Value Added Models and the Implementation of the National Standards of K-12 Physical Education

Clancy M. Seymour
Canisius College

Mark J. Garrison
D'Youville College

The implementation of value-added models of teacher evaluation continue to expand in public education, but the effects of using student test scores to evaluate K-12 physical educators necessitates further discussion. Using the five National Standards for K-12 Physical Education from the Society of Health and Physical Educators America (SHAPE), physical educators in New York State were polled about the most important goals of physical education and how value-added models may be affecting physical education practices. Participants were drawn using a proportionate stratified random sample (n=489). Standard 5 was selected as the most important by 36% of physical educators who responded, while standard 3 was chosen as most important by 33% of respondents. Thirty eight percent of physical educators reported that their performance reviews were based on student growth scores on written tests, 27% reported that their district selected fitness tests, standardized tests in English Language Arts or mathematics were reported being used by 18% of respondents, and performance-based assessments were reported being used by 17% of those completing the survey. The authors concluded that the affective domain (crucial to SHAPE standard 5) appears to be overlooked by policies that use student performance data to determine teacher effectiveness.

Introduction

In response to the Race to the Top (RTTT) federal competition grant, teacher evaluation policies in New York State have moved in a new direction since September 2012. As a winner of the RTTT federal funds, the State of New York has adopted the use of Value-Added Models (VAMs) to evaluate teachers. A VAM utilizes a student's change of test scores over time to demonstrate a teacher's effectiveness. Baker, Oluwole, and Green (2013) explained that VAM models use:

> assessment data in the context of a statistical model (regression analysis), where the objective is to estimate the extent to which a student having a specific teacher or attending a specific school influences that student's difference in score from the beginning of the year to the end of the year – or period of treatment (in school or with teacher). (p. 7)

Public schools in New York State began using the Annual Professional Performance Review (APPR) requiring teachers in academic content areas like English Language Arts (ELA) and mathematics to use student performance scores on standardized state tests to measure teacher effectiveness. Currently, a growth score for a teacher is obtained from a comparison of a pre and post-test. This metric accounts for 40% of a teacher's yearly "effectiveness" ranking, while the remaining 60% of a teacher's evaluation is completed using classroom observations that are locally negotiated and then approved by the State. The combined score is used to place teachers in one of four possible categories: highly effective, effective, developing, or ineffective. [1]

In academic content areas without a state-mandated test, a growth score must be calculated using assessments approved by the local district and the New York State Department of Education (NYSED) (Baker, Oluwole, & Green, 2013). Therefore academic subjects like physical education (PE) may utilize student performance scores from a fitness test, performance-based assessment, written test or a state-mandated test score from other content areas like ELA or mathematics (Rink, 2013). Upon acceptance, physical educators use pre- and post-test logic to set a quantitative goal for their students to attain. These are called Student Learning Objectives (SLOs) and are designed to mirror the same process that takes place in academic content areas with a state-mandated standardized test. A teacher is deemed effective if his or her students reach the intended benchmark documented by the instructor.

While the literature on VAMs of teacher evaluation continues to broaden, there is little research that investigates the implications of utilizing these methods in PE. Furthermore, today's educational reform agenda stressing cognitive outcomes and quantitative test scores (see Feingold, 2013) necessitates an investigation of what physical educators believe are the most important components of PE and how current educational policy may affect its implementation.

Background

In 1986, the National Association for Sport and Physical Education (NASPE) assembled a Blue-Ribbon Committee that sought to determine the appropriate outcomes of PE in K-12 schools (Metzler, 2011). This culminated in 1995 with the publication of the NASPE Content

[1]A moratorium on utilizing grades 3-8 ELA and mathematics test scores for teacher effectiveness ratings was approved by the New York State Board of Regents in December 2015. School districts will continue to collect this data and individual educators have been encouraged to use these calculations to help strengthen their teaching. Tenure and promotion decisions cannot be made by using the APPR (see Woodruff, 2016).

