The Physical Educator Vol. 76 • pp. 427–439 • 2019
https://doi.org/10.18666/TPE-2019-V76-I2-8541

PEDAGOGY

Bias in Physical Educators' Grading Practices With ELA Literacy

Clancy M. Seymour and Kristin E. Finn

Abstract

As accountability initiatives in public education such as the Common Core State Standards (CCSS) begin to gain traction, it is apparent that physical educators must integrate English Language Arts (ELA) literacy tasks into their daily instruction. This study tested whether physical education (PE) teachers have less experience and exhibit bias when grading ELA literacy tasks, compared to non-PE teachers. Participants from this study included 201 teachers from 77 schools in a local region of New York State. Results revealed that approximately 90% of PE teachers reported having little to no experience with grading ELA literacy assignments. In addition, physical educators were significantly more critical than their non-PE counterparts; ELA assignments were judged as being lower quality, being less neat, having poorer sentence structure, and showing less effort. The lack of consistency between PE teachers and non-PE teachers raises questions about validity and whether there is a need for more professional training should the CCSS mandate continue in PE.

Clancy M. Seymour is the director of Physical and Health Education Teacher Education, and assistant professor, Department of Kinesiology, Canisius College. Kristin E. Finn is a professor of Teacher Education, School of Education and Human Services, Canisius College. Please send author correspondence to seymourc@canisius.edu

The educational standards movement in the United States in the early 1990s prompted each state to define its own grade-level proficiencies (Common Core State Standards Initiative [CCSI], n.d.-b). As a result, every state by the early 2000s had developed its own learning standards. While this was a promising improvement, the lack of uniformity among states necessitated a more standardized set of goals (CCSI, n.d.-b). The Council of Chief State Officers and the National Governors Association developed the Common Core State Standards (CCSS) in 2009 to operationalize learning in K–12 English Language Arts (ELA) and mathematics (CCSI, n.d.-a). The CCSS have been adopted in 42 states and four territories in the United States (CCSI, n.d.-b).

Although standards for physical education (PE) were not included within this initiative, the CCSS defined PE as a technical subject and called for it to aid in the acquisition of the CCSS (CCSI, n.d.-c; James & Bullock, 2015). This resulted in the request that physical educators consider integrating ELA and mathematical literacy into daily lessons (James-Hassan, 2014). It should be noted that while the CCSS encourages ELA and mathematics integration, it expects only that physical educators support ELA learning.

The literature addressing the general effects of the CCSS from a positive and negative context is robust (e.g., Accountability Works, 2012; Au, 2013; Conley, 2011; Mathis, 2010; Saltman, 2014), while there is a dearth of research examining the CCSS's implications on PE. At the same time, prescriptive publications that provide useful strategies for physical educators to utilize and integrate the CCSS and ELA are plentiful. For example, Ballinger and Deeney (2006), James and Bullock (2015), James-Hassan (2014), Mercier, Whitley, and Manson (2014), and Scrabis-Fletcher (2016) have suggested techniques such as class projects, entry and exit slips, summary paragraphs, and readings that PE instructors can require K–12 students to complete in and out of class.

The adoption of the CCSS reinforces an interdisciplinary approach to literacy that has been stressed in PE teacher education for quite some time (James & Manson, 2015; Mercier et al., 2014). In a review of the integration literature in PE from 2004 to 2013, Marttinen, McLoughlin, Fredrick, and Novak (2017) found that it may be helpful for physical educators to consider interdisciplinary

themes to reinforce PE's standing as part of a well-rounded education. However, Marttinen et al. (2017) cautioned that the most common subject for integration is mathematics and that the literature on ELA interventions is limited. This is an important point given that the CCSS defines PE as a technical subject with a supportive role for physical educators in reaching the ELA standards.

