

TEACHER EVALUATIONS

Administrators' Perceptions of Physical Education Teacher Evaluation

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

Abstract

Purpose: Using a mixed methods approach, this study aimed to develop a better understanding of school administrators' perceptions of teacher evaluation systems, specific to physical education (PE). **Method:** This study used two sources of data collection: (a) a survey sent to administrators ($n = 19$) in one urban school district and (b) a formal semistructured interview with 10 administrators from the original surveyed group. **Results:** Statements from the survey indicated that administrators valued PE, but could not control policies concerning amount and quality of PE. Administrators also felt that they were not properly trained in PE teacher evaluation and needed help. Four common themes that emerged from the interview data were (1) I value PE, but I live in reality (administrators value PE, but practice within their reality); (2) good teaching is good teaching; (3) I know my limitations, and I want/need help (relative to teacher evaluation in PE); and (4) where's the training? (administrators felt there were shortcomings in their training specific to teacher evaluation). **Conclusion:** Evidence from this study suggests more training for administrators in PE teacher evaluation is needed, along with a content-specific evaluation instrument.

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. 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

People's experiences allow them to build different perspectives or mental models of how things occur in the world (Dervin, 1998; Starbuck & Milliken, 1988). Moreover, these mental models may act as perceptual filters that help to determine what people notice and how they interpret it. Sense-making theorists, such as Starbuck, Milliken, and Dervin, assume that people see what they know and that their actions are based on experience from their past. To understand individuals' perceptions further and how administrators interpreted teacher evaluation specific to physical education (PE) within the school as an organization, this study used sense-making theory as its framework (Dervin, 1998; Starbuck & Milliken, 1988).

With the absence of standardized testing in PE classrooms (National Association for Sport and Physical Education & American Heart Association, 2012), the sole measure of teacher performance often relies on the teacher evaluation process in place at the district level and on the expertise of the school-level administrator charged with conducting the evaluation. Generally, the school principal is the instructional leader and holds the role for successfully guiding the professional growth of the teachers within the school and for determining whether teachers will keep their job (Donaldson, 2009; Millman, 1981). Protheroe (2002) stated that a well-executed evaluation of teaching calls for the understanding of standards for student learning, an in-depth understanding of what good teaching looks like in all classrooms across each subject taught, and a strong ability to communicate and provide appropriate feedback. Moreover, principals must have an in-depth knowledge of each teacher's performance, skills, and areas of strength and weakness.

Although PE teachers and the subject itself remain marginalized within the school setting (Sheehy, 2011), teacher evaluation systems remain an important measure in determining effective teaching and professional growth and in promoting student learner outcomes. With school administrators as the likely school-site evaluators in PE teacher evaluation, identifying their perceptions of the overall evaluation process is important in determining administrators' value orientation toward the subject itself (Tziner, Murphy, & Cleveland, 2001).

Kersten and Israel (2005) examined K–8 principals' perceptions of general teacher evaluation and revealed an array of mixed results

toward the topic. Principals reported that even though they invested a large portion of their day implementing the district-required evaluation system, they felt that it had little effect on improving individual teacher instruction or student achievement. These findings are similar with those in another study, which determined that principals' attitudes toward teacher evaluation can affect the evaluation process in determining effective teacher practices (Tziner et al., 2001). Consequently, when administrators feel there is low motivation or incentive for teacher evaluation, it may become a cursory procedure with no appreciable effect on teaching practice. Thus, this may lead to a school culture in which teacher evaluation is not taken seriously and opportunity for instructional improvement is missing (Donaldson, 2009).

Teacher Evaluation Systems

Many teacher evaluation systems are in use across the country. With regard to PE, no individual systems cater specifically to the subject area. Consequently, PE teachers are evaluated with the same generic systems as other subject areas, three of which will be reviewed in the following section. The three commonly used teacher evaluation systems that will be reviewed are (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).

The TAP program was created to assist school administrators in restructuring the evaluation and rewards system within the school to recruit and retain teachers (Little, 2009). The 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. When school districts adopt this system, they have 1 year as an introductory period before monetary consequences are put into place by the administrators (National Institute for Excellence in Teaching, 2013).

Danielson's (1996) FFT was created to improve teacher instruction. The four domains of the system 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 across a rating scale: (a) unsatisfactory, (b) basic, (c) proficient, and (d) distinguished.

The Marzano evaluation model was designed based upon Marzano's previous works associated with teaching areas and behaviors (Marzano Research Laboratory, 2013). Marzano's model includes four domains that target different areas of teaching: (a) classroom strategies and behaviors, (b) preparing and planning, (c) reflecting on teaching, and (d) collegiality and professionalism.

Issues Related to Teacher Evaluation

With teacher evaluation policies raising fundamental questions about what effective teaching is and how it can be measured, the ability of school administrators to determine effective teaching behaviors across multiple subject areas is critical. Kimball and Milanowski (2009) indicated that the higher skilled that evaluators are in determining teacher behaviors across the many dimensions in teacher evaluation systems, the more accurate evaluation scores will be in determining teacher effectiveness. However, a major concern with teacher evaluation systems nationwide is that evaluators are failing to identify and remove low-performing teachers (Jacob & Lefgren, 2008). A recent analysis of state evaluation scores in Illinois showed that 83% of the state's districts had never given an unsatisfactory rating to a tenured teacher (Sartain, Stoelinga, & Krone, 2010).

