

What They Think About How They Are Evaluated: Perspectives of New York State Physical Educators on Teacher Evaluation Policy

Clancy M. Seymour, *Canisius College*
Mark J. Garrison, *D'Youville College*

Abstract

While research on high stakes testing continues to expand, little is known about how the use of student test scores to evaluate teachers is affecting physical education (PE). A proportionate, stratified random sample of physical educators in New York State was drawn ($n=489$) to survey them about their district's practices and their attitudes about the State's new teacher evaluation policy. Results indicated that 38 percent of respondents reported their district used students' written PE test results for teacher evaluation purposes, while 27 percent indicated that their district used student fitness tests for teacher evaluation purposes. Eighteen percent of respondents reported that their district used state-mandated English Language Arts (ELA) and mathematics test scores in calculating physical educators' effectiveness ranks. While few reported using performance-based measures, 94 percent of respondents indicated these as the preferred means of assessment in PE. Eighty-three percent of respondents predicted that the new teacher evaluation system would not improve PE.

Keywords Assessment in physical education; Teacher effectiveness; Value-added models

Clancy M. Seymour & Mark J. Garrison (2016). What They Think about How They're Evaluated: Perspectives of New York State Physical Educators on Teacher Evaluation Policy. *International Journal of Education Policy & Leadership* 11(6). URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/727>

IJEPL is a joint publication of PDK International, the Faculty of Education at Simon Fraser University and the College of Education and Human Development at George Mason University. By virtue of their appearance in this open access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings 90 days after initial publication. Copyright for articles published in IJEPL is retained by the authors. More information is available on the IJEPL website: <http://www.ijepl.org>



While the value of PE has been periodically questioned over its history (e.g., Berg, 2010; Claxton, 2012; Van Dalen, 1971), the current era of high stakes testing especially challenges the field in a variety of ways. In particular, the No Child Left Behind Act (NCLB) and the Race to the Top (RTTT) initiative have significant implications for K-12 PE.

By linking federal funding to schools' adequate yearly progress in reading and mathematics, NCLB has created an environment in which such classes as physical education, music, and art are viewed as nonessential and secondary to the academic mission of the school. (Trost & Van Der Mars, 2009, p. 60)

Educators have thus lamented the “diversion of resources away from physical education to ‘tested’ subjects” (Ennis, 2011, p. 16). School administrators may choose to further reduce PE offerings and schedule more time for the academic content areas that are tested (Ennis, 2011; Rink, 2013). And in response to the federal government's RTTT program, 30 states now use student test scores as a significant component of a teacher's evaluation (Rink, 2013), raising the stakes associated with student testing even higher (Baker, Oluwole, & Green, 2013). This may have implications for PE programs across the country, and given the limited time afforded physical educators to spend with their students, it might encourage instructors to demonstrate student cognitive growth via paper and pencil tests instead of psychomotor or physical growth through performance-based assessments (Seymour & Garrison, 2015).

The emergence of current teacher evaluation policy in New York State

Current teacher evaluation policy in New York State is a result of the RTTT federal initiative that intends to correct the inadequacies of NCLB by rewarding academic progress instead of establishing proficiency benchmarks (Martin, 2012). One key initiative of RTTT was to evaluate the performance of teachers and principals using value-added models (VAMs). A VAM is a proxy measure of teacher performance based on a change in student test scores over time. Corcoran (2010) explains that

[i]n theory, a teacher's value-added is the unique contribution she makes to her students' academic progress. That is, it is the portion of her students' success (or lack thereof) that cannot be attributed to any other current or past student, school, family, or community influence. (p. 4)

