



Students' Attitudes Toward Assessment in Physical Education

Authors' contribution:

- A) conception and design of the study
- B) acquisition of data
- C) analysis and interpretation of data
- D) manuscript preparation
- E) obtaining funding

Efstathia Evangelou^{*B,D,C} , Nikolaos Digelidis^{A,D,C} ,
Charalampos Krommidas^{C,D} 

Department of Physical Education and Sport Sciences, University of Thessaly, Greece

Received: 18.07.2022

Accepted: 02.11.2022

***Correspondence:** Efstathia Evangelou, University of Thessaly, Department of Physical Education and Sport Sciences; 42100 Karies, Trikala, Greece; E-mail: efevangelou@uth.gr

Abstract

Despite the recent increase in attention on studies focusing on different types of assessment of student participation in physical education (PE), there is little research concerning students' attitudes toward the PE assessment process. This study aimed to investigate students' attitudes toward assessment procedures in PE by developing and validating a quantitative instrument concerning crucial areas of pedagogy such as curriculum, teachers, grading, and fitness tests. It further investigates important attitude determinants such as grade level and gender differences. The scale was distributed to 1,403 students in the 5th, 6th, 7th, and 10th grades. Principal components analysis showed an eigenvalue factor greater than 1, which accounted for 47% of the total variance. Confirmatory factor analysis yielded satisfactory indices that confirmed the unidimensional model with high internal consistency ($\alpha = .84$). A two-way analysis of variance revealed a significant difference between the 7th and 10th grades. At the same time, students expressed a high level of positive attitudes toward the different aspects of assessment procedures. The study examined various aspects of students' attitudes toward assessment and resulted in a reliable, valid, and easy-to-use quantitative instrument that can be used by educators and researchers to assess students' attitudes toward assessment in PE.

Key words: Assessment; physical education; reliability; validity

Introduction

Assessment is a crucial and indispensable part of the teaching and learning cycle (Chróinín & Cosgrave, 2013) that “influences the teaching and learning process and defines its product” (Tolgfors, 2018, p. 311). The term assessment has historically been criticized, as many use it to refer to grading, while others claim that assessment and grading do not encourage students to maintain an active lifestyle after finishing school (Leirhaug, 2016). Despite all of this, assessment is one of the three interrelated message systems that transfer formal knowledge in schools, the others being curriculum and pedagogy (Bernstein, 1975). Furthermore, assessment has been suggested to be the most powerful of the three (Chan et al., 2012). The assessment process should place children at the center and include them in decision-making to make assessment a meaningful experience that enhances their learning and motivation (Chróinín & Cosgrave, 2013; Tolgfors & Öhman, 2015). Recently, studies have increasingly focused on

different assessment types regarding student participation in PE (Lorente-Catalan & Kirk, 2014). However, there is little research concerning students' attitudes toward the PE assessment process. Leirhaug and Annerstedt (2016) underline students' greater involvement in assessment processes as a crucial area of development.

A range of assessment strategies have been formed for the needs of PE. On the one hand, traditional assessment with the use of physical fitness tests is a popular and mandatory part of the curriculum, with supporters arguing that it promotes a healthy lifestyle and physical activity and motivates and encourages goal-setting, self-monitoring, and self-testing skills while contributing to positive attitudes that increase cognitive and affective learning (Cale & Harris, 2009; Simonton et al., 2019). Despite their benefits, there is some controversy concerning the type, validity, and reliability of fitness tests and whether they promote healthy and active lifestyles (Cale & Harris, 2009; Simonton et al., 2019). On the other hand, alternative assessment forms such as self-assessment, peer assessment, teacher observation, event tasks, portfolios, checklists, essays, and oral discourses have attracted attention (Mintah, 2003). Quality assessment includes different strategies (López-Pastor et al., 2013) and focuses on assessment for learning that is authentic and integrated, valid, and socially just while contributing to quality curriculum and pedagogy (Hay & Penney, 2009). Transformative assessment is another critical pedagogical approach in the assessment process; it involves negotiation between the teacher and student (Tolgfors, 2019). The assessment practice can be adapted to different students' needs and circumstances through student influence, thus achieving "social justice" (Tolgfors, 2019).

