriate	Classroom Strategies and Behaviors	Maintaining a Lively Pace
Instruction	Appropriate closure is provided	NA	NA
Evidence of Student Learning	Assessment is based on mastery of learning expectations that are aligned with local, state, and national standards	Planning and Preparing	Attention to Established Content Standards

Table 4 (cont.)

NASPE		Marzano	
Domain	Key item	Domain	Item
Evidence of Student Learning	There is ongoing formal and informal assessment	Classroom Strategies and Behaviors	Tracking Student Progress Celebrating Success
Evidence of Student Learning	Assessment criteria is communicated to students	Classroom Strategies and Behaviors	Providing Clear Learning Goals and Scales (Rubrics)
Evidence of Student Learning	Multiple assessment strategies and tools are used (formative and summative) to monitor student learning	NA	NA
Evidence of Student Learning	Students are able to articulate relevance and transfer of learning	Classroom Strategies and Behaviors	Previewing New Content
Evidence of Student Learning	Student progress is documented in a retrievable record-keeping system	NA	NA
Evidence of Student Learning	Student progress and achievement is communicated regularly to relevant stakeholders	Collegiality and Professionalism	Promoting Positive Interactions With Students and Parents
Management/Organization	Lesson plans and curriculum are aligned with current local, state, and national standards	Planning and Preparing	Attention to Established Content Standards
Management/Organization	Instructional area is safe, orderly, and supports learning activities	Classroom Strategies and Behaviors	Organizing the Physical Layout of the Classroom
Management/Organization	Adequate and developmentally appropriate equipment is accessible and utilized	NA	NA

Table 4 (cont.)

NASPE		Marzano	
Domain	Key item	Domain	Item
Management/ Organization	Instructional support materials are utilized to enhance the lesson	Planning and Preparing	Use of Available Traditional Resources
Management/ Organization	Students understand and adhere to class rules, routines, and behavioral expectations	Classroom Strategies and Behaviors	Establishing Classroom Rules and Procedures
Management/ Organization	Class routines maximize instructional time	Classroom Strategies and Behaviors	Establishing Classroom Rules and Procedures
Management/ Organization	There is a behavior management plan that is fair, firm, and equitable	Classroom Strategies and Behaviors	Establishing Classroom Rules and Procedures Applying Consequences for Lack of Adherence to Rules and Procedures
Management/ Organization	Appropriate behaviors are reinforced consistently	Classroom Strategies and Behaviors	Applying Consequences for Lack of Adherence to Rules and Procedures Acknowledging Adherence to Rules and Procedures
Management/ Organization	Students are actively monitored and closely supervised using effective management strategies	NA	NA

Table 4 (cont.)

NASPE		Marzano	
Domain	Key item	Domain	Item
Management/ Organization	Students are appropriately grouped	Classroom Strategies and Behaviors	Organizing Students to Interact With New Knowledge Organizing Students to Practice and Deepen Knowledge
Management/ Organization	Effective and smooth transitions are apparent	Classroom Strategies and Behaviors	Maintaining a Lively Pace
Management/ Organization	Allocated time is used effectively and efficiently, allowing students to remain focused on the lesson and task expectations	NA	NA
Management/ Organization	Students are engaged in relevant, meaningful physical activity a minimum of 50–60 % of the instructional time.	NA	NA
Management/ Organization	Accurate records are maintained	NA	NA
Learning Climate	Lifelong physical activity and skillful movement are promoted	Classroom Strategies and Behaviors	Organizing the Physical Layout of the Classroom
Learning Climate	There is a safe, secure learning environment that promotes success, appropriate risk taking, positive self-expression, and enjoyment	NA	NA
Learning Climate	High expectations for learning and behavior are evident	NA	NA

Table 4 (cont.)

NASPE		Marzano	
Domain	Key item	Domain	Item
Learning Climate	Climate of courtesy and respect is established	Classroom Strategies and Behaviors	Understanding Students' Interests and Backgrounds Using Verbal and Nonverbal Behaviors That Indicate Affection for Students
Learning Climate	Students support the learning of others	Classroom Strategies and Behaviors	Understanding Students' Interests and Backgrounds Using Verbal and Nonverbal Behaviors That Indicate Affection for Students

Note. NA = not present.

Both the NASPE tool and the Marzano system emphasize the use of clear expectations, objectives, and instructional goals within the Instruction domain and the Classroom Strategies and Behaviors domain. Moreover, both systems support the use of developmentally appropriate learning tasks that are accessible for all learners. Furthermore, progressions, chunking, and proper feedback during lessons are important common parts of both instruments. Although important areas of this domain share many commonalities, the data suggest that NASPE key items are not present in 42.9% of Marzano's instructional domain. Some key items from the NASPE instrument that are missing from the Marzano system are (a) appropriate introduction and closure of the lesson, (b) the use of accurate and current content, and (c) accommodations and modifications for students with disabilities.