Standards. The content standards specified, "what students should know and be able to do" (NASPE, 1995, p. vi) in PE. Despite being a significant milestone for the PE profession, continued efforts to make public education more accountable in the late 1990s prompted No Child Left Behind (NCLB) legislation, which was enacted in January of 2002. As a result of the law's focus on mathematics and ELA test scores, many PE professionals became concerned that PE and health experiences would become undervalued in K-12 schools (Cook, 2005; Filburn & Fletcher, 2008; McKenzie & Lounsbery, 2009; NASPE, 2010; Smith & Lounsbery, 2009; Trost & Van Der Mars, 2009).

As a result, NASPE appointed the K-12 National Physical Education Standards Review Committee in the summer of 2002 and the second edition of the NASPE Content Standards were revised and published in 2004 (NASPE, 2004). In this new version, NASPE commented on NCLB, citing passages from the law itself reinforcing the need to "close this achievement gap, with accountability, flexibility, and choice so that no child is left behind" (NASPE, 2004, p. 2). Although the NCLB deadline has since passed, President Obama's own version of NCLB (RTTT) has resulted in a large number of states enacting VAMs to compete for federal RTTT revenue (Rink, 2013). This has prompted a similar response from the PE profession, with a third edition of the national standards approved and published by the newly reorganized professional association entitled SHAPE (formerly NASPE) in May 2014.

The PE standards movement in the State of New York shares a history similar to SHAPE and largely driven by the accountability era in public education since 1980. However, the notable exception rests in its amendment cycle. The latest version of the New York State Standards in PE has not been revised since 1996 (The University of the State of New York, 1996). Therefore, revised twice prior to its current 2014 edition, we have selected SHAPE's National Standards for K-12 PE to be the focus and comparative tool for this analysis.

Given the above account, it seems likely that the implementation of PE standards in the classroom may be affected by a policy like the APPR. For example, are there standards that receive more attention as a result of the APPR, especially when PE opportunities are not equally distributed as is the case with "children from communities that struggle with poverty and reduced school funding" (Seymour & Garrison, 2015, p. 404)? That said it would be useful to examine if in complying with the APPR, do urban, suburban, and rural school districts exhibit different patterns in PE with respect to the national standards and the goals they reflect. Finally, could the

APPR, and similar policies, both directly and indirectly change the focus of PE programs and the utilization of the national standards?

The following research questions for this study were developed to address these concerns:

1) What do physical educators rank as the most important components of PE?
2) Do the rankings of the most important components of PE correlate with the type of school district (urban, suburban, and rural) reported?
3) Is there a difference between what physical educators rank as the most important components of PE and the type of metric they report using in their school district for the APPR?

The results of this analysis are derived from a larger representative study surveying physical educators about their perspectives and practices with the APPR in New York State (Seymour, 2014).

Method

Participant Selection

Following Institutional Review Board (IRB) approval, the study was conducted during the 2013-2014 public school year from an online, anonymous survey distributed to physical educators across New York State via email. A proportionate, stratified, random sample of New York State public school physical educators were polled about their perspectives and their district's practices with the APPR. The 11 geographic zones adopted by the professional organization of PE in the State of New York were the designated strata utilized in this study (New York State Association for Health, Recreation, and Dance (NYS AHPERD), n.d.). Email addresses of nearly 50% of physical educators were manually retrieved from schools based on a list of PE professionals in New York State provided by the New York State Education Department (NYSED) (see Table 1). Demographic information such as race, ethnicity, gender, and age was not the focus of the study, and therefore was not collected.

The survey was distributed during an eight-week period in multiple waves to 20% of randomly sampled physical educators from each stratum. The distribution cycle was staggered into 4, two-week phases where 5% of physical educators randomly sampled from each stratum were emailed. This yielded a 5% response rate (n=489) with a maximum margin of error of 4.32% ($p < .05$). Proportionality was achieved by obtaining representative thresholds (5%) for each of the 11 NYS AHPERD zones/strata (Table 1).

Data Collection

Responses from each zone were calculated after each of the four phases of email solicitations requesting participation. Physical educators randomly sampled in each phase were eliminated from future samples of that zone. All randomly selected physical educators sampled in 5 of the 11 zones, which had not met threshold were sent a final email. Responses were collected from 489 physical educators, 5% of the total PE population in the State (Table 1).