Another critical issue within current teacher evaluation practices is the lack of confidence that teachers have in their evaluators (Norris et al., 2017). Halverson, Kelley, and Kimball (2004) pointed out that many teachers were able to identify when their evaluators lacked pedagogical content knowledge within their particular subject area. Thus, teachers felt that their evaluators were not qualified to evaluate instructional content decisions. Evaluators' lack of ability to make valid evaluations often led to very little critical feedback in written evaluation results (Halverson et al., 2004).

Value-Added Models

Value-added models (VAMs) of teacher evaluation have become a popular method for measuring the value a teacher adds to student learning (Amrein-Beardsley & Collins, 2012; Berliner, 2014).

VAMs use a complex statistical formula to determine if teachers are contributing to a student's growth from year to year. Many districts use these models to make high-stakes decisions such as promotion, tenure, pay, and termination. Rothstein et al. (2010) stated that although VAMs contribute to stronger analyses of school progress and program influence and to increased validity of evaluations, these models alone are not reliable and valid indicators of teacher effectiveness. Most recently, VAMs have been criticized for assessing teachers based upon student outcomes and not considering the countless variables outside of the classroom that contribute to student success (Amrein-Beardsley & Collins, 2012; Berliner, 2014). The use of VAMs has the potential to affect PE teachers, on the basis that PE teachers are now starting to be held accountable for school-wide standardized testing data.

No known studies to date have examined administrators' perceptions of evaluating physical educators. Therefore, the purpose of this study was to determine K-8 school administrators' perceptions of conducting formal teacher evaluation of PE teachers in today's context of high-stakes accountability approaches to teacher evaluation. The research question guiding this study was, how do administrators perceive the evaluation process specific to PE? The emphasis of this question was based on four specific foci: (a) perceptions of the value of PE, (b) perceptions and understanding of measures used to determine teacher effectiveness in PE, (c) perceptions and understanding of measures used to determine student achievement in PE, and (d) perceived ability to conduct teacher evaluation in PE.

Method

Recruitment and Procedures

With district and university research approval, an e-mail requesting participation in an online survey addressing administrators' perceptions of PE and the teacher evaluation process was sent to all principals ($N = 38$) in one urban school district in the western United States. Administrators were given a link to a consent form and an online survey. The SurveyMonkey online survey program was used to administer the survey. Of the 38 administrators who received the invitation to participate, 19 fully completed the survey. Administrators who completed the survey were invited to partici-

pate in a formal interview. Of these 19 participants, 10 agreed to be interviewed. Those to be interviewed filled out their contact information to be accessed by the research team.

Participants and Setting

Nineteen administrators, 14 female and five male, completed the survey questionnaire component of the study. Four of the 19 participants had 1 to 5 years of administrative experience, seven had 5 to 10 years, and 8 had over 10 years' experience. All but one of the participants indicated that their ethnic background was Caucasian; one indicated a biracial background. Ten of the 19 administrators who completed the survey questionnaire agreed to a postsurvey interview. Seven were female, and three were male. Table 1 presents the demographic information of the interview participants. Pseudonyms were used to conceal the identity of the interview participants.

Table 1
Administrator Interview Demographics

Administrator	Gender	Ethnicity	Experience	Taught PE
Jack	M	Caucasian	8 Years	No
Barbara	F	Caucasian	Over 10 Years	No
Audrey	F	Hispanic	9 Years	Yes
Karen	F	Caucasian	Over 10 Years	No
Fred	M	Caucasian	8 Years	No
Eric	M	Caucasian	Over 10 Years	No
Paris	F	Caucasian	Over 10 Years	No
Pamela	F	Caucasian	Over 10 Years	No
Jill	F	Caucasian	3 Years	No
Susan	F	Hispanic	2 Years	No

Note. M = male; F = female.

The district comprised 25 schools: 19 elementary schools and six middle schools. The district served 17,756 students. The district reported ethnic backgrounds of the students as 65.31% Caucasian, 15.33% Hispanic, 8.16% African American, 8.46% Asian/Pacific Islander, and 2.7% American Indian. ESL students make up 2.9% of the district population. Of all of the students in the district, 23.4%

were eligible for free or reduced lunch. The district was “moving toward” using high-stakes teacher evaluation, but had not yet fully implemented it at the time of data collection.

Instruments

Data from two sources were collected in this study: (a) a survey questionnaire developed by the researchers and (b) semistructured interviews. The survey questionnaire was used to target the entire administrator population from the participating district to learn about their perceptions of teacher evaluation specific to PE. The interviews were conducted to examine the administrators’ perceptions of teacher evaluation in more depth.

Short survey. The survey served three purposes: (a) to gather demographic information, (b) to examine administrators’ level of agreement on statements concerning PE teacher evaluation, and (c) to determine which participants were willing to be interviewed.

The participants provided answers to 12 statements (items) using the following Likert-scale responses: *strongly disagree* (1), *disagree* (2), *agree* (3), and *strongly agree* (4). All of the items were in reference to teacher evaluation of PE teachers. Following are two example items: “Teacher evaluations are a reliable measure of teacher effectiveness in physical education” and “I am skilled in accurately employing the current formal teacher evaluation tool (e.g., Marzano) when evaluating the performance of classroom teachers.”

The researcher created the items for the survey using general teacher evaluation literature as a framework (see Table 2).

