

PEDAGOGY

Physical Education Teacher Perceptions of Teacher Evaluation

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

Abstract

Purpose: *The purpose of this study was to develop a better understanding of current PE teachers' perceptions of teacher evaluation systems. **Method:** A mixed methods approach was used and two sources of data collection were used: (a) a short survey of PE teachers (n = 22) in one urban school district and (b) a formal semistructured interview with 10 PE teachers from the same school district. **Results:** Statements from the survey showed low level of agreement in the following areas: (a) PE was a priority, (b) evidence of student growth and achievement were important factors of teacher evaluation in PE, and (c) PE teachers were confident in evaluation system and evaluators ability to determine effectiveness of PE teacher. Three common themes were revealed from the interviews: (a) PE is valued, but not prioritized; (b) teacher evaluation in PE is greatly needed, yet not transparent; and (c) physical educators are not confident in their evaluators. **Conclusion:** Results are consistent with current literature suggesting that (a) PE remains a marginalized subject with low priority; (b) teacher evaluation systems are not tailored toward noncore subjects; (c) and PE teachers, like classroom teachers, are not confident that their evaluators can give a fair and valid assessment.*

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

Swanson (1999) defined performance as the way something or someone functions. Performance can negatively or positively affect a single person, other people, or an entire organization. Schools are organizations that are committed to teacher performance, including human resource development, student performance, management, and quality improvement linked to teacher evaluation. This serves true in physical education (PE) as well. Currently, PE teacher performance and effectiveness are evaluated using the same evaluation systems used across all subject areas. Therefore, to further understand PE teacher evaluation systems, the Performance Improvement Theory (PIT) was used as a framework for this study (Swanson, 1999). An individual's growth of learning is the focal point of the PIT framework. When this occurs, people are empowered to create results and make a difference.

Teacher Evaluation Systems

The No Child Left Behind Act (2002) and more recently President Obama's Race to the Top Initiative (2009) have caused stakeholders in education to raise their interest in teacher accountability. Because of the elevated accountability and higher emphasis on showing evidence of teacher effectiveness and student achievement, school districts have adopted various performance-based and value-added models of teacher evaluation (Little, Goe, & Bell, 2009).

Performance-based teacher evaluation systems are meant to measure teacher effectiveness in the classroom by using multiple rating categories across general bodies of knowledge and skills for teaching (Shakman et al., 2012). Two commonly used performance-based teacher evaluation systems are the Marzano teacher evaluation system (Marzano, 2003) and the Danielson framework for teaching (Danielson, 1996). These systems both use a series of domains that target different areas of teaching (e.g., planning and preparation, classroom environment). Moreover, there are specific task elements within these domains that are measured across different rating scales. These systems are typically used by school administrators during formal classroom observations. Evaluation outcomes are used to rate teacher effectiveness based on proficiency levels scored by the evaluator (Little et al., 2009). However, evidence suggests that principals receive little training with these systems and that scor-

ing outcomes are based upon subjectivity (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007; Jacob & Lefgren, 2008).

Furthermore, current evaluation systems often lack alignment with subject matter curricular standards (Jerald, 2009). With so many different variations of evaluation system practices (Little et al., 2009), most district-level teacher evaluation policies provide little guidance on what criteria to follow when observing, along with how to use and share feedback from the evaluation process (Brandt et al., 2007). These inaccuracies, coupled with lack of support and insufficient training, may affect school cultures in which administrators and teachers alike struggle with the process, do not take the evaluation process seriously, or see little practical value in the evaluation process (Donaldson, 2009).

There is evidence that traditional teacher evaluation programs have been based on competing conceptions of teaching (Darling-Hammond, Wise, & Klein, 1999; Loup, Garland, Ellet, & Rugutt, 1996; Peterson, 1995). Moreover, Halverson, Kelley, and Kimball (2007) found that some teachers felt their evaluators lacked in pedagogical content knowledge and were not qualified to evaluate them on instructional content decisions. Similarly, Zimmerman and Deckert-Pelton (2003) found that teachers viewed their evaluators as going through the motions and not showing that they were putting forth much effort into the teacher evaluation. In addition, these actions gave mixed feelings as to whether the evaluation system was tailored toward them or was a generic approach (Zimmerman, 2003). Although research has shown a disconnect between teachers and their evaluators, some evidence indicates that teachers value communication and trust between themselves and the evaluator (Davis, 1988; Valentine, 1992).