A review of the recent literature reveals that most of the studies dealing with PE assessment examine teachers' perspectives through qualitative methods (Chróinín & Cosgrave, 2013; Leirhaug & MacPhail, 2015; Leirhaug et al., 2016; Svennberg et al., 2018). In contrast, quantitative methods have been used by only a limited number of studies to test teachers' attitudes toward fitness tests (Keating et al., 2008; Keating & Silverman, 2004). Teachers' perceptions of different assessment practices were examined using questionnaires developed and validated by the researchers (Mintah, 2003; Borghouts et al., 2017), with missing steps of validity.

There is limited research concerning students' views of PE assessment. On the one hand, qualitative methods were used to demonstrate students' beliefs about different assessment processes or grading in PE, understand students' thoughts on the researched subject, or reveal what different types of assessment promote (Leirhaug, 2016; Redelius & Hay, 2009; Tolgfors, 2018; MacPhail & Halbert, 2010). Researchers have used a range of qualitative methods such as interviews, semi-structured interviews, focus groups, and lesson observations (Redelius & Hay, 2009; Tolgfors, 2018; MacPhail & Halbert, 2010). Despite the remarkable value of assessment in pedagogy, it has been established that the assessment of students does not always serve the pursued outcomes, as accountability mechanisms influence the messages transmitted (Chan et al., 2011). Instead of having a pedagogical and educational orientation, they turn to value participation, effort, and behavior (Chan et al., 2011).

While the quantitative methodology has been the most common form used in education research, especially with Likert-type scale questionnaires (Phillips & Silverman, 2012), the use of quantitative questionnaires in research on students' perceptions of assessment in PE is also quite limited. A literature review detected quite a few studies that used a questionnaire. The researchers developed tools for the specific research purpose to explore students' perceptions of different assessment forms or attitudes toward fitness tests (Leirhaug, 2016; James et al., 2005; Mercier & Silverman, 2014). While the analyzed questionnaires offered several valuable findings, some of them did not follow all of the necessary validation steps.

It is important to understand students' attitudes as attitudes can mediate student achievements (Subramaniam & Silverman, 2000). Understanding what students think and feel toward assessment in PE can assist PE teachers in planning a more effective educational process, meaningful assessment, and curriculum for students. Although we can form a reasonable opinion of what teachers think of different assessment methods through modern research, the same does not apply to students' views.

Purpose of Research

The primary purpose of this study was to assess and investigate what students think and feel about assessment procedures in PE. Due to the lack of validated quantitative measurements, an instrument was developed and validated according to the rules of psychometry. Because the validity and reliability of the measurements could depend on factors such as the respondents' age and gender (Jansson et al., 2019), the second purpose was to investigate the suitability of the developed instrument according to grade level and gender differences and further explore potential developmental differences in students' attitudes toward assessment according to grade level and gender (Subramaniam & Silverman, 2000). The development of a new measurement to assess students' attitudes toward assessment in PE could provide a valuable methodological contribution to further studies (Jansson et al., 2019).

Methodology

Instrument Development

The first step in developing a tool for assessing students' perceptions of assessment in PE was to conduct an extensive literature review. In this manner, different assessment procedures in PE were identified, and the methods and instruments used in previous studies, as well as aspects of the assessment process in need of exploration, were detected (James et al., 2005; MacPhail & Halbert, 2010; Redelius & Hay, 2009; Phillips & Silverman, 2012; Subramaniam & Silverman, 2000). Then, sentences were formed about assessment concerning the pedagogical process of assessment and curriculum. Critical aspects considered were the teacher's role, the importance of grading, and fitness tests. An eight-item instrument emerged called "Students' attitudes toward assessment in PE (SATA-PE)"; it concerned the following themes: (a) curriculum, (b) teachers, (c) curriculum grading, (d) teacher grading, and (e) fitness testing (Table 3).