The researchers experienced unforeseen email retrieval issues in the New York City zone. A large proportion of schools in the New York City Department of Education (NYCDOE) do not provide teacher names and/or email addresses on school or department websites. It was determined from the small percentage of available email addresses that the first initial and last name followed by @nyc.schools.gov was the email naming convention utilized by the NYCDOE. To correctly predict the email addresses of over 2,700 physical educators alongside the potential confusion of other subject teachers within this region proved to be a challenge. Therefore using the above email naming convention as a guide, the survey was sent to all physical educators in this zone while still pursuing the original threshold (136 responses or 5%). The survey was no longer distributed once the target was obtained.

Table 1:
Distribution of PE Teachers and PE Teacher Response Threshold by phase and NYS AHPERD Zone

Zone	Teacher Totals (%)	Phase 1	Phase 2	Phase 3	Phase 4	Responses Obtained
*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 (Seymour & Garrison, 2015)

Instrumentation

Validation of the survey began with a pilot in the summer of 2013. The survey was designed to answer the research questions and evaluate practices and attitudes of physical educators towards the APPR. During the pilot, a focus group of physical educators in the local zone of the researchers were questioned about the survey's authenticity and feasibility. The researchers adopted all proposed amendments by the focus group, which included both item revision recommendations and supplementary questions about appropriate techniques that can be used for teacher evaluation in PE. To further strengthen the survey's validity, the instrument was also reviewed by two teacher education faculty with expertise in PE teacher education and curricular development. To simplify and better align the instrument to the research questions and objectives of the study, additional edits to the survey language were also made.

Table 2: Survey of Physical Educators Perspectives and Practices with the APPR (A, B, C)

A. Listed below are the Society of Health and Physical Educators America's National Standards for K-12 PE. Please rank them in the order you believe represents their importance.
The physically literate individual: 1.) Demonstrates competency in a variety of motor skills and movement patterns. 2.) Applies knowledge of concepts, principles, strategies and tactics related to movement and performance. 3.) Applies knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness. 4.) Exhibits responsible personal and social behavior that respects self and others. 5.) Recognizes the value of physical activity for health, enjoyment, challenge, self-expression and/or social interaction.
B. What type of assessment is your school utilizing in physical education to demonstrate growth (SLOs) as outlined by the APPR?
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.
C. Please identify if your school district is urban, suburban, or rural.
Urban Suburban Rural

Adopted from a larger study (Seymour, 2014), three survey items linked to the research questions were utilized for this analysis (Table 2). Item A linked to research question 1 and required respondents to rank SHAPE's National Standards of K-12 PE in order of importance. This approach did counter SHAPE's recommendations not to prioritize the standards in order of importance (SHAPE, 2014), yet was still pursued because not all PE programs and states comply and utilize the same standards (Baghurst, Langley, & Bishop, 2015). The regulation of PE occurs differently in each state and in many cases each state has individual standards in PE (Baghurst, Langley, & Bishop, 2015). At the same time, in New York State where the study took place, the State Learning Standards in PE have not been revised since 1996 (The University of the State of New York, 1996). The selection of SHAPE's National Standards for K-12 PE provided the most current standards to use in research regarding the effects of VAMs on the evaluation of PE professionals.

Survey item B which aligned to research question 2 asked respondents to report their district's teacher evaluation practices in accordance with the APPR (Table 2). Linked directly to research question 3, survey item C required physical educators to report their school district type (urban, suburban, rural). No duplicate surveys were found and any incomplete submissions were discarded.

Results

The 5% target threshold of responses was achieved in all zones (see Table 1) and the results were analyzed using descriptive and inferential statistics. School type as reported by physical educators (n=489) indicated that the sample consisted of 172 physical educators working in an urban district, 208 physical educators working in a suburban district, and 109 physical educators working in a rural district (Table 3).

Research question 1 asked physical educators to prioritize the national standards using a 1-5 ranking depicted in Table 3 below. Results demonstrated that standard 5 was ranked most important by the greatest number of physical educators (36%) followed closely by standard 3, ranked the most important by 33% of respondents. A much smaller percentage of physical educators reported ranking standard 4 (14%) or standard 1 (13%) as most important. Standard 2 was recognized as most important by the lowest percentage (5%) of physical educators across all settings.