Value-Added Models

One newer type of teacher evaluation system is a value-added model. Compared to traditional teacher observation instruments that assess the degree to which teachers perform to a set of standards in the classroom, value-added models use complicated statistical formulas to predict student achievement gains from year to year (Little et al., 2009). One of the most widely used value-added models is the Education Value-Added Assessment System (EVAAS) created by Statistical Analysis Systems (Amrein-Beardsley & Collins, 2012).

The authors of the EVAAS claim that this system provides valuable diagnostics about student achievement growth over time, growth that can be traced back to the teacher. Moreover, these diagnostics are used to measure teacher effectiveness in the classroom (Little et al., 2009). Rothstein et al. (2010) noted that although value-added models contribute to stronger analyses of school progress, program influences, and increased validity of evaluations, these methods alone are not reliable and valid indicators of teacher effectiveness.

Herlihy (2012) examined state and local efforts to investigate validity and reliability of scores from teacher evaluation systems. Findings indicated that few of the school personnel considered the negative unintended consequences of systems that were perceived as being arbitrary. In addition, administrators from many states reported understanding the importance of implementing better systems, but lacked the knowledge needed to measure the validity and reliability of scores produced by their systems. In a similar study, Berliner (2014) pointed out the problematic issue of many other possible influencing variables associated with value-added models such as the difficulty in assessing teachers based upon student outcomes when countless variables beyond the classroom affect achievement inside the classroom.

Teacher Evaluation in Physical Education

Similar to music and art, PE remains a marginalized subject receiving low priority and concern within school curricula (Prince et al., 2008). With the overweight and obesity epidemic plaguing the nation and school PE playing a key role in counteracting this dilemma (Pate et al., 2006), it is more important now than ever before that PE teachers receive quality evaluations and be given opportunities for professional growth and development.

There are no known studies in which researchers examine how physical educators perceive current teacher evaluation systems. Therefore, the purpose of this study was to examine current PE teachers' perceptions of teacher evaluation to understand teacher evaluation within the context of PE better. Five questions served to guide this study: (a) What are physical educators' perceptions about the value of physical education in schools? (b) What are physical educators' perceptions about quality and standards-based physical education? (c) What are physical educators' perceptions about teacher

evaluation in physical education? (d) What are physical educators' perceptions about those who evaluate physical education teachers? (e) What are physical educators' perceptions about the use of evidence of student learning in the process of teacher evaluation?

Method

Recruitment and Procedures

With university and district research approval, an e-mail was sent to all PE teachers ($N = 33$) in one urban school district in the Western United States requesting their participation in an online survey addressing PE teachers' perceptions of teacher evaluation in PE. They were sent the link to the informed consent form followed by the survey. SurveyMonkey was used in administering the survey. Of the 33 PE teachers who received the invitation to participate in the survey, 22 fully completed the survey (two surveys were eliminated as a result of the participants only completing the demographic portion of the form). Teachers who completed the survey were invited to participate in a formal interview. Of the 22 PE teachers who completed the survey, 10 agreed to participate in a formal interview. Agreement was made by participants checking yes or no at the end of the survey in reply to an interview invitation. Participants filled out their contact information to be accessed by the research team.

Participants and Setting

Twenty-two PE teachers, 11 female and 11 male, completed the survey questionnaire component of the study (Phase I). Eight of the 22 participants had 1–5 years teaching experience, three 5–10 years, and 11 over 10 years ($M = 14.55$, $SD = 7.26$). All but one of the 22 participants indicated their ethnic background as Caucasian; one indicated a biracial background. Ten of the 22 who completed the survey questionnaire agreed to be interviewed (Phase II of the study). Eight were female and two male. Nine of the 10 interviewed had 8 or more years of teaching experience; the other was a first year physical educator. Six taught PE at the elementary school level (K–5) and four taught at the middle school level (6–8). Demographic information for the Phase II participants is presented in Table 1. Pseudonyms were used to conceal the identity of the interview participants.