Pilot study and expert assessment. A pilot study that provided the researcher with an opportunity to obtain feedback on the survey items and adjust the interview protocol was conducted prior to the formal study. Two administrators were chosen from outside the district and served as pilot study participants.

Two content experts in the field of PE teacher education also evaluated the items on the survey for content appropriateness. Minor changes were made to the instrument based on the pilot study and content expert evaluation of the short survey. Those suggestions were directed at content overlap and redundancy of statements.

Internal consistency reliability. The items on the survey were evaluated for internal consistency, and the entire sample showed adequate consistency across the items (overall Cronbach’s alpha of

.81). Cronbach's alpha determines the internal consistency or average correlation of items in a survey instrument to measure its reliability.

Semistructured Interview

The general interview guide was developed using the classroom teacher evaluation literature as a framework. The research team created interview questions based on four foci: (a) perceptions of the value of PE, (b) perceptions and understanding of measures used to determine teacher effectiveness in PE, (c) perceptions and understanding of measures used to determine student achievement in PE, and (d) perceived ability to conduct teacher evaluation in PE. The interview questions were developed to reveal administrators' perceptions of their districts' current high-stakes teacher evaluation systems in general and specific to the PE context. For example, one question related to the "classroom lesson observation" asked administrators, "Specific to your district's evaluation tool, what are key elements you look for when scoring the physical education teacher?" Two content experts in the field of PE teacher evaluation evaluated the questions for content appropriateness and did not have recommendations for changes.

Each interview was conducted by the primary investigator (PI) on the campus of each participant. Interviews lasted approximately 1 hr. They were recorded with a digital voice recorder. The PI, using Mac OS X Mountain Lion Dictation, then transcribed digital voice records.

Data Analysis

Surveys. Descriptive statistics were calculated for each question (see Table 2). In addition, for each survey statement, the percentage of respondents who scored it as *agree* or *strongly agree* was calculated.

Interviews. Interview data were analyzed using constant comparison methods (Bogdan & Biklen, 2007). Common themes were identified and coded using the Dedoose (www.dedoose.com) online qualitative analysis software program (Dedoose Version 4.5, 2013). Two research team members independently reviewed all transcripts. Team members used frequent peer debriefing sessions to determine that all of the themes had emerged from the data, to negotiate themes, and to minimize researcher bias (Lincoln & Guba, 1985). They used member checks to determine whether themes and interpretations

of participant statements were accurate (Kvale & Brinkman, 2009). They did this by sending all of the final transcripts to the participants to ensure that accurate statements had been made. The participants did not recommend changes. Finally, team members used a negative case analysis to ensure the accuracy of findings, with no negative cases emerging (Bogdan & Biklen, 2007). Negative case analysis is the method of searching and discussing the data for elements that do not support themes or patterns that emerge from the data analysis.

Results

Findings of this study are presented through statistical data from the survey and a discussion of its meaning, along with a discussion of the common themes that emerged from the qualitative data. The general themes identified from the data sources include (a) I value PE, but I live in reality; (b) good teaching is good teaching; (c) I know my limitations, and I want/need help; and (e) where's the training?

Survey Data

Table 2 presents mean and standard deviation scores for each survey questionnaire statement and the percentage of respondents who scored each statement with *agree* or *strongly agree*.

Table 2
Administrators' Perceptions of PE Teacher Evaluation

Survey statement	<i>M</i>	<i>SD</i>	%
1. Physical education is just as important to whole child development as are "core subjects" (e.g., mathematics)	3.53	0.51	100
2. Teacher evaluation is as important for physical education teachers as it is for teachers of "core subjects"	3.84	0.37	100
3. Physical education teachers should be held to the same expectations as teachers of other school subjects	3.68	0.48	100
4. Teacher evaluations are a useful tool for professional growth in physical education	3.63	0.50	100
5. Teacher evaluations are a reliable measure of teacher effectiveness in physical education	3.53	0.51	100

Table 2 (cont.)

Survey statement	<i>M</i>	<i>SD</i>	%
6. Current teacher evaluation practices impact teachers' classroom practices positively	3.58	0.51	100
7. Evidence of student growth and achievement is (or "should be"?) an important factor of teacher evaluation in physical education	3.37	0.68	84.21
8. My district's teacher evaluation system (e.g., Marzano) can accurately assess/determine the pedagogical content knowledge of physical education teachers	3.00	0.67	78.95
9. I am skilled in accurately employing the current formal teacher evaluation tool (e.g., Marzano) when evaluating the performance of classroom teachers	3.28	0.46	89.47
10. I am skilled in accurately employing the current formal teacher evaluation tool (e.g., Marzano) when evaluating the performance of physical education teachers	3.00	0.59	78.95
11. I am highly skilled in using the data collected through the district's current teacher evaluation system for use in postobservation feedback sessions with my physical educator(s)	3.11	0.81	68.42
12. Feedback sessions/postevaluation conferences are valuable to the professional growth of my physical education teacher(s)	3.63	0.50	100

Note. Likert scale ranged from 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*. $n = 19$. Percentage indicates participants that either agreed or strongly agreed with statement.

Administrators valued PE and felt it was as important as core subject areas (e.g., mathematics). In addition, administrators believed that PE teachers needed to be held to the same high expectations as teachers of other subject areas. Furthermore, administrators found that teacher evaluations in PE were a useful tool for professional growth and a reliable measure of teacher effectiveness.