As a winner of RTTT federal funding, New York State has agreed to adopt teacher evaluation procedures that emulate VAMs of teacher effectiveness. Public schools thus began using the Annual Professional Performance Review (APPR) in the fall 2012.¹ In core academic subjects, standardized test results on state-mandated tests are used as a means for calculating student “growth”; these test-based growth scores account for 40 percent of evaluation points used to place teachers into one of four “effectiveness” rankings: highly effective, effective, developing, or ineffective. Classroom observations that are locally negotiated and approved by the state make up the remaining 60 percent of a teacher's evaluation. Effective teaching is, however, defined as “growth”

in student test scores, and as used in New York, this metric overrides a district's evaluation; in other words, a teacher who is rated highly effective by his or her principal may still be deemed ineffective under the APPR on the basis of her students' test scores (Garrison, 2015). In PE, where a standardized test is not used, 40 percent of the teacher effectiveness score must be calculated using student performance scores on an assessment approved by both the New York State Education Department (NYSED) and the local district (Baker, Oluwole, & Green, 2013). To meet this requirement, the physical educator must develop what are called Student Learning Objectives (SLO). A SLO is a quantitative goal that a teacher sets for his or her students to attain on a selected assessment using pre- and post-test logic. This process allows the physical educator to document student performance scores and simulate what standardized test results do in core academic subjects such as mathematics. Teacher proficiency is established if the targeted percentages of students reach the goal. Physical educators can use student results from a performance-based assessment, written test, or fitness test, any of which must be approved by both NYSED and the local district. Physical educators may also be evaluated on the basis of student performance on the State's English Language Arts (ELA) or mathematics tests (Rink, 2013). While research on VAMs and related teacher evaluation schemes is fairly extensive (Baker, Oluwole, & Green, 2013), we know of no studies examining how value-added models of teacher evaluation are affecting physical educators. Further, with 60 percent of states utilizing this type of metric for teacher effectiveness (see Rink, 2013), it is important to examine the effects of this policy more closely.

Drawn from a larger study (Seymour, 2014), this article reports the results of a representative survey of physical educator's practices and perspectives in the wake of the APPR in New York State. The research questions were:

1. What types of teacher evaluation mechanisms are school districts using in New York State to evaluate physical educators, as a result of the APPR?
2. Do physical educators believe that the APPR as enacted in their district is a sound method for evaluating teachers in their profession?

Method

Participant Selection

Following Institutional Review Board (IRB) approval, the study was conducted during the 2013–2014 public school year using an anonymous online survey distributed via email to physical educators from New York State. A proportionate, stratified random sample of physical educators was drawn, in order to survey these professionals about their district's practices and their attitudes in light of the State's new teacher evaluation policy. A list of 9,737 K-12 public school physical educators was obtained from NYSED. Strata were organized using the 11 geographic zones adopted by the professional organization of PE in the state of New York (NYS AHPERD, n.d.). A sample of 5 percent ($n=498$) yielded a maximum margin of error of 4.32 percent ($p < .05$). We manually retrieved the email addresses of nearly 50 percent (4,188) of physical educators via public school district websites and/or phone calls to schools (see Table 1). Participant background data such as race, ethnicity, gender, and age

was not collected because the research questions did not ask about variables typically associated with representation and all members of the population had an equal chance of being selected to participate in the study.

Over the course of an eight-week distribution cycle, the survey was distributed to 20 percent of teachers randomly selected from each stratum in multiple waves. The distribution cycle was divided into four two-week phases where 5 percent of randomly selected physical educators from each stratum received the survey by email. By obtaining representative thresholds for each of the 11 NYS AHPERD zones, proportionality was achieved (see Table 1).

Data Collection

After each phase, responses from each zone were calculated. Each random sample of physical educators in each zone was eliminated from future samples of that zone. A final email was sent to all randomly selected physical educators in phases 1–4 in the 5 out of the 11 zones that had not met threshold. The responses of 489 physical educators (5% of the total PE population in the state) were subsequently collected (see Table 1).