Table 1
Teacher Interview Demographics

Teacher	Gender	Ethnicity	Experience	Grade level
Jennifer	F	Caucasian	9 Years	K–5
Tracy	F	Caucasian	Over 10 Years	6–8
Kevin	M	Caucasian	Over 10 Years	6–8
Elizabeth ^a	F	Biracial	1 Year	K–5
Kelly	F	Caucasian	Over 10 Years	K–5
Marie	F	Caucasian	Over 10 Years	K–5
Brittany	F	Caucasian	Over 10 Years	K–5
Johnny	M	Caucasian	Over 10 Years	6–8
Sharon	F	Caucasian	8 Years	K–5
Sheila	F	Caucasian	Over 10 Years	6–8

^aThe only nontenured participant.

The district comprises 25 schools, with 19 elementary schools and six middle schools. The district serves 17,756 students, with student ethnic heritages reported 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 students in the district, 23.4% are eligible for free and reduced lunch. The district is “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) survey questionnaire developed by the researchers and (b) semistructured interviews. The survey questionnaire was used to target the entire PE teacher population from the participating district to learn about their perceptions of PE teacher evaluation. The interviews were conducted to examine the PE teachers’ perceptions of teacher evaluation in more depth.

Short survey. The survey served three specific purposes: (a) to gather demographic information about the 22 study participants, (b) to examine teachers’ level of agreement for statements concerning

evaluation of PE teachers, and (c) to determine which participants were willing to be interviewed.

The participants provided answers to 10 statements (items) using the following Likert-scale responses: *strongly disagree* (1), *disagree* (2), *agree* (3), and *strongly agree* (4). All of the items referenced teacher evaluation of PE teachers. The following are example items: “Physical education is a priority in my school curriculum” and “I am confident that my administrator/evaluator is able to determine my effectiveness as a physical educator.” Items from the survey were created by the research team using general teacher evaluation literature as a framework.

Pilot study and expert assessment. A pilot study was conducted prior to the formal study that provided an opportunity for the research team to obtain feedback on the survey items and adjust the interview protocol. Two PE teachers from outside the district were recruited to serve as pilot study participants.

Two content experts also evaluated the items on the survey for content appropriateness. Minor changes were made to the instrument based on the pilot and content expert evaluation of the short survey.

Internal consistency reliability. The items on the survey were evaluated for internal consistency across the entire sample, showing adequate consistency across the 18 items for this small sample, with an overall Cronbach’s alpha of .79. Cronbach’s alpha determines the internal consistency or average correlation of items in a survey instrument to measure its reliability.

Semistructured Interviews

The research team developed a general interview guide for the study using the classroom teacher evaluation literature as a framework (Kvale & Brinkmann, 2009). The interview protocol included demographic questions and a series of open-ended questions. The research team created interview questions based on five foci: (a) perceptions of the value of PE within the school, (b) understanding of quality and standards-based PE, (c) perceptions of current teacher evaluation practices in PE, (d) confidence of the evaluator, and (e) evidence of student learning. The interview questions were created to reveal PE teachers’ perceptions of their district’s current high-stakes teacher evaluation practices in general and those specific

to the PE context. For example, one question related to “feedback sessions,” and teachers were asked, “Do you feel that the feedback session meetings are valuable to your growth as a teacher?” Two content experts also evaluated the interview questions for content appropriateness and did not make any substantive recommendations for changes.