Statements 8, 10, and 11 received less than an 80% response rate of either *agree* or *strongly agree* (see Figure 1 for percentages). All three of these statements focused on how the administrators perceived the evaluation system used within their respective district. Statement 8 focused on whether or not the administrator felt the current evaluation system could accurately assess or determine the pedagogical content knowledge of PE teachers. Although the response rate was high (> 75%) for administrators agreeing with this statement, 21.05% of the administrators disagreed. Statement 10 focused on whether or not the administrator felt skilled at employing the current teacher evaluation system within a PE classroom setting. There was evidence that 21.05% of the teachers either disagreed or strongly disagreed with this, meaning they did not feel skilled at employing teacher evaluation within a PE classroom.

The largest percentage of administrators either disagreeing or strongly disagreeing occurred with Statement 11 (31.58%). This statement focused on understanding how administrators perceived their ability to use data collected from the formal evaluation and give feedback to the PE teacher during the postevaluation conference. With 100% of the administrators either agreeing or strongly agreeing (see Statement 12 in Figure 1) that the feedback session was important to the professional growth of the PE teacher, Statement 11 responses reflected a slight disconnect with how administrators could effectively use the evaluation data during the feedback session.

Interview Results

Four common themes emerged from the interview data: (1) I value PE, but I live in reality (administrators value PE, but practice within their reality); (2) good teaching is good teaching; (3) I know my limitations, and I want/need help (relative to teacher evaluation in PE); and (4) where's the training?

Theme 1: I value PE, but I live in reality. Administrators in this study valued PE. Moreover, it was apparent that they understood that PE is beneficial in the education of the whole child. The administrators also described the cognitive benefits that physical activity contributes toward students' academic achievement. Karen has been an administrator for over 10 years. She stated,

I believe that physical education plays an important part in a child's overall education, it plays a role in helping us accomplish everything else that we try to do with kids. Keeping the kids active and teaching them about healthy lifestyles will benefit in the classroom. I obviously can't control the amount of time spent in PE, but I can encourage brain breaks and other types of physical activity during the school day. I think that we see positive results in reading and math and everything else that they do.

Administrators in this study also identified the effect that PE has on public health. They understood the overweight and obesity crisis that affects youth with sedentary lifestyles. Jill, who has been an administrator for 3 years, was animated when asked if PE was as important as other subjects within her school. She stated,

I do! Absolutely. Especially because our students, the kids, don't play outside as much as they should. They don't play outside as much as we used to. We know that there is a problem with childhood obesity, we see it. Here in school we see it. Kids are sitting more, playing video games more, on the computer all the time, watching TV. They spend the majority of the day inside. So absolutely, I feel that physical education is just as important as other subjects.

Administrators may have held a strong value for PE, but they also indicated that budget cuts and other factors controlled by policymakers are out of their control. Susan is relatively new to administration with only 2 years of experience. She stated,

To me, I think it is very important. One, it is a state standard, and it is part of the whole child, and the whole education process. Because of budget cuts, kids are only able to attend it once a week or once every six-day cycle, depending on the school district.

Although administrators in this study valued PE, they lacked knowledge as to what curricular models and instructional strategies were currently being used within their schools. Consequently, when asked about curriculum in PE, none of the administrators were able

to discuss the specifics of the PE curriculum at their school. Eric, who has over 10 years of experience, seemed a little uncertain about what curriculum model was being used in his school. He stated, “I know that elementary, they use the Pangrazi model. I am assuming that we use the same model in middle school, but please don’t quote me on that one.”

Audrey is a biracial female with 9 years of administrative experience. She indicated that she had prior teaching experience in PE. When asked about curriculum in her school, she stated,

Oh, um, you know, that’s an interesting question. I don’t believe that there is one. I mean, it’s not the Pangrazi, it’s not the dynamic physical education, it’s kind of a morphed version of that, they do use part of that, I would say a more morphed version of the dynamic.

Quality PE is another area in which administrators seemed to have a lack of knowledge and understanding. Moreover, when asked about the definition of quality PE, administrators seemed to relate their answer toward student participation and engagement levels. When asked about the definition of quality PE, Barbara, who has over 10 years of experience as an administrator, replied,

I don’t think that quality physical education instruction is that much different than quality math instruction. It’s about having an understanding of knowing what you want students to learn and do, and giving them lots of opportunities to practice doing that. So in PE that means giving students the opportunity to be moving and participating.

When asked about quality PE, Jack, an administrator with 8 years of experience, added, “Just like we want kids to be involved in their learning in the classrooms, it’s getting them active, it’s skill development, it’s practice and then application of that skill.”

The administrators emphasized professional development (PD) as an area that promotes growth with their teachers. Moreover, they felt that PD was important. However, they seemed to perceive PD for physical educators with an “us and them” approach and were at a loss to be of assistance to the physical educators at their school. Thus, the administrators seemed to understand PD for core subject teachers,

but had minimal input regarding how much and what kinds of PD that PE teachers received. Fred has been an administrator for 8 years, and he responded about PD with the following statement:

If it is, it's typically done at the district level. We don't provide it at the site level. Usually PE along with our other specialty areas if we're doing staff development days, we have one district staff development day in October. Typically, the PE teachers will get together and work on something and they will usually ask the principal's permission to go do this, and I always say absolutely, it's going to be more meaningful than working on our academic improvement plan. So I think anytime they can, they have to seek it out though and I know on Wednesdays, which is our early release day, PE teachers will try get together and go over things. They will try to have departmental meetings among themselves and I think a lot of it is their own initiative in what they do, which is too bad, but that is just the way it works around here.