Table 1. Distribution of PE teachers and PE teacher response threshold, by NYS AHPERD zone

Zone	PE teacher totals and state percentages	Number of emails sent by phase				Response obtained (5%)
		Phase 1	Phase 2	Phase 3	Phase 4	
*Southeastern	954 (9.80)	48	48	48	69	48
*Capital	724 (7.44)	36	36	36	42	36
*Central North	693 (7.12)	35	35	35	38	35
Central South	470 (4.83)	24	24	24	24	24
Central Western	937 (9.62)	47	47	47	47	47
Western	806 (8.28)	40	40	NA	NA	40
Northern	190 (1.95)	10	10	10	10	10
*Nassau	871 (8.95)	44	44	44	160	44
*Catskill	392 (4.03)	20	20	20	34	20
*Suffolk	976 (10.02)	49	49	49	106	49
*New York City	2,724 (10.02)	136	136	2,367	NA	136
Total	9,737 (100)	489	489	2,680	530	489

*These zones were sent additional emails in phase 3 and/or 4 to obtain threshold. Zones with NA reached response threshold before completion of 4 phases of survey distribution

There were unforeseen email retrieval issues in the New York City Zone. Many schools in the New York City Department of Education (NYCDOE) do not post on school or department websites the names and email addresses of PE teachers. From the small percentage of email addresses that were accessible, it was determined that schools in the NYCDOE use the first initial and last name followed by @nyc.schools.gov as their email naming convention. With over 2,700 physical educators in this zone, alongside other subject teachers within the region, it was a challenge to correctly predict email addresses. To resolve these issues, the survey was sent to all physical educators in this zone, using the above email naming convention.

To preserve proportionality, the original threshold (136 responses or 5%) was still pursued, and once obtained, the survey was no longer distributed.

Instrumentation

A survey was developed to answer the research questions and ascertain district teacher evaluation practices and physical educator's attitudes in light of the State's new teacher evaluation policy. The survey was first piloted to a focus group of physical educators in the Western zone during the summer of 2013. The researchers adopted the item revisions recommended by the focus group of physical educators. Supplementary questions about appropriate techniques that can be used for teacher evaluation in PE were also suggested and added. Additionally, two education professionals reviewed the draft survey. The first was a teacher educator with an expertise in curriculum, and the second was a physical educator teacher educator. Further edits to language were made to simplify and better align the instrument to the research questions and objectives of the study. The survey was not intended to measure any psychological construct (for example, self-efficacy) and instead only sought to discover physical educators' practices and perspectives regarding teacher evaluation policy. Therefore, standard psychometric efforts at construct validation were not undertaken.

The survey consisted of 19 questions that were directly linked to the research questions. Survey items 2–7 and 14–16 were aligned to research question 1, while items 1, 3, and 8–12 answered research question 2. Using a Likert type scale, questions 1–12 asked respondents to indicate their level of agreement with 12 statements about the APPR and physical education.

As research methodologists have shown, there is a “well-documented tendency for respondents to be more willing to endorse strongly positive categories than strongly negative ones” (Foddy, 1994, p. 162). To compensate for this predisposition, especially given the controversial nature of the subject matter, 12 items relating to potential outcomes of the APPR were developed. The tone of these statements was carefully crafted as desirable (items 6, 7, 12), undesirable (items 3, 8, 9, 11), and neutral (items 1, 2, 4, 5, 10) to offset the perception that only a critical view of the new policy was sought and counterbalance the known propensity for respondents to be more willing to agree or strongly agree with statements irrespective of their content (Foddy, 1994). In addition, survey items were arranged to ensure that a pattern of desirable, undesirable, and neutral was not exhibited or apparent to respondents.

Finally, items 13–17 asked respondents to select, from among several constructed options, their assessment preferences, and to report their district's teacher evaluation practices (see Table 2). While protecting participant anonymity, surveys were coded to avoid duplicate submissions (questions 18 and 19). Incomplete submissions were discarded and no duplicate surveys were found.