The primary investigator (PI) conducted each interview. Interviews lasted approximately 1 hr each and were conducted at each participant’s school location. 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. 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 4.5, 2013). Two research team members independently reviewed all transcripts. Team members used frequent peer debriefing sessions to determine that all themes had emerged from the data, to negotiate themes, and to minimize researcher bias (Lincoln & Guba, 1985). Member checks were used in determining whether themes and interpretations of participant statements were accurate (Kvale & Brinkman, 2009). This was done by sending all final transcripts to participants to ensure accurate statements had been made. No changes were recommended by the teachers. Finally, a negative case analysis was used 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

The research team examined PE teachers’ perceptions of teacher evaluation specific to PE. Findings from the 10-item survey questionnaire (see Table 2) will be presented followed by the themes that emerged from the analysis of the individual interviews. The general

themes identified from the data sources are discussed including (a) valued, but not prioritized?; (b) teacher evaluation in PE is greatly needed, yet not transparent; and (c) who do you trust? (physical educators are not confident in their evaluators).

Survey Results

Mean and standard deviation scores for each survey questionnaire statement and the percentage of respondents who scored each statement with *agree* or *strongly agree* are presented in Table 2.

Table 2

Physical Education Teachers' Perceptions of Physical Education Teacher Evaluation

Survey statements	<i>M</i>	<i>SD</i>	%
1. Physical education is a priority in my school curriculum.	2.59	0.67	50.00
2. Teacher evaluation is necessary in physical education.	3.41	0.50	100.00
3. Physical education teachers need to be held to the same expectations as teachers of other school subjects.	2.95	0.84	63.64
4. Teacher evaluations are a useful tool for professional growth in physical education.	3.18	0.59	90.91
5. Teacher evaluations are a reliable measure of teacher effectiveness in physical education.	2.71	0.85	63.64
6. Teacher evaluation does have an effect on teacher practices.	3.05	0.67	77.27
7. Evidence of student growth and achievement is an important factor of teacher evaluation in physical education.	2.71	0.85	54.55
8. I completely understand the current teacher evaluation system used within my district and know exactly what is expected of me.	2.73	0.83	59.09
9. I am confident that my administrator/evaluator is able to determine my effectiveness as a physical educator.	2.75	0.85	54.55

Table 2 (cont.)

Survey statements	<i>M</i>	<i>SD</i>	%
10. Feedback sessions/post-evaluation conferences are valuable to my growth as a physical education teacher.	3.14	0.64	86.36

Note. Likert scale ranged from 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*. $n = 22$. Percentage indicates participants that either agreed or strongly agreed with statement.

All of the PE teachers either agreed or strongly agreed that teacher evaluation in PE was necessary. They also firmly believed that teacher evaluation was important for their professional growth. Some areas of the survey that received low support from the PE teachers were Statements 1, 7, 8, and 9. Only 50% of the respondents viewed that their program was a school priority (Statement 1). A high percentage of the PE teachers did not believe that evidence of student growth and achievement were important factors of teacher evaluation in PE. Statements 8 and 9 directly reflected the PE teachers' confidence in the evaluation system and in the evaluator's ability to determine the effectiveness of the physical educator. Just fewer than 60% of the physical educators completely understood the current system being used in their district. Finally, there was evidence that PE teachers were not very confident that their evaluators can determine their effectiveness as a physical educator.

Interview Results

Three themes were identified related to PE teachers' views of the teacher evaluation process: (a) valued, but not prioritized?; (b) teacher evaluation in PE is greatly needed, yet not transparent; and (c) who do you trust? (physical educators are not confident in their evaluators).

Valued, but not prioritized. Physical educators felt that PE is valued within their schools, yet not valued enough to make it a priority. As a group, the teachers claimed that PE priority ranking in the school was based on the value that the school administrator placed on PE. Tracy explained that from her current and past experiences, it depends on how the school principal feels personally about PE.

Furthermore, if the principal shows value toward PE, the rest of the faculty will follow suit. Tracy stated,

I think if they view it [principal] as not important, then the teachers view it that way too. They'll pull kids out whenever they want, whereas we want them in our classes working hard, learning, team building, whatever we're doing. If everybody thinks they can pull the kids whenever they want, our program isn't valued. Having an administration that values that makes a big difference.

Similarly, Marie stated,

Well, I do feel like my principal feels it's very important, but, unfortunately, when our district ran into money trouble and we had to eliminate one of the days, I feel like priority-wise it did go down. They've put so much emphasis on testing and the standards and all that kind of thing now that I think, unfortunately, we are sort of like a second step.