When asked about PD, Jill, who has 3 years of experience, added,

I don't know, but I think that our district will provide things for them. Two times a year we have our staff development days, and there are no students. Teachers are in professional development all day long, and oftentimes PE teachers will go and do their own professional development and not be here at our site. So I know that they are offered professional development, but I know it's not a huge priority. I think a lot of times they have to seek it outside of the district.

Theme 2: Good teaching is good teaching. As a group, the administrators believed that the evaluation system used within their district was intended to fit all content areas. They also believed that good teaching is good teaching and, regardless of content area, teaching behaviors can be measured all the same. Jack has 8 years of experience, and he stated,

PE just like in every other grade level or any other content area, I am not going to be the content expert, but even without that depth of knowledge, it's again about the delivery of

good instruction. Whether it be reading or writing or math, whatever it may be, we kind of look for the same things.

Paris, who has over 10 years of experience, added,

I think I am pretty comfortable and as an administrative group we are pretty comfortable that we can make this apply to any classroom whether it be PE, art, music, because a lot of the elements, there is content, there is routine events happening constantly in the classroom.

According to the administrators in this study, they measured good teaching and teacher effectiveness by the engagement of the students. For instance, administrators appeared to gauge the level of student activity as a primary indicator of good teaching and teacher effectiveness. Thus, when asked about the measurement of teacher effectiveness during a formal observation, Barbara, with over 10 years of experience, stated, "It's measured by the level of engagement the kids are engaged in." Audrey, an administrator with 9 years of experience, also stated, "I would say that this is a hard one. I would say data based on the understanding of. . . . I would say the level of engagement/the level of activity versus sitting time." Karen, who has over 10 years' teaching experience, added,

I want to see that everyone is participating. So if there is an act or activity the teacher is expecting everyone to do I will actually count how many kids are doing it and divide it by the total number of kids. What I shoot for, and I think that this is a reasonable goal, is somewhere between 80% and 100%, participating 100% of the time.

In addition to engagement levels of students as a measure of good teaching and teacher effectiveness, administrators also looked at procedures such as safety and classroom management as key areas. Susan, who has 2 years of experience, stated,

Besides student engagement, another thing I would say would be is the classroom organized, is it safe, are there sticks flying across the room, is a lesson going smooth, are all the kids being monitored, are they getting feedback, are the kids moving, is there talking going on, are the kids off task?

When asked about measures of good teaching and teacher effectiveness, Paris, who has over 10 years of experience, added, “Classroom management in PE. It would look like this to me: Is it safe, are students being held accountable, it is optional whether or not the students are participating?”

Although administrators believed that the evaluation system used within their district was intended to fit all content areas, they also indicated that it lacked subject matter specificity, which makes it difficult for someone without a PE background to assess PE teachers fairly. Barbara (over 10 years of experience) stated,

I don't have the knowledge or skill to really and truly evaluate my physical education teacher in the way that someone with a PE background could do. I don't try to cover that up either. I am pretty honest with my special area folks. If I don't know, I ask them to tell me what I am looking for.

Audrey (9 years of experience) added,

I feel like this is our model, and now how are we going to make teachers try to fit into this model? I have trouble with that anyway, but, for example, let's just say I am going to choir, how are they going to provide an answer to question six about test and hypothesis? How about cognitive complex and provide what those two evidences are? You know, I sometimes feel like we're trying to fit a square peg into a round hole.

Theme 3: I know my limitations, and I want/need help.

Administrators in this study acknowledged their limitations relative to teacher evaluation specific to PE. They pointed out that their pedagogical content knowledge may be lacking in the field of PE and that they might need more training specific to the field and outside sources of expertise. Karen, with over 10 years of experience, stated,

If I were a PE teacher, I would want more specific feedback in my area, but realistically I don't know how possible that is without bringing in an expert in art, or PE, or music to evaluate them. I think it would be very beneficial if at a future principals meeting the district would have a physical

education teacher come out and talk to us about and explain to us exactly what we should be seeing at the different grade levels and this is also the design of the format of our lessons so that it would strengthen my knowledge of [PE], because I kind of feel like it's rudimentary right now, I know that there is a scope and sequence.

Barbara (over 10 years of experience) stated,

I would never want to give up the evaluation of the physical education teacher, but I wonder if there wasn't someone that came and looked at it really from that physical education standpoint about the quality of what's happening for kids as far as instruction, if there wasn't someone that could help do that.

As a group, the administrators stated the importance of the feedback sessions with teachers after administrator observations. As a group, the administrators stated the importance of the feedback sessions with teachers after administrator observations; they felt it was important for teachers to reflect on their own teaching. Moreover, administrators counted on feedback sessions to assist in the PD of the PE teacher. These sessions were often teacher led with the administrator relying on the expertise of the teacher in the discussion of lesson results. Eric (over 10 years of experience) stated,

Why did you pull this group aside and have them do that, you know what I mean? So there is a lot for me, you know, my postevaluation conferences with my specialty areas are those types of questions because I want to get their level of thinking, because they have a level of thinking and expertise about their subject area that I just don't have because I haven't had those experiences. My classroom teachers, a lot of times I know why they did this or why they did that. I don't have to ask. I know why you pulled that kiddo, and I know why you did this. I don't particularly know that for PE and music. I'm like, oh really; their feedback conferences are really about giving them the opportunity to talk about their rationale and chain of thinking and those types of things.