Conversely, critics of the reform argue that content areas such as PE have not been considered in the development of the CCSS, which devalues the relationship between the mind and body (Magnotta & Darst, 2015). Additionally, Lounsbery and McKenzie (2015) and Seymour and Garrison (2015) argue that the current educational reform agenda has ". . . triggered a possible shift in the goals of PE . . . and have physical educators scampering to comply with mandates that pressure them to adopt traditional methods of assessment that focus on the cognitive domain" (Seymour & Garrison, 2015, p. 406).

Incorporating ELA literacy into PE classes may also be problematic given the current educational climate in public education regarding teacher evaluations. For example, 30 states now use student test scores (value-added model) as part of a teacher's effectiveness rating (Rink, 2013). Seymour and Garrison (2016) recently showed that only 38% of New York State physical educators reported utilizing students' written PE test results for teacher evaluation purposes and 18% reported the use of state-mandated ELA and mathematics tests scores that align to the CCSS. Therefore, in addition to their own disciplinary content, physical educators are being asked to infuse literacy concepts into their daily instruction (Mercier et al., 2014). As a result, physical educators could be utilizing assessments that focus on literacy, yet may need more training with ELA integration strategies and confidence applying these principles in the gymnasium (Constantinou & Wuest, 2015). Furthermore, the integration of literacy content requires that PE teachers learn to assess and evaluate these tasks, which may require additional training and resources for professional development (Collier, 2011; Mercier & Doolittle, 2013). Thus, it seems logical that we question how accurately PE teachers can assess ELA literacy content. This study examined the validity of PE teachers assessing ELA literacy content. Specifically, we compared PE and non-PE teachers on grading an ELA literacy assignment. The following research questions were tested:

1. Do elementary and secondary PE teachers report equivalent experience with grading ELA literacy assignments, compared to non-PE teachers?
2. Are ELA literacy assignments comparably graded by elementary and secondary PE teachers and non-PE teachers?
3. Do elementary and secondary PE teachers rate the quality of work on ELA literacy assignments the same as non-PE teachers do?

Method

Participants

Participants in this study were 201 teachers working in 28 high, 18 elementary, 13 K–8, and 18 middle schools in the Northeastern United States. Over half the teachers (62%) were female. About 8% of teachers were under 30 years of age, 30% were between ages 30 and 39 years, nearly half (48%) were between ages 40 and 49 years, and 14% were 50 years or older. Regarding subject matter, 15% of teachers indicated typically teaching ELA, 17% mathematics, 13% science, 10% history, 6% physical education/health, and 38% other (e.g., language, special education, music, technology). Concerning number of teaching years, experience ranged from first year to 45 years (M = 16.2, SD = 7.0) of teaching.

Procedure and Measures

Surveys were administered to teachers electronically and in person. Invitations were electronically sent to 80 physical educators, and 68 voluntarily participated in the study (85% response rate). The online data (n = 68) were collected through the Qualtrics survey management system. In-person data (n = 133) were collected from teachers during a mandatory monthly staff meeting at their respective schools.

Teachers read and evaluated a short handwritten essay written by a sixth-grade student. Teachers were falsely led to believe that in an effort to deemphasize classroom testing, this study was conducted to determine whether individual pieces of authentic classroom writing assignments could be used to validly assess student achievement levels. For an increase in the authenticity of the task, a photograph of the student, fictitious birthdate, age, and grade information

accompanied the essay. After reading the essay, teachers completed a brief survey that assessed ratings of the student's work.

To decrease confounding factors and increase validity, the study investigators included in the research design four essays and four students. Thus, each teacher was randomly assigned to one student–essay combination. The brief essays were typical writing samples obtained from three independent Grade 6 classrooms (one private school, two public schools). As part of a classroom writing exercise, teachers in each school asked students to write a brief essay about health and fitness. The study investigators selected four essays to be used in this study. They chose the essays because they were deemed to be of average writing quality indicative of a Grade 6 student. Three external educators who have expertise in writing skills at the middle school level judged face validity of the essays. They judged essays on overall quality, sentence structure, and word choice. They deemed all essays to be within the range of average quality; no essay received either extremely high or extremely low quality ratings from the external judges. Headshot photographs of four female middle school students were taken by one of the investigators after obtaining parental permission. All of the girls attended schools other than the ones that participated in this study and were the same race (White) with straight shoulder-length hair. The study investigators chose these girls because they were physically deemed to be typical of middle school girls of this age.