Brittany, compared the overall value of PE in her school to that of a grading system. She felt that the value of the program was a *B*, but very low in priority. She stated,

Well, I'd say it's—if you rated them A, B, C, I'd say I'd probably give it a B as far as the overall feeling, the general—the way people value it. I think they value what we do. It's just I think that it's not as important as math and reading and those things to, I would say, most people. I feel that they believe it is not as important.

The PE teachers felt that there was an “us and them” mentality within the culture of their school. They grouped themselves with other “specialty areas.” Johnny explained his feelings:

I just think a lot of people don't have the same level of respect for physical education that I do. I think there's a lot of pressure on administration to—with the test scores and that's what a lot of parents are interested in, seeing those

higher test scores. To get that, I think they try and cram as much academics as they can, and focus groups, and ways to improve kids' level in reading and math. With that, I think there's less of an emphasis on the special areas, PE especially, but music and art would be lumped into that as well.

Greatly needed, yet not transparent. As a group, the PE teachers felt that teacher evaluation in their subject area was very important. Moreover, the teachers felt that teacher evaluation improved accountability and helped with instruction. Elizabeth stressed that accountability was an important component in the teacher evaluation process:

I think teachers should be held accountable for teaching what they're supposed to be teaching and for being good teachers and for improving their teaching. On that hand, yes, I think teacher evaluation is important for anyone who's responsible for teaching something to students—that they should be held accountable.

Similarly, Marie added,

I feel like if we are gonna be considered part of a child's overall educational experience, then we as physical educators need to be held accountable for what we teach. I think we need to have a well-rounded curriculum, just like every other subject content area. Evaluation is one way for us to be held accountable for what we teach.

The PE teachers did not understand their current evaluation system. They understood the concept, but did not know what was expected from them or how they were to prepare for the evaluation process. As a group, they felt that there was a lack of transparency and understanding of expectations. Elizabeth mentioned,

No [laughing]. It's my first year as a PE teacher and it's my first year using it. I was trained on it a little bit in new teacher orientation. I had a little bit of prior knowledge about it and then at both my sites there are teachers at the school that are

trained to be mentors for everyone to learn the system. I've gone to different sessions with them to get a little bit more help and just to understand a little bit more, but no, I think that the system is kind of exhaustive. There are 60 different elements you're supposed to know and do and know how to do the right thing you could be evaluated on. No, I don't feel like [laughing] I completely understand it.

Brittany added, "It's more applicable to classrooms. Learning how it applies to what we're doing in PE, like putting up our goals is confusing."

As a group, the teachers stated that they were uncertain as to what evidence of student learning and teacher effectiveness was expected from them during the evaluation process. Johnny stated,

I think they look at overall behavior management, which I think most of us are pretty good at. I think in PE, you have to be if you wanna have kids be safe. So many kids moving in a small environment, behavior management's important. Are the kids organized? Are they—are the activities safe? Are the kids listening? Are they following directions? I think if they're not in PE, that's gonna stick out right away with all the kids moving, 25 or 30 kids moving in a small area.

Brittany agreed, "Well, I think the key areas that they are able to focus on are the management, the behavior management, seeing how things are set up, making sure it's safe."

"Who do you trust"? PE teachers as a group did not have confidence in their evaluator. They felt that the administrators may not have the proper training or knowledge of pedagogy in a PE classroom to give a fair assessment. Sharon stated,

I am not confident. We actually just had a meeting about this with the specials, which is art and music and PE, and she was saying—my evaluator was saying, "If I had to sub for PE, I would just not be confident." I just thought, "If you're not confident teaching my subject, why are coming to evaluate me when you're not?" I'm just not confident in her ability.

Sheila added,

Well, just in talking to a few people, they've presented lessons that maybe take place out in the field with jogging, for instance. They've shared stories of the administrator's literally about 100 yards away with the iPad, 100 yards from where the students are out moving in the field and the PE teacher's moving in the field. They're rating them and I don't know how you could possibly rate someone, or hear what they're saying, or see everything that they're doing when you're 100 yards away from the action. When I think about it in a classroom that would never be, you'd never be 100 yards away from a teacher that you're evaluating.