Jack (8 years of experience) stated,

It is really hard for us to dive deep into each subject area. Sometimes we hear that as a criticism, like hey, you don't know my content as well as I do. I shouldn't. I shouldn't know your content as well as you do and if I do, um, you are the grade level content expert, the content expert in your subject area.

Theme 4: Where's the training? As a group, administrators identified a weakness in the area of teacher evaluation training. Thus, administrators noticed within their administrative certification process a shortcoming in regard to how to conduct effective teacher evaluation. Furthermore, the administrators within this study felt that they resorted to "on-the-job training" upon entering into their first leadership position. Paris, with over 10 years of experience, stated,

I would say the administrative license share program didn't do jack for me in regards to [laughter] preparing me to be an instructional leader. I would hope that they're getting better. You get out in the field, and it's kind of like a sink or swim. That's where you'll get your experience.

Karen (over 10 years of experience) stated,

Well, it paled in comparison to on-the-job training, so maybe I had one three-credit class that talked about supervision. The university administrator training program did not prepare me for everything I needed to know about supervision and evaluation.

Audrey (9 years of experience) added,

We received very little training during my master's program. And then, when ordered to become a qualified evaluator, we went to two-day training. Now this model, the Marzano, we've received extensive training. Well over 70 hours, very extensive training. So with this new model we received a lot of training, but when I went to become an administrator we didn't, we got little to none or very inadequate training.

Susan, who has been an administrator for 2 years, added,

My schooling for administration, I didn't receive any on teacher evaluation. I think that we took one class where we did research on them, but I don't remember ever covering them. It is interesting that I did not have any training during my school. You would think that you would get training on teacher evaluation. We did talk about the needs of looking at teachers and evaluating them, but not actually, like, looking at different tools and comparing them or learning how to use them.

Discussion

The good news is that evidence from this study supports that the administrators as a group valued PE and understood the benefits it had on students. This aligns with the findings from Sallis, McKenzie, Kolody, and Curtis (1996), who noted that a high percentage of school administrators believe that PE can enhance concentration, decrease discipline problems, and improve academic performance. Hence, one would think that with a high percentage of administrators who value the subject, it would be a higher priority within the curriculum. In reality, the No Child Left Behind Act (2002) has played a huge role in reducing PE in the schools, making it a low priority within the curriculum (Barosso, McCullum-Gomez, Hoelscher, Kelder, & Murray, 2005; Prince et al., 2008). Administrators understood that even though they value PE, their hands are tied because of national, state, and district policies, the constant pressure to improve students' academic achievement scores, along with increasingly prevalent use of high-stakes teacher evaluation practices (Amrein-Beardsley & Collins, 2012; Berliner, 2014; Herlihy, 2012; Hill, Kapitula, & Umland, 2011).

Administrators defined quality PE in their own words, and all but one administrator was unfamiliar about which curricular model was used in their schools. This evidence suggests a major disconnect with the reality of what these administrators see and what they are supposed see in a quality PE setting. These results are consistent with results in studies that found a high percentage of principals were unaware and far removed from the day-to-day realities in PE (Locke,

1974; Lounsbery, McKenzie, Trost, & Smith, 2011). This poses a major problem, as the delivery of quality PE is vital for enhancing physical activity opportunities for students, benefiting the overall health of students, and developing skills, attributes, and behaviors to be active for life in students (Le Masurier & Corbin, 2006; McKenzie, Marshall, Sallis, & Conway, 2000). Furthermore, quality PE programs can have positive outcomes on student achievement and overall well-being (Coe, Pivarnik, Womack, Reeves, & Malina, 2006). Thus, the Centers for Disease Control and Prevention (2009) has recognized school PE as an important and available resource for promoting physical activity and healthy behaviors among children. Finally, it is important that the principal, as the school curriculum leader, is aware of current curriculum trends within PE and thus knows that students are receiving quality PE (San Diego State University, 2007).

The administrators identified PD as being important in the professional growth of the PE teachers. However, administrators approached PD with an us-versus-them lens. Moreover, they had a strong grasp on the types and amount of PD for core subject teachers, but when asked about PD for physical educators, they were unable to provide the same information about their PE teachers. This is concerning, as PD has been linked to teacher development and to student learning (Huffman & Thomas, 2003). Wang and Ha (2008) pointed out that a major issue related to PD for PE teachers is the lack of support from the school-site principal. Furthermore, it is suggested that multiple stakeholders are to be involved with the PD of the PE teacher if it is to be effective; additionally, PD should be considered through multiple lenses and aligned with elements such as district policies and curriculum requirements (Wang & Ha, 2008).

Administrators were convinced that the Marzano teacher evaluation system used within their district was applicable to all content areas, including PE. The phrase *good teaching is good teaching* was used often during the interview process. Thus, it is assumed that the same teaching behaviors that are observable within a classroom are observable within a PE setting (e.g., gymnasium, playing field). Locke (1974) explained the complexities of the PE classroom and discussed the major differences it has from those present in the classroom. Moreover, students in PE are active and mobile, some-

times within areas spanning an acre. Furthermore, students are not confined to a desk, which makes classroom management and direct instruction more challenging. Finally, although teaching behaviors may be defined the same across all subject matters, the ecology of a PE setting and how those behaviors may look are different.