Type of teacher. As part of their background information, teachers indicated the subject area they typically teach. Response options included ELA, mathematics, science, history, physical education/health, and other. This variable was dichotomized, which allowed for comparisons between teachers of PE and all other subjects.

Grading experience. A single item, "How much experience do you have grading essays from students in Grades 6–8?" measured grading experience. Teachers responded using a 4-point scale ranging from 1 = *none* to 4 = *a lot*.

Ratings of the work. Teachers graded the essay using seven quality ratings. Overall quality was judged on a 5-point scale (1 = *below average* and 5 = *above average*). Teachers assigned a grade to the essay using a 10-point scale (1 = *F* and 10 = *A*). They also judged essays on sentence structure, vocabulary/word choice, organization,

and creativity. Each rating was based on a 5-point scale ranging from *poor* to *excellent*.

Teachers also judged on a 5-point scale (1 = *extremely low* and 5 = *extremely high*) the amount of effort the student expended when writing the essay. They assessed neatness of the student's work on a 5-point scale (1 = *extremely low* and 5 = *extremely high*). Finally, teachers were asked if they would recommend that the student receive remedial help with writing (1 = *definitely not* and 4 = *definitely yes*).

Analysis

For the first research question, a chi-square test of association examined whether PE teachers reported less experience grading essays than did non-PE teachers. For all other research questions, the data were analyzed with a two-way analysis of variance (ANOVA) model for unequal cell *n*s. Type of teacher and essay were the factors of classification, and ratings of the work were the dependent measures. The primary research analyses addressed whether PE teachers gave equal ratings of work quality, compared to non-PE teachers. While not a primary research question, including the essay factor in the model ensured that the relationship between type of teacher and essay ratings was not specific to any particular essay (i.e., no interaction between type of teacher and essay). Given the nondirectional hypotheses for the main effect of type of teacher, two-tailed tests of significance were used with the alpha level set to .05. Type I sums of squares tested for main effects—controlling for prior main effects in the analysis—and for interactions that held constant the main effects.

Results

Table 1 shows the results for the first research question. The significant chi-square test of association indicated that PE teachers reported different levels of experience grading essays, compared to non-PE teachers, $\chi^2(3) = 22.58$, $p < .01$. Approximately 90% of PE teachers indicated they had either no or little experience grading essays, compared to 60% of non-PE teachers who also reported no or little experience. Further, no PE teachers perceived they had a lot of experience grading essays, compared to 18% of non-PE teachers.

Table 1
Percentages for PE and Non-PE Teachers Who Reported Experience Grading Student Essays

	Type of teacher		
Experience grading essays	**PE**	**Non-PE**	**Total**
None			
Count	42	44	86
% within experience grading essays	48.8	51.2	100.0
% within type of teacher	55.3	35.5	43.0
Little			
Count	26	31	57
% within experience grading essays	45.6	54.4	100.0
% within type of teacher	34.2	25.0	28.5
Some			
Count	8	27	35
% within experience grading essays	22.9	77.1	100.0
% within type of teacher	10.5	21.8	17.5
A lot			
Count	0	22	22
% within experience grading essays	0.0	100.0	100.0
% within type of teacher	0.0	17.7	11.0
Total			
Count	76	124	200
% within experience grading essays	38.0	62.0	100.0
% within type of teacher	100.0	100.0	100.0

The 2 × 4 (Type of Teacher × Essay) ANOVA tested the hypotheses that ELA literacy assignments would be graded and judged differently by PE teachers compared to non-PE teachers. This series of tests examined whether PE teachers assigned different grades and directly assessed writing components of the essay differently than did non-PE teachers (Research Questions 2 and 3). Table 2 shows mean group differences and effect sizes for teacher type.