Jennifer agreed, "I don't think they know enough about the curriculum or the standards to say, oh, you missed standard number two. You're working on this, this, and this. I don't think they really know those things."

The teachers also agreed that outside expertise and more administrators training is needed for teacher evaluation specific to PE to be effective. Kelly stated, "I really do think there needs to be training on it. What does this look like in the classroom? Well, this is what it looks like in PE." Marie also stated,

I would like to have more of an assessment with someone who knows PE. I would love that, other than my principal, who has never been teaching PE. I'd like someone who has some expertise to help me.

Johnny added,

You know, it's hard because I know their time is so—they're pulled in so many directions.

Maybe a special area orientation, maybe before the school year, maybe a one hour deal where they spend 20 minutes on, "Here's the district's art curriculum. Here's how most of our teachers in the district teach art. Here's why they do these certain things." Then another 20 minutes for music and another 20 for PE, just so they have a background on what we

do, why do it, and just a general overview of the curriculum that we're all using because it is so much different than the classroom teachers.

Discussion

PE teachers' perceptions of current teacher evaluation practices in their schools were examined. Specific to teacher evaluation, survey results and the three themes that emerged from the interviews will be discussed.

The survey results indicate that only 50% of the PE teachers felt that PE was a priority in their schools. This outcome is consistent with studies that suggest that PE is of low priority in schools (National Association for Sport and Physical Education & American Heart Association [NASPE & AHA], 2012; Prince et al., 2008; Puhse & Gerber, 2005; Sheehy, 2011). Sallis, McKenzie, Kolody, and Curtis (1996) found that administrators do support PE programs, but had multiple barriers that inhibited them from fully funding, implementing, and sustaining quality programs in their schools. Fast-forward 20 years and administrators still face multiple barriers including budget cuts and standardized testing requirements that leave administrators little ability to fully support or highly prioritize quality PE programs in schools.

A majority (54.55%) of the teachers felt that evidence of student growth was not an important factor of teacher evaluation in PE. This has been a major concern in PE due to many years of nonexistent accountability measures and, more currently, the development of the Race to the Top (RTTT) initiative in 2010. Showing evidence of student growth is a new accountability measure and may not be understood by most PE teachers (Mckenzie & Lounsbery, 2013; Rink, 2013; Ward, 2013). Rink (2013) indicated that a majority of the product measures in PE have been focused merely on motor and skill development, which happens within the dynamics of the gym, and because of this, showing evidence of student growth for evaluation purposes may be challenging for teachers.

Fewer than 60% of the PE teachers felt they understood their current evaluation system. Additionally, 54.55% of teachers did not agree that their administrators were able to determine their effectiveness as teachers. Similar outcomes were discovered in multiple stud-

ies related to classroom teachers' perceptions of the transparency of teacher evaluations and yielded similar results (Darling-Hammond et al., 1999; Halverson et al., 2004; Loup et al., 1996; Zimmerman, 2003). Teachers did not feel that they understood the evaluation process, did not trust their evaluators, and did not understand what was expected to achieve high marks during the observation. Themes that emerged from the interview data will be discussed in the following sections.

Valued, but Not Prioritized

This theme is consistent with current literature that suggests PE remains a marginalized subject area and is not given priority within the school curriculum (NASPE & AHA, 2012; Prince et al., 2008; Puhse & Gerber, 2005; Sheehy, 2011). PE teachers as a group believed that value for PE was placed on whether the school principal valued the subject. There is evidence that administrators value PE and believe that it is beneficial in enhancing concentration, decreasing discipline problems, and improving academic performance (Sallis et al., 1996). Furthermore, the problem may not lie within the realm of whether a school administrator values PE, but it may lie with the policy makers (Hardman & Marshall, 2000).

In an effort to increase classroom learning time with hopes of higher academic performance, state and district policy makers have drastically reduced the amount of PE students receive in schools (NASPE & AHA, 2012). The NASPE and AHA (2012) *Shape of the Nation Report* outlines many state-level loopholes that reduce the effectiveness of policy efforts ensuring that quality PE is present in schools. This evidence may support a top-down effect on policies alleviating a lot of the decision making or flexibility school administrators have regarding PE.