Results

The findings were analyzed using descriptive statistics. The target threshold of responses (5%) with representative proportions from all zones was obtained (see Table 1). The resulting sample of respondents consisted of 172 urban, 208 suburban, and 109 rural physical educators ($n=489$). Results indicated that 38 percent of respondents

Table 2. Survey of physical educators' perspectives and practices with the APPR

Strongly Disagree	Disagree	Agree	Strongly Agree	Unsure
1	2	4	5	3
1. Current educational policy (APPR) will serve to change the purpose of physical education.				
2. As a result of the APPR, my school district is asking me to engage more in mathematics, English language arts, and science content in my physical education class.				
3. It is appropriate for physical educators to be teaching mathematics, English language arts, and science content in their physical education class.				
4. I will have more paper and pencil assessments in my physical education classes as a result of the APPR.				
5. My approach to lesson planning will have more of an academic focus (mathematics, English language arts, and science) as a result of the APPR.				
6. The APPR will allow me to focus more on psychomotor skills.				
7. The APPR will allow me to focus more on affective skills.				
8. The APPR encourages physical educators to “game the system” (for example, setting a low pre-test score that will ensure growth).				
9. The APPR will encourage physical educators to spend more time with students who are most likely to show growth on measures of success in class.				
10. It is appropriate to use fitness tests to evaluate the effectiveness of physical educators.				
11. It is appropriate to evaluate the quality of physical educators based on measures of academic success.				
12. The APPR will improve the quality of public K-12 physical education in New York State.				
13. Listed below are the Society of Health and Physical Educators (SHAPE) America's National Standards for K-12 PE. Please rank them in the order you believe represents their importance.				
<ol style="list-style-type: none"> 1. Competency in a variety of motor skills and movement patterns. 2. Knowledge of concepts, principles, strategies and tactics related to movement and performance. 3. Knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness. 4. Responsible personal and social behavior that respects self and others. 5. The value of physical activity for health, enjoyment, challenge, self-expression, and/or social interaction. 				
14. If you were the sole decision maker, which assessment would you choose to reflect the goals that are stated below? National Content Standard in PE 1.—Competency in a variety of motor skills and movement patterns.				
<ol style="list-style-type: none"> A. Performance-based (i.e., watching a student perform the skill) B. Written test or assessment C. Fitness test 				

Table 2. (continued)

<p>15. If you were the sole decision maker, which assessment would you choose to reflect the goals that are stated below? <i>National Content Standard in PE 2.—Knowledge of concepts, principles, strategies and tactics related to movement and performance.</i></p> <p>A. Performance-based (i.e., watching a student perform the skill) B. Written test or assessment C. Fitness test</p>
<p>16. What type of assessment is your school utilizing in physical education to demonstrate growth (SLOs) as outlined by the APPR?</p> <p>A. Performance-based (i.e., watching a student perform the skill) B. Written test or assessment C. Fitness test D. ELA and/or mathematics, etc.</p>
<p>17. Please identify if your school district is urban, suburban, or rural.</p> <p>Urban Suburban Rural</p>

Question 18 and 19 were coding questions used to search for duplicate submissions.

were evaluated in part on the basis of student results from PE content related paper and pencil tests, while 27 percent of those surveyed indicated their district used student fitness tests for this purpose. Eighteen percent of respondents reported their district used state-mandated ELA and mathematics test scores, and 17 percent of physical educators identified the use of performance-based assessments in calculating physical educator performance ranks.

Seventy percent of physical educators who responded indicated they either agreed or strongly agreed with the prediction that there would be increased use of paper and pencil assessments in PE class in response to the APPR. Similarly, 70 percent of physical educators reported that they were being asked to engage in more mathematics and ELA instruction as a result of the APPR. When questioned about their approach to lesson planning, 54 percent of surveyed physical educators agreed or strongly agreed that their lessons will take on more of an academic focus (mathematics, ELA, science, etc.) as a result of the APPR. Over 70 percent of respondents disagreed or strongly disagreed with the statement predicting that the APPR would allow for an emphasis on the psychomotor domain, while 60 percent of physical educators disagreed or strongly disagreed with the statement predicting that the APPR would allow for an emphasis on the affective domain.