To examine the instrument's content validity, a panel of three research experts with more than 20 years of experience in behavior change, attitudes, and motivation who have published articles in peer-reviewed journals on questionnaire development and psychometrics reviewed the final items. It resulted that the scale's items referred to students' attitudes toward assessment in PE and were appropriate for both elementary and middle school. A 5-point Likert scale was chosen, ranging from 5 (strongly agree) to 1 (strongly disagree).

The SATA-PE scale was pilot tested on a convenience sample of 75 students from the 5th and 6th grades. This grade level group was chosen because more clarifications would be needed, more questions would arise, and the completion time would be longer than with the other grade levels. The purpose of this pilot study was to check any ambiguous or non-discriminant items, assess the clarity of the completion instructions, and estimate the required completion time. The pilot study sample was excluded from the main study as only the SATA-PE questionnaire was tested, and not all the necessary data were according to the original research design.

Reliability and Validity Study

In order to determine reliability and validity, the constructed scale was administered to students in the 5th, 6th, 7th, and 10th grades. The study was conducted in two waves to ensure a representative sample size. The initial sample comprised 638 students; after a year, a second wave followed with a sample of 765 students.

Participants

Twenty-two schools were selected from different municipal districts in Athens, the capital city of Greece, representing 80 classes, with a total of 1,403 students (730 boys and 684 girls) participating. The students were attending 5th or 6th grades ($n = 256$), the 7th ($n = 437$) and 10th ($n = 721$) grades. They were aged 10, 11, 12, or 15 years old ($M = 13.29$, $SD = 1.85$). Twenty-two schools were selected from different municipal districts in Athens, the capital city of Greece, representing 80 classes. The first author distributed the questionnaires to the students and informed them about the purpose of the study. The following instructions were provided: "In the following sentences, circle what best matches what you think of..." The response alternatives were explained. The students were further informed that there were no right or wrong answers, and that they were free to ask questions.

Ethical Clearance

The research received approval from the Institute for Educational Policy of Greece, and data collection was conducted after receiving approval from the University of Thessaly Institutional Research Ethics Committee. Participants were fully informed of the nature of the study, and their participation was voluntary; written consent was obtained from all participants and their parents/guardians.

Data Analysis

To determine the instrument's reliability and validity, the Statistical Package for Social Sciences (SPSS v.26) and AMOS v.18.0 were used in the following steps: (a) principal components analysis, which functions as an initial step in factor analysis as it reveals much about the number and nature of factors (Tabachnick & Fidell, 2001); (b) confirmatory factor analysis to confirm the factorial validity; and (c) Cronbach's alpha, inter-item correlation, and corrected item-total correlation to calculate the internal reliability consistency of items. Furthermore, the Pearson correlation coefficient was calculated to examine concurrent validity. Beyond that, a two-way analysis of variance (two-way ANOVA) was conducted to reveal significant differences between gender, grade level, and interaction

between gender and grade level. Finally, descriptive statistics were used to investigate students' answers and their perspectives on assessment in PE.

Results

Principal Components Analysis

Principal components analysis (PCA) was performed on the eight items of the SATA-PE tool to test conceptual variables. Principal components analysis is a reliable method for developing objective tests and reveals a factor solution (Tabachnick & Fidell, 2001). The results showed an eigenvalue factor greater than 1, which accounted for 47% of the total variance. That factor had eight items with high loads that indicate attitude toward assessment in the PE. This factor was called "Students' attitudes toward assessment in PE (SATA-PE)." Table 1 presents the item loadings and the communality values.

Confirmatory Factor Analysis

Confirmatory factor analysis was conducted on the eight-item scale to cross-validate PCA, confirming a uni-dimensional model. The following eight-item model yielded satisfactory indices ($\chi^2_{20} = 251.753$, $p < .001$, TLI = .907, CFI = .934, RMSEA = .091, RMSEA 90% CI = .081-.101). Figure 1 presents the item standardized regression weights.