Table 2
Means and Effect Sizes for PE and Non-PE Teachers on Ratings of Work

Outcome	PE teachers		Non-PE teachers			Cohen's
	n	*M*	*n*	*M*	*SD*	*d*
Overall quality*	76	2.32	125	2.67	.87	.40
Overall grade	76	4.93	122	5.14	1.79	.12
Sentence structure*	76	2.08	124	2.47	.83	.47
Vocabulary/word choice	75	2.36	124	2.52	.75	.21
Organization	75	2.31	125	2.52	.89	.24
Creativity	74	2.22	125	2.26	.89	.05
Neatness*	73	2.73	124	3.27	.69	.78
Effort*	76	2.63	124	3.12	.58	.85
Remedial*	76	2.84	125	2.59	.71	.35

Note. Pooled within-group standard deviation in 2-way model.

*Statistically significant difference between PE and non-PE teachers.

A significant difference between teacher groups was found for ratings of the overall quality of the essay, $F(1, 193) = 7.98$, $p < .01$. Examination of the cell means showed that PE teachers assigned lower quality ratings to the essays, compared to non-PE teachers. The Cohen's *d* effect size was .40, indicating a moderate effect. Although PE teachers judged the essays to be lower in quality, the grade assigned to the essay was not significantly different for PE and non-PE teachers, $F(1, 190) = 0.61$, $p > .05$. On average, PE and non-PE teachers gave an average grade of C+ to the essays. Interactions of teacher type with essay were not significant for quality rating or essay grade.

In addition to these overall assessments, teachers judged essays on specific aspects of writing such as sentence structure, vocabulary/word choice, organization, and creativity. Essays graded by PE teachers were regarded as having worse sentence structure than those graded by non-PE teachers, $F(1, 192) = 10.89$, $p < .01$. A moderate effect size of .47 was found for the difference. There were no significant differences between PE and non-PE teachers on the other writing characteristics. Finally, all interactions of Teacher Type × Essay were nonsignificant, indicating that any assessment differences between

PE and non-PE teachers did not vary as a function of which essay they graded.

Beyond overall grading and structural ratings of the essays, teachers indicated the degree to which the student work was neat, perceived amount of effort the student put forth, and the extent to which they would recommend remedial assistance or tutoring in writing. Results from the ANOVAs showed statistically significant differences between PE and non-PE teachers on all three measures; there were no significant interactions of Teacher Type × Essay. Regardless of which essay was graded, PE teachers judged the handwritten work to be less neat than did non-PE teachers, with a large effect size of .78, $F(1, 189) = 28.15$, $p < .01$. PE teachers, compared to non-PE teachers, reported that the students put forth less effort on the assignment, $F(1, 192) = 33.69$, $p < .01$. This effect was also large ($d = .85$). Finally, PE teachers, compared to non-PE teachers, were more likely to recommend that students receive tutoring or remedial help with writing, $F(1, 193) = 5.94$, $p < .01$, $d = .35$.

Discussion

This study reports what experiences physical educators have with grading ELA literacy assignments and determines validity or reliability issues, compared with experiences of non-PE teachers. As a whole, PE teachers reported significantly lower levels of experience with grading these types of tasks, compared to non-PE teachers. More specifically, not a single physical educator in this study indicated a high degree of practice with grading ELA literacy projects. This is cause for concern given the trends in teacher evaluation where PE teachers are being held accountable for content they may not directly teach (Mercier et al., 2014; Rink, 2013; Seymour & Garrison, 2016).

The comparison of grading ELA literacy tasks between physical educators and non-PE teachers produced important findings. While PE and non-PE teachers generally graded essays with the same grade (C+), there were significant differences on their subjective judgments about the work. As a whole, physical educators were harsher in grading the overall quality of the essays. They judged the students' work to be of lesser quality, to have worse sentence structure, and to be less neat. Furthermore, physical educators were more likely to perceive that students put forth less effort on the assignment and more often would recommend remedial writing assistance.