Table 3: Reporting of Physical Educators Most Important Standard by Type of School District

SHAPE Standard…The physically literate individual …	Urban	Suburban	Rural	Total	(%)
1. Demonstrates competency in a variety of motor skills and movement patterns	24	22	17	63	12.88
2. Applies knowledge of concepts, principles, strategies and tactics related to movement and performance	9	11	4	24	4.91
3. Demonstrates the knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness	46	73	40	159	32.52
4. Exhibits responsible personal and social behavior that respects self and others	27	25	17	69	14.11
5. Recognizes the value of physical activity for health, enjoyment, challenge, self-expression and/or social interaction	66	77	31	174	35.58
Total	172	208	109	489	100

(SHAPE, 2014, p. 12.)

Research question 2 sought to compare a physical educator's ranking of SHAPE's national standards to the type of school district, and the assessment utilized by that school district to evaluate physical educators (Table 4). The most common district measure reported by physical educators to comply with APPR was a written test (38%). Student fitness test scores were indicated in use by 27% of those surveyed. District utilization of State-mandated ELA and mathematics test scores to calculate physical educator teacher effectiveness was reported by 18% of respondents and performance-based assessments were reported in use by 17% of physical educators.

Table 4: Reported Assessments Utilized to Document Student Growth (SLOs) for APPR in New York State

	Performance Based Assessment	Written Test	Fitness Test	ELA or Math	Total
Responses	83	184	134	88	489
Percentages	16.97	37.63	27.4	18	100

Physical educator rankings of the standards by the type of school district are summarized in Table 3. For teachers from urban and suburban settings, standard 5 was ranked highest at 38% and 37% respectively, while standard 2 was prioritized the least by both groups of teachers at 5%. On the other hand, 37% of physical educators who worked in a rural district prioritized

standard 3 the most and scored standard 2 the least (4%). District differences were analyzed further using a chi-square test to determine if there were significant differences among groups. Results revealed no significant differences between district types in ranking of national standards.

To answer research question 3, a chi-square analysis was performed to determine if there was a difference between physical educators' ranking of the national standards and the type of metric reported in use currently by the school district to comply with the APPR. There was no statistically significant correlation between variables. The reported APPR assessment selected by districts did not correlate with what physical educators rank as the most important standards for PE.

Discussion

When asked to rank the national standards, over 67% of physical educators ranked standard 5 (35.58%) and 3 (32.52%) as the most important compared to nearly 33% of physical educators who rated standard 1, 2, and 4 as most important. Most physical educators appeared to believe their work should focus on the value of physical activity for health, enjoyment, challenge, self-expression and social interaction (standard 5) and building knowledge and skills to achieve and maintain a health-enhancing level of physical activity and fitness (standard 3). Although both standard 3 and 5 are crafted to emphasize the importance of physical activity and an active lifestyle, standard 3 targets academic knowledge related to PE, while standard 5 is geared towards the affective domain. Therefore the question that must be asked is: are written tests used to measure teacher effectiveness stressing cognitive knowledge about fitness equally to the value of a healthy lifestyle? It is concerning that although standard 5 was more frequently ranked the most important by surveyed PE professionals, tasks that focus on the affective domain are not used to document student growth linked to teacher proficiency as determined by the APPR.

At the same time, while our analysis revealed no significant difference among physical educators from different regions and no association was found between what physical educators rank as the most important standards for PE and the type of assessment their district utilized for the APPR, there are trends in the data that should not be overlooked. Decisions about VAMs of teacher evaluation do not appear to be connected to PE teacher professional judgments about the

proper goals of and best practices for PE. It should not be ignored that 18% of physical educators reported their district used ELA and mathematics test results to document student learning in PE as required by the APPR. This suggests that despite their views about the most important standards of PE, many physical educators must adapt to a system that uses less than ideal measures to determine teacher effectiveness within this policy context. It also substantiates Seymour and Garrison's (2015) claims that assessments in PE may not align with the curriculum, but instead are chosen simply because they are easiest to use to satisfy the APPR. Indeed it could be argued that the assessment tail may be wagging the curriculum dog in K-12 public schools of New York State (Seymour & Garrison, 2015).