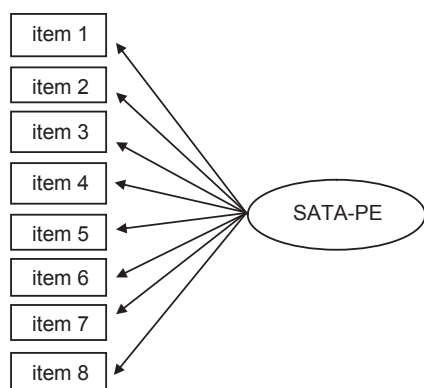


Figure 1. Confirmatory factor analysis (CFA) of SATA-PE, standardized regression weights

Reliability

Cronbach's alpha was computed to assess whether the data from the eight items of the SATA-PE scale formed a reliable scale. The alpha of the eight items was .84, indicating that the items form a scale with good internal consistency and reliability. Eight of the 10 items that make up the tool contribute positively and, if removed, reduce its degree of reliability. The reliability analysis results are presented in Table 1, with the questionnaire listed in Table 3.

Concurrent Validity

Concurrent validity was examined using a semantic differential scale for the first wave of the sample ($n = 638$). The scale used the question, "Generally, assessing our performance in a PE lesson is..." and responses were measured using a semantic differential format (Osgood et al., 1967) on five bipolar adjectives (good-bad, fair-unfair, useful-useless, harmful-beneficial, enjoyable-unenjoyable). All adjectives were rated on five-point semantic differential scales and achieved satisfactory internal reliability ($\alpha = 0.80$). Pearson product correlation was then computed and revealed that there was a significant relationship between the two scales ($r = .609$, $p < .01$), with a large effect size ($r^2 = .37$), which explained 37% of the variance.

Gender and Grade Level Differences

Two-way ANOVA was conducted to note gender and grade level differences in students' attitudes toward PE. The results indicated no significant interaction between gender and grade level effects on students' attitudes toward assessment in PE. No significant gender differences were noted, either. Significant differences were noted according to grade level with a small effect size $F(2,1397) = 4.26$, $p = .014$, partial $\eta^2 = .006$ (Table 2). The assumption of independent observations, homogeneity of variances, and normal distributions of the dependent variable for each group were checked and met. The post-hoc tests showed a significant difference between the 7th and 10th grades. There was no significant difference between elementary and high school. Table 3 provides the means and standard deviations for grade level by gender.

Students' Perspectives on Assessment in PE

From the percentages of students' answers (Table 4), it is concluded that the students generally had a positive attitude toward PE assessment since all of their answers agreed with a high rate. More specifically, concerning the curriculum, the students agreed (strongly agreed, 24.5%, or agreed, 31%) that the assessment process helped them

Table 1. Item loadings and reliability of the SATA-PE

Items	Factor Loading	Communality	Cronbach's α
	SATA-PE		
1	.74	.55	.84
2	.66	.44	
3	.78	.61	
4	.80	.64	
5	.53	.28	
6	.67	.45	
7	.67	.44	
8	.60	.36	
Eigenvalue	3.76		
% of variance	46.99		

Table 2. Two-way analysis of variance for students' attitudes toward assessment in PE as functions of grade level and gender

Variable	df	MS	F	p	η^2
Grade level	2	2.53	4.26	.014	.006
Gender	1	1.90	3.20	.074	.002
Grade level*Gender	2	.336	.567	.568	.001
Error	1397	.594			