Administrators also agreed, however, that the current system lacked content matter specificity, which makes it more difficult for them to assess the PE teacher fairly. Kimball (2002) similarly reported that administrators felt comfortable giving generic forms of feedback on various teaching strategies, but lacked content knowledge, which made it difficult for them to give a content-related evaluation.

Administrators acknowledged that there is a lot of room for improvement in regard to the evaluation of PE teachers. The administrators agreed that they lack content knowledge, which thus makes it difficult for them to give valid assessments of teaching performance in PE. These outcomes are consistent with those in the literature that points out the lack of content knowledge that administrators face when evaluating teachers not only in PE, but also in all subject areas (Donaldson, 2009; Halverson et al., 2004; Kimball, 2002). Related to the lack of content knowledge that administrators felt that they possessed, all of them agreed that more training specific to PE content and instructional methods would be beneficial to conducting more effective teacher evaluations.

Feedback is an important component of the teacher evaluation process (Danielson, 1996). Administrators in this study agreed that the feedback sessions were beneficial regardless of the administrators' level of content knowledge. They pointed out that these sessions are often teacher led. Moreover, this further demonstrates the lack of pedagogical content knowledge from the administrators' standpoint. With feedback and reflection being a major component in the professional growth of teachers, the ability of the administrator to discuss instructional strategies based on specific content is vital.

Prior to becoming an administrator, administrators must go through a mandatory certification process. The current results suggest that administrators felt that during their certification program they did not receive adequate training to conduct teacher evaluations. These outcomes are consistent with those in studies that show evidence of principal candidates and existing principals being

ill-prepared and inadequately supported to act as curriculum leaders while fulfilling all of the other demands of the job (Levine, 2005; Peterson, 2001). Moreover, they agreed that they received the bulk of their PD in regard to most procedural responsibilities including teacher evaluation from on-the-job training. Donaldson (2009) discussed the internal constraints to differentiation in teacher evaluation, pointing out that “without high-quality professional development, evaluators will not evaluate accurately and the evaluation will likely have little impact on teaching or learning” (p. 9).

Strengths and Limitations of the Study

This study was one of the first to examine administrators’ perceptions of teacher evaluation specific to PE. This study may be groundbreaking and encourage future studies in the area of teacher evaluation specific to PE. Moreover, with the direction in which accountability measures for nontested subjects are headed, it is important that researchers conduct more research studies in this area.

The identifiable limitations within this study were (a) a small sample size, (b) a limited number of interviews, and (c) specificity to one school district. This may have caused a lack of generalizability to the larger population (Locke, Silverman, & Spirduso, 2010). Furthermore, having one interview inhibited the opportunity for subsequent follow-up questions, which may have reduced the chances for richer data (Patel, Doku, & Tennakoon, 2003).

Conclusion

School administrators have the responsibility of determining the effectiveness of PE teachers within public schools. Administrators understanding the importance of PE is just not enough. They need to be advocates and supporters of quality PE programs and understand what is going on within the PE curriculum in their schools. Unfortunately, in today’s world, some children are suffering from overweight and obesity, diabetes, and other diseases brought on by sedentary and unhealthy lifestyles. Having effective physical educators teaching children is vital in combating this epidemic. Moreover, administrators are key players in determining PE teacher effectiveness; they are evaluating and making high-stakes decisions. Thus, it is vital that they receive the proper training and education on effec-

tive PE classroom practices, new trends in PE curricula, and the ecology in a PE setting, to determine these high-stakes decisions. Finally, administrators need to be given the proper instrumentation to give a valid evaluation to PE teachers. Although generic systems may suffice in the eyes of policymakers and stakeholders for now, future instrumentation must reflect what PE teachers and evaluators consider being the essential elements of instruction in a PE setting and must yield reliable results.

This study is one of the first to examine administrators' perceptions of teacher evaluation specific to PE. Evidence from this study suggests that more training for administrators in PE teacher evaluation, along with a content-specific evaluation instrument, is greatly needed. More research is necessary in the area of PE teacher evaluation, as the call for evidence-based outcomes in teacher effectiveness and student achievement is a central part of educational systems.

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>
- Barosso, C., McCullum-Gomez, C., Hoelscher, D., Kelder, S., & Murray, N. (2005). Self-reported barriers to quality physical education by physical education specialist in Texas. *Journal of School Health*, 75, 313–319. <https://doi.org/10.1111/j.1746-1561.2005.tb07348.x>
- 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>
- Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods*. Boston, MA: Pearson Education.
- Centers for Disease Control and Prevention. (2009). *Youth physical activity: The role of schools*. Retrieved from https://www.cdc.gov/healthyschools/physicalactivity/toolkit/factsheet_pa_guidelines_schools.pdf