Limitations

As is customary with survey research, the researchers were unable to determine the accuracy of responses and it was assumed that respondents provided honest submissions. In addition, using a proportionate stratified random sample meant participants should have refrained from forwarding the survey to other colleagues. This was communicated to subjects several times throughout the study, but it was still possible that this took place without our knowledge. The survey itself may also have contributed to inaccuracies in the data. Item A asked physical educators to rank the national content standards in order of importance. A ranking question is an ordinal measurement and the distance between levels (1-5) cannot be determined (Foddy, 1994). As explained previously, this question conflicted with recommendations of SHAPE about prioritizing the standards (SHAPE, 2014). As a result, physical educators familiar with this provision may have been uneasy about answering the question. In the future this item could be amended asking the physical educator to estimate the instructional importance of and/or time spent on each of the standards.

Conclusion

It appears that current policies like the APPR have not disrupted the perspectives of physical educators about the important components of PE. This is a good sign and means that although VAMs may be indirectly shifting the focus of PE, the personal views of physical educators have not wavered. Therefore sustained research on how teacher evaluation policies—whatever form they take—is warranted in light of what these findings suggest about possible

effects on PE.

Finally, recent accountability trends in public education may be imposing indirect curricular changes to all content areas that favor quantitative test scores (Feingold, 2013). For subjects like PE, this may hinder affective domain learning which many consider to be critical (see Hellison, 2011). Further demonstrating the affective domain's importance to K-12 PE, the professional association itself—SHAPE—has devoted two of its five national standards (standard 4 and 5) to its role in a quality K-12 PE program. Consequently, Metzler (2014) has cautioned the PE profession that the current direction in public education may lead to questionable curricular content no longer developed by PE experts, but instead determined by educational policymakers or the highest bidding textbook publishing companies. The profession and in particular physical educators must recognize these concerns and be purposeful when selecting assessments to document teacher proficiency under this current policy context.

References

Baker, B. D., Oluwole, J. O., & III, P. C. G. (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*(0), 5. doi:10.14507/epaa.v21n5.2013

Baghurst, T., Langley, J., & Bishop, J. C. (2015). Physical educators' perceptions of their use of NASPE Standards. *The Physical Educator*, *72*, 324–341.

Cook, G. (2005). Cut to fit. *American School Board Journal*, *192*(8), 16–19.

Feingold, R. S. (2013). Vision in an age of accountability. *Quest (00336297)*, *65*(4), 385–393. doi:10.1080/00336297.2013.834832

Filburn, A. D., & Fletcher, J. S. (2008). NCLB: The next step in the diminishment of physical education and health. *Illinois Journal for Health, Physical Education, Recreation & Dance*, *61*, 58–60.

Foddy, W. (1994). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. New York: Cambridge University Press.

Hellison, D. (2011). *Teaching responsibility through physical activity* (3rd edition). Champaign, IL: Human Kinetics.

McKenzie, T. L., & Lounsbery, M. A. F. (2009). School physical education: The pill not taken. *American Journal of Lifestyle Medicine*, *3*(3), 219–225. doi:10.1177/1559827609331562

Metzler, M. W. (2011). *Instructional models for physical education* (3rd edition). Holcomb Hathaway, Incorporated.

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. doi:10.1080/02701367.2014.872932

National Association for Sport and Physical Education. (1995). *Moving into the future: National physical education standards: A guide to content and assessment*. WCB McGraw-Hill.

National Association for Sport and Physical Education. (2004). *Moving into the future: National standards for physical education* (2nd edition). Reston, VA.

The New York State Association for Health, Physical Education, Recreation, and Dance (NYS AHPERD). *About*. (n.d.). Retrieved November 2, 2013, 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., & Garrison, M. J. (2016). What they think about how they are evaluated: perspectives of New York State Physical educators on teacher evaluation policy. *International Journal of Education Policy and Leadership*, *11*(6).

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.

Smith, N. J., & Lounsbery, M. (2009). Promoting physical education: The link to academic

achievement. *Journal of Physical Education, Recreation & Dance*, *80*(1), 39–43. doi:10.1080/07303084.2009.10598266

The University of the State of New York. (1996). *Learning Standards for Health, Physical Education, and Family and Consumer Sciences*, (revised edition). Albany, NY

Trost, S. G., & Van Der Mars, H. (2009). Why we should not cut P.E. *Educational Leadership*, *67*(4), 60–65.

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/.