Greatly Needed, yet Not Transparent

PE teachers felt that the current teacher evaluation system was confusing and left them unsure as to what was expected of them. This is similar to what other researchers have found, in that classroom teachers felt that teacher evaluation systems were not tailored toward them (Darling-Hammond et al., 1999; Halverson et al., 2004; Loup et al., 1996; Zimmerman, 2003). Additionally, they felt that there was a lack of connection with the teacher, the evaluator, and the evalua-

tion process itself. In view of this evidence, this may be more critical within PE based on the fact that school districts use generic teacher evaluation systems that are not tailored specifically for individual subject areas. With this in mind, teacher behaviors desired based on the generic evaluation system may look or be pedagogically different within the dynamics of a PE classroom (Locke, 1974).

Physical Educators Are NOT Confident in Their Evaluators

As a group, all of the PE teachers stated that they were not confident that their evaluators could give them a fair and valid assessment. These results are consistent with other studies of classroom teachers' perceptions of their evaluators (Brandt et al., 2007; Halverson et al., 2004). Furthermore, classroom teachers felt that their principals were not adequately qualified to evaluate the subject area. These PE teachers perceived, similar to teachers in recent studies, that evaluators who lacked instructional skills did not have the ability to evaluate instructional content decisions or pedagogical content knowledge (Brandt et al., 2007; Halverson et al., 2004; Prince et al., 2008). This disconnect between teacher and evaluator points toward the increasing evidence that there are many inaccuracies, insufficient training, and lack of support for administrators expected to execute effective teacher evaluations (Darling-Hammond et al., 1999; Loup, et al., 1996; Peterson, 1995).

Limitations of the Study

Limitations of this study should be mentioned. This was a small study with a limited sample size and number of interviews. It also had participants from only one school district. Of the total participants, only one participant was not Caucasian. Also, all but one of the participants from the interviews were tenured faculty members. This affects the generalizability of the findings (Locke, Silverman, & Spirduso, 2010).

Conclusion

There are no known studies in which researchers examine the perceptions of physical educators toward teacher evaluation. This study may be the first to expose the perceptions of PE teachers toward performance-based teacher evaluation systems that may have implications at schools in the high-stakes decisions. With the cur-

rent push for teacher accountability and evidence of teacher effectiveness, teacher evaluation systems specific to PE will be beneficial for all stakeholders. This study may serve as the foundation needed to encourage the creation of effective teacher evaluation systems and future research studies on them.

As PE teachers continue to be held more accountable, and as the call for evidence-based teacher evaluation persists, more research in this area is needed. The research team examined perceptions of PE teacher evaluation through the lens of the physical educators themselves. A major disconnect was apparent among physical educators and current teacher evaluation systems. The fact that physical educators did not clearly understand what was expected of them and did not feel confident in their evaluators is concerning in today's context of high-stakes accountability approaches to teacher evaluation.

Furthermore, as PE continues to be a primary source of physical activity for youth in schools in the United States (Pate et al., 2006), a more transparent and consistent evaluation system needs to be in place to ensure valid and reliable measures of teacher effectiveness are used in PE. Also, administrators need to be educated on the ecology of a PE setting and given more training in regard to content knowledge and pedagogical content knowledge in PE. In addition, PE teachers need to be given extensive training on the evaluation systems, to develop a clear understanding of performance expectations to be met.

Implications from this study call for researchers, policy makers, and practitioners to work together to create an effective PE teacher evaluation system that will become a fundamental part of quality PE programs.