Question 14 and 15 polled physical educators about the types of assessments they might select for their PE program if they were the decision maker. The questions asked respondents to choose what assessment they believed best aligned to standards one and two established by the Society of Health and Physical Educators (SHAPE). Standard 1 states: "The physically literate individual demonstrates competency in a variety of motor skills and movement patterns" (SHAPE America, 2014, p. 12). Results showed that 94 percent of physical educators polled preferred a performance-based assessment versus a written or fitness test, if given the choice. Standard

2 states: “The physically literate individual applies knowledge of concepts, principles, strategies and tactics related to movement and performance” (p. 12). About half of physical educators who responded either favored a performance-based assessment or a written test, while only 3 percent indicated that, if provided the option, they would use a fitness test.

Nearly 70 percent of physical educators surveyed indicated they agreed or strongly agreed that the APPR encourages setting a low pre-test score to ensure growth when compared to post-test scores. More than 35 percent of respondents agreed or strongly agreed with the prediction that the APPR will encourage physical educators to spend more time with students capable of demonstrating the most growth, while 53 percent questioned indicated disagreeing or strongly disagreeing with this prediction. Nearly 80 percent of those surveyed indicated moderate or strong opposition to using student fitness tests to evaluate the effectiveness of physical educators. Finally, 83 percent of physical educators expressed disagreement or strong disagreement with the prediction that the new teacher evaluation system would improve the quality of PE in New York State, and almost 90 percent of respondents disagreed or strongly disagreed with being evaluated based on the academic achievement of students. In fact, in both instances, over 45 percent of physical educators strongly disagreed, compared to less than 1 percent who strongly agreed. These findings are even more significant since it has been found that “the endorsement of a strongly positive category does not mean as much as the endorsement of a strongly negative category” (Foddy, 1994, p. 162).

Discussion

We predicted that the APPR would prompt physical educators to utilize more cognitive assessments in PE. The results from this study support our prediction. A written test in PE was the most commonly reported assessment utilized to comply with the APPR, while a performance-based assessment was reported least frequently. Moreover, a number of physical educators reported being evaluated on the basis of student scores on New York State’s mandated Common Core ELA or mathematics assessments, subjects they did not teach. Seventy percent of physical educators surveyed also reported believing that the state’s teacher evaluation policy would increase paper and pencil assessments in PE class.

We also predicted that New York’s APPR would be perceived by physical educators to limit their focus on the psychomotor and affective domains in PE class. This prediction was also supported. The majority of physical educators surveyed indicated moderate or strong disagreement with the prediction that APPR will cause physical educators to focus on the affective and psychomotor domains.

More insight into this trend can be found in physical educator’s assessment choices for SHAPE Standard 1 and 2 (discussed above). Those surveyed overwhelmingly chose performance-based assessments, even for Standard 2, which is knowledge-based and targets the cognitive domain. This is further evidence that physical educators’ preferred methods of assessment are in contradiction to the State’s current teacher evaluation policy. This also raises additional concerns regarding the possibility that New York’s APPR may—intentionally or unintentionally—alter the nature and purpose of PE.

Interestingly, and in contrast to the surveyed opinions of physical educators in New York regarding their use, fitness tests were the second most reported assessment utilized to comply with the State's APPR. This tendency may reflect a tension for the profession. By using fitness tests for the APPR, teachers might preserve an emphasis on the psychomotor domain but risk an ineffective teacher rating, as student fitness levels are linked to many out-of-school factors. Alternatively, teachers can adopt a written test that may be less related to the content of PE, but easier to use to document student growth. It appears that the APPR in New York has highlighted a key issue for PE: "[o]ne of the problems in physical education has been the lack of practical, reliable, and valid measures of program objectives other than fitness" (Rink, 2013, p. 410).