These inconsistencies between teachers suggest that physical educators have little experience with grading ELA literacy assignments in PE and/or may find this challenging. It is unknown whether the critical appraisal among PE teachers is a more valid assessment of the ELA literacy task. That is, PE teachers may have a more accurate view of the work quality and effort than non-PE teachers, although this seems less plausible given PE teachers' self-reported lack of experience grading essays on the merits of writing. On the other hand, one may expect physical educators to be less critical because of their limited knowledge about literacy instruction and what is developmentally appropriate for a child. Similarly, as a result of their experience grading student writing assignments, non-PE teachers may be more gracious in their feedback. Future research comparing the grading of more extreme student writing exemplars (i.e., very high or low quality) between PE and non-PE teachers may reveal more about this issue. In addition, more research comparing the results of grading ELA literacy tasks among trained PE teachers, untrained PE teachers, and non-PE teachers would contribute more to this discussion.

In either case, the reading and writing levels of students within any classroom can vary significantly (Kozub & Hodge, 2014). This poses a challenge for classroom teachers who specialize in ELA pedagogical principles, let alone physical educators who do not specialize in ELA learning yet are being asked to support it. Therefore, PE teachers may have unrealistic expectations for ELA-oriented tasks in PE class, which may result in bias. For example, physical educators in this study were asked to grade components of writing such as sentence structure, vocabulary/word choice, organization, and creativity. However, little consideration was given to whether physical educators know how to identify student writing exemplars that are deficient in these traits (i.e., what is poor sentence structure?). Better support would clearly define these writing attributes for evaluation so that PE and non-PE teachers can be compared accordingly. This is important if the CCSS ELA literacy mandate is to continue, and it suggests that physical educators need more training on implementing and assessing ELA integrative tasks in PE class. This could include weekly journals, writing projects, and entry/exit slips in response to PE-related content. Future studies investigating how PE teachers can

work with their non-PE teacher counterparts in creating and evaluating some of these literacy tasks may be beneficial.

Overall, this study had limitations that may have affected generalizability. First, broader conclusions may not be possible given the small region in New York State where the study was conducted. Second, the online format of our study may have been less realistic in terms of grading writing samples. However, our examination of the online and in-person data revealed the same pattern of results for both formats of data collection.

In conclusion, as the current focus on the CCSS in educational reform endures, the need for physical educators to integrate ELA literacy tasks will likely continue. At the same time, the PE profession's limited selection of valid and reliable assessments other than fitness (Rink, 2013) in today's accountability era in public education means that physical educators may be evaluated on content that they do not teach (Mercier et al., 2014; Rink, 2013; Seymour & Garrison, 2016). While interdisciplinary instruction in PE is not a new approach, a focus on ELA literacy tied to the CCSS is a new concept that requires careful thought. For this initiative to be supported in an effective way, it appears that physical educators need to be trained on administering and evaluating ELA literacy tasks, not only for their own benefit, but also to enhance learning for their students.

References

Accountability Works. (2012). *National cost of aligning states and localities to the Common Core Standards*. Retrieved from Pioneer Institute website: http://www.accountabilityworks.org/photos/Cmmn_Cr_Cst_Stdy.Fin.2.22.12.pdf

Au, W. (2013). Coring Social Studies within corporate education reform: The Common Core State Standards, social justice, and the politics of knowledge in U.S. schools. *Critical Education*, *4*(5). Retrieved from http://ices.library.ubc.ca/index.php/criticaled/article/view/184278

Ballinger, D. A., & Deeney, T. A. (2006). Physical educators as teachers of literacy. *Journal of Physical Education, Recreation, & Dance, 77*(5), 18–23. https://doi.org/10.1080/07303084.2006.10597873