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>

- Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods*. Boston, MA: Pearson Education.
- 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.
- Darling-Hammond, L., Wise, A., & Klein, S. (1999). *A license to teach: Raising standards for teaching*. San Francisco, CA: Jossey-Bass.
- Davis, J. W. (1988). The relationship between selected performance evaluation procedures and principals' perceptions about performance evaluation. *Dissertations Abstracts International*, 50, 837A.
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Retrieved from Center for American Progress website: <https://www.americanprogress.org/issues/education/reports/2009/06/25/6243/so-long-lake-wobegon/>
- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In M. DiPaola, P. B. Forsyth (Series Eds.), W. K. Hoy, & C. G. Miskel (Vol. Eds.), *Research and theory in educational administration. Educational administration, policy, and reform: Research and measurement* (pp. 153–188). Charlotte, NC: Information Age Publishing.
- Hardman, K., & Marshall, J. (2000). The state and status of physical education in schools in international context. *European Physical Education Review*, 6, 203–229. <https://doi.org/10.1177/1356336X000063001>

- Herlihy, C. (October, 2012). *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, Arizona.
- 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>
- Jerald, C. (2009). *Aligned by design: How teacher compensation reform can support and reinforce other educational reforms*. Washington, DC: Center for American Progress.
- Kvale, S., & Brinkman, S. (2009). *InterViews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from ERIC database. (ED543776)
- Locke, L. F. (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.
- Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest districts. *Journal of Personnel Evaluation in Education*, 10, 203–226. <https://doi.org/10.1007/BF00124986>
- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mckenzie, T., & Lounsbery, M. (2013). Physical education teacher effectiveness in a public health context. *Research Quarterly for Exercise and Sport*, 84, 419–430. <https://doi.org/10.1080/02701367.2013.844025>

- 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.
- Pate, R. R., Davis, M. G., Robinson, T. N., Stone, E. J., McKenzie, T. L., & Young, J. C. (2006). Promoting physical activity in children and youth: A leadership role for schools: A scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism (Physical Activity Committee) in collaboration with the Councils on Cardiovascular Disease in the Young and Cardiovascular Nursing. *Circulation*, *114*, 1214–1224. <https://doi.org/10.1161/CIRCULATIONAHA.106.177052>
- Peterson, A. (1995). Teacher–researcher compatibility: A view from both sides. *Remedial and Special Education*, *16*, 364–367. <https://doi.org/10.1177/074193259501600606>
- 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: Center for Educator Compensation Reform, U.S. Department of Education.
- Puhse, U., & Gerber, M. (2005). *International comparison of physical education: Concepts, problems, prospects*. Oxford, United Kingdom: Meyer & Meyer Sport.
- Race to the Top. (2010). *Race to the top: Timeline and process*. Retrieved from <http://racetotop.com/about-rttt/timeline>
- Rink, J. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, *84*, 407–418. <https://doi.org/10.1080/02701367.2013.844018>
- 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>

- Shakman, K., Riordan, J., Sanchez, M., DeMeo-Cook, K., Fournier, R., & Brett, J. (2012). *An examination of performance-based teacher evaluation systems in five states* (Issues & Answers Report, REL 2012-No. 129). Retrieved from Institute of Education Sciences website: http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/rel_2012129.pdf
- Sheehy, D. (2011). Addressing parents' perceptions in the marginalization of physical education. *Journal of Physical Education, Recreation, and Dance*, 82(7), 42–44. <https://doi.org/10.1080/07303084.2011.10598657>
- Swanson, R. (1999). The foundations of performance improvement and implications for practice. In R. Torrance (Ed.), *The theory and practice of performance improvement* (pp. 1–25). <https://doi.org/10.1177/152342239900100102>
- Valentine, J. W. (1992). *Principles and practices for effective teacher evaluation*. Needham Heights, MA: Allyn and Bacon.
- Ward, P. (2013). The role of content knowledge in conceptions of teaching effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84, 431–440. <https://doi.org/10.1080/02701367.2013.844045>
- Zimmerman, S. (2003). Five steps for improving teacher evaluation: Focusing on the continual improvement of teaching and learning. *The Professional Educator*, 25(2), 43–54.
- Zimmerman, S., & Deckert-Pelton, M. (2003). Evaluating the evaluators: Teachers' perceptions of the principals' role in professional evaluation. *NASSP Bulletin*, 87(636), 28–37. <https://doi.org/10.1177/019263650308763604>