The Evidence of Student Learning domain in the NASPE tool is similar to Marzano's Classroom Strategies and Behaviors domain. Evidence of its presence was found in 70% of the Marzano system. The Marzano system emphasizes clear learning goals and scales, along with assessments that are linked to content standards. Furthermore, one area that the NASPE tool supports that is not mentioned within the Marzano system is the use of multiple assessment strategies, along with documentation of student progress within a retrievable records system.

The NASPE Management/Organization key items are evident in 64.3% of the Marzano system. Moreover, both systems support establishing rules and classroom procedures, along with an organized classroom environment. The Marzano system does not emphasize active monitoring and supervision nor pinpoint the amount of time students are expected to be physically active during instructional time.

The NASPE Learning Climate key items are evident in 60% of the Marzano system. Thus, both systems emphasize promoting an environment of courtesy and respect. Furthermore, two key items from the NASPE tool that are absent from the Marzano system are the emphasis of (a) a safe and secure learning environment and (b) high expectations for learning and behavior.

NASPE Tool Key Items Present in 100% of the Three Systems

Evidence supports that 15 key items (37.5%) from the NASPE tool are present in 100% of the three teacher evaluation systems. Of the 15 key items in the NASPE tool present within all three systems, seven (46.7%) are from the Instruction domain, one (6.7%) is from the Evidence of Student Learning domain, five (35.7%) are from the Management/Organization domain, and two (13.3%) are from the Learning Climate domain.

NASPE Tool Key Items Missing From the Three Systems

Five key items (12.5%) were missing from $\geq 75\%$ of the three teacher evaluation systems: (a) lesson introduction is appropriate; (b) appropriate closure is provided; (c) student progress is documented in a retrievable record-keeping system; (d) students are engaged in relevant, meaningful physical activity a minimum of 50–60% of instructional time; and (e) lifelong physical activity and skillful movement are promoted.

Discussion

Teacher evaluators can use generic teacher evaluation systems to evaluate teacher effectiveness across multiple subject areas, including PE (Danielson, 1996, 2007; Marzano Research Laboratory, 2011). Two of the systems had a high percentage of the NASPE tool key items present (TAP, 87.5%, FFT, 82.5%). In contrast, the Marzano model had the least number of NASPE key items present (62.5%). Based on the variance in the presence of the NASPE tool key items, these generic teacher evaluation systems may not comprehensively address PE.

Evaluators need to be trained and highly knowledgeable in the subjects they observe. Brandt et al. (2007) discussed the importance of evaluators being trained and proficient in the subjects that they observe to avoid subjective and biased evaluations. Brandt et al. also stated that district evaluation systems typically provide little guidance about what to observe across subject areas. These statements are relevant to this document analysis study in that the language across all of the teacher evaluation systems of inquiry is not PE specific. With the need for quality physical educators in every classroom, it is important that PE teachers are properly evaluated on teacher effectiveness within their subject area. The researchers and the experts who contributed to this analysis are trained professionals within PE and were able to connect the language within the three systems' domains to the context of PE. This is not the case with most evaluators, who lack that specific training (Brandt et al., 2007).

The document analyses determined that a high percentage of the NASPE tool key items are present within the three teacher evaluation systems. However, key items missing from the three teacher evaluation systems are in areas that are pertinent to delivering quality PE lessons or that are specific to this content. Consequently, the five NASPE key items left out of a majority of the three teacher evaluation systems are key concepts that are taught by PETE programs.

Strength and Limitations

An identifiable strength of this study was the knowledge and ability of the research team to identify and connect the language from the three teacher evaluation systems to PE.

An identifiable limitation of this study was the small sample of systems. Many known teacher evaluation systems are used through-

out the United States. The three systems examined in this study were limited to one county in one state in the western United States.

Conclusion

PE as a school subject is not excused from the recent call for evidence of student learning and teacher effectiveness across all subject areas. Thus, more measures of accountability will be expected in the near future. In light of this, the call for more training of school administrators and evaluators is essential in allowing PE teachers the opportunity to receive a fair and valid evaluation.

This study can stimulate more research within the area of teacher evaluation in PE. Furthermore, it appears important to develop an instrument that could complement currently used systems or stand alone to measure teacher effectiveness specific to PE.

In conclusion, the TAP and FFT tools appear to be appropriate for evaluating PE teachers, contingent on at least two criteria: (a) school administrators having completed proper training on the use of the evaluation tool and (b) evaluators having adequate PE-specific content knowledge and pedagogical content knowledge.

References

- Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12), 1–27. <https://doi.org/10.14507/epaa.v20n12.2012>
- Berliner, D. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, 116(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17293>
- Boyd, R. T. C. (1989). Improving teacher evaluations. *Practical Assessment, Research & Evaluation*, 1(7), 1–3. Retrieved from <http://PAREonline.net/getvn.asp?v=1&n=7>
- Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district guidance to schools on teacher evaluation policies in the Midwest region* (Issues & Answers Report, REL 2007-No. 030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest.

- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development. <https://doi.org/10.1037/0735-7028.38.6.652>
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44. <https://doi.org/10.14507/epaa.v8n1.2000>
- Glazerman, S., & Seifullah, A. (2010). *An evaluation of the Teacher Advancement Program (TAP) in Chicago*. Washington, DC: Mathematica Policy Research.
- Guion, L., Diehl, D., & McDonald, D. (2011). *Triangulation: Establishing the validity of qualitative studies* (FCS6014). Retrieved from University of Florida, Institute of Food and Agricultural Sciences website: <http://edis.ifas.ufl.edu/pdffiles/FY/FY39400.pdf>
- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In A. Danzig & W. Black (Series Eds.), *Research and Theory in Educational Administration. Educational administration, policy, and reform: Research and measurement* (pp. 153–188). Charlotte, NC: Information Age Publishing.
- Haystead, M. (2010). *Report on professional development: A summary of findings for ten teachers on three critical commitments to school reform*. Englewood, CO: Marzano Research Laboratory.
- Henneman, H., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay*. Philadelphia, PA: Consortium for Policy Research in Education. <https://doi.org/10.12698/cpre.2013.rb45>
- Hill, C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48, 794–831. <https://doi.org/10.3102/0002831210387916>

- Kane, T., Taylor, E., Tyler, J., & Wooten, A. (2010). *Identifying effective classroom practices using student achievement data* (Paper Number 15803). Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w15803>
- Kennedy, M. (2008). Sorting out teacher quality. *Phi Delta Kappan*, 90(5), 59–63. <https://doi.org/10.1177/003172170809000115>
- Koppich, J. (2008). *Toward a more comprehensive model of teacher pay* (Working Paper 2008-06). Nashville, TN: National Center on Performance Incentives.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Little, O. (2009). *Teacher evaluation systems: The window for opportunity and reform*. Washington, DC: NEA Research.
- Maricopa County Education Service Agency. (2013). Rewarding excellence in instruction and leadership. Retrieved from <http://mcesa.schoolwires.net/domain/106>
- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R., & Haystead, M. (2010). *Final report: A second year evaluation study of Promethean ActivClassroom*. Retrieved from Marzano Research Laboratory website: http://www.marzanoresearch.com/documents/Continuation_Study_2010.pdf
- Marzano, R., & Haystead, M. (2011). *2010–2011 Adams 50 instructional model study*. Retrieved from Marzano Research Laboratory website: <http://www.sbsadams50.org/wp-content/uploads/2011/10/ACSD50-Instructional-ModelFinal-Validity-Report-2011.pdf>
- Marzano Research Laboratory. (2010). *Phase II – What works in Oklahoma schools?* Retrieved from http://www.marzanoresearch.com/documents/Phase_II_OK_State_Report.pdf
- Marzano Research Laboratory. (2011). *Research base and validation studies on the Marzano evaluation model*. Retrieved from http://www.marzanevaluation.com/files/Research_Base_and_Validation_Studies_Marzano_Evaluation_Model.pdf
- Marzano Research Laboratory. (2013). *Marzano Teacher Evaluation Model*. Retrieved from <http://www.marzanevaluation.com>

- National Association for Sport and Physical Education. (2003). *National standards for beginning physical education teachers* (2nd ed.). 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. (2007). *Physical education teacher evaluation tool*. Reston, VA: Author.
- National Association for Sport and Physical Education. (2012). *Shape of the nation report: Status of physical education in the USA*. Reston, VA: American Alliance for Health, Physical Education, Recreation, and Dance.
- National Education Association. (2014). New policy statement on teacher evaluation and accountability-Adopted as amended. Retrieved from <http://www.nea.org/grants/46326.htm>
- National Institute for Excellence in Teaching. (2013). Teacher Advancement Program. Retrieved from <http://www.tapsystem.org>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 § 1425 (2002).
- Oliva, M., Mathers, C., & Laine, S. (2009). Effective evaluation. *Principle Leadership*, 9(7), 16–21.
- Prince, C., Schuermann, P., Guthrie, J., Witham, P., Milanowski, A., & Thorn, C. (2008). *The other 69 percent: Fairly rewarding the performance of teachers of non-tested subjects and grades*. Washington, DC: U.S. Department of Education, Center for Educator Compensation Reform.
- Sartain, L., Stoelinga, S., & Brown, E. (2009). *Evaluation of the Excellence in Teaching pilot: Year one report*. Chicago, IL: Consortium on Chicago School Research.
- Schacter, J., & Thum, Y. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, 23, 411–430. <https://doi.org/10.1016/j.econedurev.2003.08.002>
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (p. 244). Beverly Hills, CA: Sage.

- Solomon, L., White, J., Cohen, D., & Woo, D. (2007). *The effectiveness of the Teacher Advancement Program*. Santa Monica, CA: National Institute for Excellence in Teaching.
- U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Retrieved from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Zimmerman, S., & Deckert-Pelton, M. (2003). Evaluating the evaluators: Teachers' perceptions of the principals' role in professional evaluation. *NASSP Bulletin*, 87, 28–37. <https://doi.org/10.1177/019263650308763604>