- Coe, D., Pivarnik, J., Womack, C., Reeves, M., & Malina, R. (2006). Effect of physical education and activity level on academic achievements in children. *Medicine and Science in Sports and Exercise*, 38, 1515–1519. <https://doi.org/10.1249/01.mss.0000227537.13175.1b>
- 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>
- Dervin, B. (1998). Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36–46. <https://doi.org/10.1108/13673279810249369>
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Retrieved from Center for American Progress website: http://www.americanprogress.org/issues/2009/06/pdf/teacher_evaluation.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.
- Herlihy, C. (2012, October). *State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems*. Paper presented at the High-Stakes Teacher Evaluation: High Cost — Big Losses conference, Tucson, AZ.
- 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>
- Huffman, D., & Thomas, K. (2003). Relationship between professional development, teachers' instructional practices, and the achievement of students in science and mathematics. *School Science and Mathematics*, 103, 378–387. <https://doi.org/10.1111/j.1949-8594.2003.tb18123.x>

- Jacob, B., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26, 101–136. <https://doi.org/10.1086/522974>
- Kersten, T., & Israel, M. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing*, 36(1–2), 47–67.
- Kimball, S. (2002). Analysis of feedback, enabling conditions, and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16, 241–268. <https://doi.org/10.1023/A:1021787806189>
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making with a standards-based evaluation system. *Education Administration Quarterly*, 45(1), 34–70. <https://doi.org/10.1177/0013161X08327549>
- Kvale, S., & Brinkman, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage.
- Le Masurier, G., & Corbin, C. (2006). Top ten reasons for quality physical education. *Journal of Physical Education, Recreation, and Dance*, 77(6), 44–53. <https://doi.org/10.1080/07303084.2006.10597894>
- Levine, A. (2005). *Educating school leaders*. Retrieved from Education School Project website: [http://edschools.org/pdf/Fin al313.pdf](http://edschools.org/pdf/Fin%20al313.pdf)
- 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.
- Locke, L. (1974). *The ecology of the gymnasium: What the tourists never see*. Paper presented at the meeting for the Southern Association for Physical Education for College Women, Gatlinburg, TN.
- Locke, L. F., Silverman, S. J., & Spirduso, W. W. (2010). *Reading and understanding research* (3rd ed.). Thousand Oaks, CA: Sage.
- Lounsbery, M., McKenzie, T., Trost, S., & Smith, N. (2011). Facilitators and barriers to adopting evidence-based physical education in elementary schools. *Journal of Physical Activity and Health*, 8(Suppl. 1), S17–S25. <https://doi.org/10.1123/jpah.8.s1.s17>

- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano Research Laboratory. (2013). *Marzano teacher evaluation model*. Retrieved from <http://www.marzanoevaluation.com/>
- McKenzie, T., Marshall, S., Sallis, J., & Conway, T. (2000). Student activity levels, lesson context, and teacher behaviors during middle school physical education. *Research Quarterly*, 71, 249–259. <https://doi.org/10.1080/02701367.2000.10608905>
- Millman, J. (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- National Association for Sport and Physical Education & American Heart Association. (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 Institute for Excellence in Teaching. (2013). Teacher Advancement Program. Retrieved from <https://www.tapsystem.org/>
- Norris, J., van der Mars, H., Kulinna, P., Amrein-Beardsley, A., Kwon, J., & Hodges, M. (2017). Physical education teacher perceptions of teacher evaluation. *The Physical Educator*, 74, 41–62. <https://doi.org/10.18666/TPE-2017-V74-I1-6882>
- Patel, M., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment*, 9, 229–238. <https://doi.org/10.1192/apt.9.3.229>
- Peterson, K. (2001, February). *The professional development of principals: Innovations and opportunities*. Paper commissioned for the first meeting of the National Commission for the Advancement of Educational Leadership Preparations, Racine, WI.
- Protheroe, N. (2002). Improving instruction through teacher observation. *Principal*, 82(1), 48–51.
- Rothstein, R., Ladd, H., Ravitch, D., Baker, E., Barton, P., Darling-Hammond, L., . . . Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers*. Retrieved from Economic Policy Institute website: <http://www.epi.org/publication/bp278>
- Sallis, J., McKenzie, T., Kolody, B., & Curtis, P. (1996). Assessing district administrators' perceptions of elementary school physical education. *Journal of Physical Education, Recreation, and Dance*, 67(8), 25–29. <https://doi.org/10.1080/07303084.1996.10604833>

- San Diego State University. (2007). *Physical education matters* (Activity Matters for California Kids Policy Brief). Retrieved from California Endowment website: http://www.calendow.org/uploadedfiles/publications/by_topic/disparities/obesity_and_diabetes/pe%20matters%20long%20versionfinal.pdf
- Sartain, L., Stoelinga, S., & Krone, E. (2010). *Rethinking teacher evaluation: Findings from the first year of the Excellence in Teaching Project in Chicago Public Schools*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago Urban Education Institute.
- Sheehy, D. (2011). Addressing parents' perceptions in the marginalization of physical education. *Journal of Physical Education, Recreation, and Dance*, 82(7), 42–56. <https://doi.org/10.1080/07303084.2011.10598657>
- Starbuck, W., & Milliken, F. (1988). Executive perceptual filters: What they notice and how they make sense. In D. Hambrick (Ed.), *The executive effect: Concepts and methods for studying top managers* (pp. 35–65). Greenwich, CT: JAI Press.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationship between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, 9, 226–239. <https://doi.org/10.1111/1468-2389.00176>
- Wang, C., & Ha, A. (2008). The teacher development in physical education: A review of the literature. *Asian Social Science*, 4(12), 3–18.