Collier, D. (2011). Increasing the value of physical education: The role of assessment. *Journal of Physical Education, Recreation, & Dance*, *82*(7), 38–41. https://doi.org/10.1080/07303084.2011.10598656

Common Core State Standards Initiative. (n.d.-a). About the standards. Retrieved June 21, 2017, from http://www.corestandards.org/about-the-standards/

Common Core State Standards Initiative. (n.d.-b). Development process. Retrieved June 21, 2017, from http://www.corestandards.org/about-the-standards/development-process/

Common Core State Standards Initiative. (n.d.-c). English language arts standards. Retrieved June 21, 2017, from http://www.corestandards.org/ELA-Literacy/

Conley, D. T. (2011). Building on the common core. *Educational Leadership*, *68*(6), 16–20.

Constantinou, P., & Wuest, D. A. (2015). Using academic language to level the playing field for English-language learners in physical education: Part 1. *Strategies*, *28*(5), 28–33. https://doi.org/10.1080/08924562.2015.1066285

James, A. R., & Bullock, K. (2015). Integrating the English language arts Common Core State Standards into physical education. *Journal of Physical Education, Recreation, and Dance*, *86*(3), 25. https://doi.org/10.1080/07303084.2014.998391

James, A. R., & Manson, M. (2015). *Physical education: A literacy-based approach*. Urbana, IL: Sagamore-Venture. https://doi.org/10.18666/TPE-2015-V72-I5-4271

James-Hassan, M. (2014). Common purposes: Using the Common Core State Standards to strengthen physical education instruction. *Strategies*, *27*(6), 8–12. https://doi.org/10.1080/08924562.2014.960123

Kozub, F. M., & Hodge, S. R. (2014). Considerations for written testing in physical education. *Journal of Physical Education, Recreation, and Dance*, *85*(6), 47–48. https://doi.org/10.1080/07303084.2014.927681

Lounsbery, M. A. F., & McKenzie, T. L. (2015). Physically literate and physically educated: A rose by any other name? *Journal of Sport and Health Science*, *4*(2), 139–144. https://doi.org/10.1016/j.jshs.2015.02.002

Magnotta, J., & Darst, P. (2015). The Common Core Learning Standards: Where does physical education fit in? *Journal of Physical Education, Recreation, and Dance*, *86*(2), 8–9. https://doi.org/10.1080/07303084.2015.988464

Marttinen, R. H. J., McLoughlin, G., Fredrick, R., & Novak, D. (2017). Integration and physical education: A review of research. *Quest*, *69*(1), 37–49. https://doi.org/10.1080/00336297.2016.1150864

Mathis, W. J. (2010). *The "Common Core" standards initiative: An effective reform tool?* Retrieved October 12, 2016, from National Education Policy Center website: https://nepc.colorado.edu/publication/common-core-standards

Mercier, K., & Doolittle, S. (2013). Assessing student achievement in physical education for teacher evaluation. *Journal of Physical Education, Recreation, and Dance*, *84*(3), 38–42. https://doi.org/10.1080/07303084.2013.767721

Mercier, K., Whitley, M. A., & Manson, M. (2014). Middle school assessments that contribute to literacy goals. *Strategies*, *27*(3), 22–27. https://doi.org/10.1080/08924562.2014.900464

Rink, J. E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport, 84*, 407–418. https://doi.org/10.1080/02701367.2013.844018

Saltman, K. J. (2014). The austerity school: Grit, character, and the privatization of public education. *Symploke*, *22*(1), 41–57. https://doi.org/10.5250/symploke.22.1-2.0041

Scrabis-Fletcher, K. A. (2016). Integrating Common Core content into physical education: Creating new knowledge connections. *Journal of Physical Education, Recreation, and Dance, 87*(7), 20–26. https://doi.org/10.1080/07303084.2016.1202797

Seymour, C., & Garrison, M. (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. https://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), 1–11. https://doi.org/10.22230/ijepl.2016v11n6a727