How do physical educators appraise the practice of using student growth scores to determine teacher effectiveness? Do they believe the new APPR is a sound method for evaluating teachers? Nearly 90 percent of those polled disagreed with using measures of student performance results to assess the competency of teachers, while 80 percent believed that the APPR would not improve the quality of K-12 PE in New York State. This is a cause for concern, but unfortunately, respondents' predictions have historical precedent. Earlier efforts to evaluate teachers based on student exam results yielded no improvement in student learning while in fact harming the quality of education in a variety of ways. Documented outcomes of the Payment by Results era in the late 19th century United Kingdom included "teaching to the test," which narrowed curricula, and a tendency among teachers to focus on students most likely to show gains on exams (Garrison, 2011).

A large portion of physical educators polled agreed or strongly agreed that the APPR encourages the teacher to set a low pre-test score that ensures student growth in performance. Conversely, participants were not as unanimous in predicting that the APPR will cause physical educators to spend more time with those students most likely to show growth. It should be noted that the unsure response for this item (11%) was the highest reported for any item in the survey. This may suggest that "gaming the system" is less an issue for PE professionals than it is for their core academic subject counterparts (see Nichols & Berliner, 2007).

Finally, on the issue of whether the APPR will change the purpose of PE, physical educators were almost evenly split. This is interesting and to some degree conflicts with other results observed in this study. For example, the majority of physical educators polled agreed or strongly agreed that they would implement more paper and pencil assessments as a result of the APPR, and they did not agree that the APPR would allow for a focus in the psychomotor and affective domains. This suggests that the new teacher evaluation policy may be shifting the focus of physical educators to the cognitive domain, which is more readily gauged by paper and pencil tests that are themselves more readily used in VAMs. Thus the profession should ask: are VAMs of teacher effectiveness, such as that adopted in New York, causing a de facto change in the aim of PE programs, that is, less emphasis on the non-cognitive domains?

Limitations

While we were able to identify and remove two instances where randomly selected respondents forwarded the survey to their colleagues, it was still possible for physical

educators to forward the survey to other colleagues without our knowledge. Another issue communicated to us by two subjects concerned item 16 that asked physical educators to identify the type of assessment being utilized in their school district to comply with the APPR. This question provided respondents with only one option, yet some physical educators may teach multiple grades, and as a result, may be subject to different types of assessments. Finally, while measures were taken in the construction of the survey to limit leading respondents to be more critical of the new policy, it was possible that participants may have felt pressure to express negative views of the APPR given the strong negative reaction to the policy by teacher unions (Harris, 2015). However, while many items elicited a clear majority perspective among physical educators in New York State with respect to several aspects of the new policy, item 9 revealed no clear uniformity of views. This is important because item 9 deals with a key debate regarding the effect of the new educational policy on teacher practices. For this item, physical educators expressed in near equal measure both agreement and disagreement with the statement that “the APPR will encourage physical educators to spend more time with students who are most likely to show growth on measures of success in class.” This can be taken as further evidence that the survey did not pressure respondents to misrepresent their attitudes about the APPR.

Conclusions

The results of this study contribute to the growing line of research on teacher evaluation and its influence on PE. Respondents reported strong reservations about using value-added logic (student growth scores) to evaluate physical educators. Low unsure response rates (below 10%) for many survey items indicate that physical educators in New York State have definite views about the State’s evaluation system (APPR). While it is possible that the adoption of the policy is too recent for it to have had significant effects on physical educators’ practices and perceptions, results do demonstrate that practitioners are focused on the affective and psychomotor domains, despite the APPR’s apparent emphasis on cognitive outcomes. The findings also suggest that physical educators’ views have not yet changed in terms of how they wish to assess their students and be evaluated. This study provides a baseline for future efforts to track changes in attitudes and reported practices over time.

In conclusion, it is apparent that physical educators are concerned about the APPR and how it may affect PE programs moving forward. As in Metzler (2014), this study raises the question of whether the future of teacher effectiveness research in PE will be determined by professionals in PE or by those driving educational policy agendas. Whatever the case, the metrics by which physical educators and students in PE are evaluated must be closely investigated. The future of PE programs and tomorrow’s physical educators may hang in the balance.

Note

1. In December 2015, the New York State Board of Regents approved a moratorium on using grade 3–8 mathematics and English language arts (ELA) test scores for teacher effectiveness ratings. The use of APPR calculations will continue with recommendations that individual educators use the metrics to help

improve their teaching practices. Decisions about tenure and promotion cannot be made using the APPR (see Woodruff, 2016).

References

- Baker, B.D., Oluwole, J.O., & Green P.C., III. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives*, 21(5). doi:10.14507/epaa.v21n5.2013
- Berg, K. (2010). Justifying physical education based on neuroscience evidence. *Journal of Physical Education, Recreation & Dance*, 81(3), 24–29, 46.
- Claxton, D. (2012). The politics of physical education. *Quest*, 64(3), 141–149.
- Corcoran, S.P. (2010). Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. *Education Policy for Action Series*. Annenberg Institute for School Reform at Brown University (NJ1). Retrieved March 11, 2015, from <http://eric.ed.gov/?id=ED522163> .
- Ennis, C.D. (2011). Physical education curriculum priorities: Evidence for education and skillfulness. *Quest*, 63(1), 5–18. doi:10.1080/00336297.2011.10483659
- Foddy, W. (1994). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. New York: Cambridge University Press.
- Garrison, M.J. (2011). On the origin and political significance of test-based teacher evaluation and compensation. *Journal of Inquiry and Action in Education*, 4(1). Retrieved September 14, 2016, from <https://journal.buffalostate.edu/index.php/soe/article/viewArticle/123> .
- Garrison, M. J. (2015). Measurement as politics by other means: The case of test-based teacher evaluation. In K. E. O'Hare (Ed.), *Teacher evaluation: The charge and the challenges* (pp. 39–58). New York: Peter Lang Publishing.
- Harris, E.A. (2015). Cuomo gets deals on tenure and evaluations of teachers. *The New York Times*. Retrieved September 14, 2016, from <http://www.nytimes.com/2015/04/01/nyregion/cuomo-gets-deals-on-tenure-and-evaluations-of-teachers.html> .
- Martin, B. (2012). An increased role for the Department of Education in addressing federalism concerns. *Brigham Young University Education & Law Journal*, (1), 79–110.
- Metzler, M.W. (2014). Teacher effectiveness research in physical education: The future isn't what it used to be. *Research Quarterly for Exercise and Sport*, 85(1), 14–19.
- Nichols, S.L., & Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, Mass.: Harvard Education Press.
- The New York State Association for Health, Physical Education, Recreation, and Dance (NYS AHPERD). (2015). *About*. Retrieved September 27, 2016, from <http://www.nysahperd.org/content/about/nys-ahperd-zone-map.cfm> .
- Rink, J.E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84(4), 407–418. doi:10.1080/02701367.2013.844018
- Seymour, C. M., & Garrison, M.J. (2015). Is the physical being taken out of physical education? On the possible effects of high-stakes testing on an embattled profession's curriculum goals. *Quest*. 67(4), 400–408. Retrieved September 27, 2016, from <http://dx.doi.org/10.1080/00336297.2015.1085422> .
- Seymour, C.M., (2014). *The changing purpose of physical education in the era of test based teacher evaluation and compensation*. (Doctoral Dissertation). D'Youville College, Buffalo, NY.
- SHAPE America. (2014). *National Standards & Grade-Level Outcomes for K-12 Physical Education*. Human Kinetics.
- Trost, S.G., & van der Mars, H. (2009). Why we should not cut P.E. *Educational Leadership*, 67(4), 60–65.
- Van Dalen, D.B. (1971). *A world history of physical education: Cultural, philosophical, comparative* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Woodruff, C. (2016). *Big APPR changes for NYS teachers, but test results still count for districts*. New York State School Boards Association. Retrieved September 7, 2016, from <http://www.nyssba.org/news/2016/01/07/on-board-online-january-11-2016/big-appr-changes-for-nys-teachers-but-test-results-still-count-for-districts/> .