Table 3. Means and standard deviations of the grade level separately by gender

Grade level	Boys			Girls		
	n	M	SD	n	M	SD
5 th -6 th	130	3.66	.65	126	3.64	.71
7 th	219	3.77	.81	218	3.63	.81
10 th	377	3.61	.78	333	3.53	.77

improve in PE lessons (skills, knowledge, behavior) and believed that the process of assessment in PE was important for learning (strongly agreed, 19.6%, or agreed, 31.5%). As far as the PE teacher was concerned, students felt that the PE teacher helped them improve through the assessment process (strongly agreed, 23.4%, or agreed, 35.6%); they were assessed according to whether they had learned what had been taught (strongly agreed, 16.7%, or agreed, 37.1%) and graded according to whether they had achieved their goals (strongly agreed, 25.7%, or agreed, 39.4%). Moreover, the students agreed with the highest rate that having a good grade in the PE lesson was important (strongly agreed, 36.2%, or agreed, 33.1%) and that being informed about the grading method in the lesson was helpful (strongly agreed, 22.7%, or agreed, 34.6%). Similarly, they agreed that they liked to test their abilities through fitness tests (strongly agreed, 34.2%, or agreed, 30.5%).

Table 4. Topics and items of the examined instrument, descriptive statistics (frequencies, percentages) of students' responses

Topics	Items	5	4	3	2	1
		n (%)	n (%)	n (%)	n (%)	n (%)
Curriculum	1. The assessment process helps me improve in PE (skills, knowledge, behavior).	344 (24.5)	435 (31)	380 (27.1)	137 (9.8)	107 (7.6)
Teacher	2. The PE teacher assesses whether I have learned what has been taught.	235 (16.7)	520 (37.1)	368 (26.2)	197 (14)	83 (5.9)
Curriculum	3. The process of assessment in PE is essential for learning.	275 (19.6)	442 (31.5)	402 (28.7)	184 (13.1)	100 (7.1)
Teacher	4. I feel that the PE teacher helps me improve through the assessment process.	328 (23.4)	500 (35.6)	283 (20.2)	201 (14.3)	91 (6.5)
Grading – curriculum	5. Having a good grade in PE is important.	508 (36.2)	465 (33.1)	280 (20)	94 (6.7)	55 (3.9)
Grading – curriculum	6. It is helpful to be informed about the manner of grading in PE.	318 (22.7)	486 (34.6)	382 (27.2)	148 (10.5)	69 (4.9)
Grading – teacher	7. The PE teacher grades me fairly on whether I have achieved my goals.	360 (25.7)	553 (39.4)	323 (23)	105 (7.5)	62 (4.4)
Fitness tests	8. I like testing my abilities through fitness tests.	480 (34.2)	428 (30.5)	291 (20.7)	117 (8.3)	87 (6.2)

Notes. Strongly agree = 5; Agree = 4; Not sure = 3; Disagree = 2; Strongly disagree = 1.

Discussion

This study's purpose was to assess students' attitudes toward assessment in PE through quantitative methods. Developing and validating a quantitative instrument assessing students' attitudes toward PE is one way to gain a quantitative measurement. According to Chan et al. (2011, p. 4), "work in this area is still relatively limited, and our knowledge in this area is still weakly conceptualized." This instrument was constructed to provide the whole picture of students' attitudes toward assessment procedures as it integrates crucial domains of pedagogy such as curriculum, grading, teachers, and fitness testing. The research concerned boys and girls of a broad age group consisting of elementary, middle, and high school grade levels. The scale's development and testing were guided through research consistent with attitude theory and psychometry, integrating the literature and empirical knowledge concerning students' attitudes toward PE assessment. The conceptual and methodological procedures appropriate for quantitative approaches were also used. Additionally, a suitable method design was chosen to develop an instrument that could address a wide age range of students who do not all necessarily have the same ability to express their feelings.

A psychometric assessment of the scale resulted in a short, easy-to-complete instrument. Since completion time is a consideration for researchers, teachers, and school administrators, this short, easy-to-complete instrument could be beneficial for them as it is not time-consuming (Mercier & Silverman, 2014). Moreover, the good internal consistency it exhibits may increase the accuracy of students' answers and enable regular use by teachers, resulting in a better understanding of what students think and feel about assessment methods and procedures in PE. This tool may help educators and researchers explore students' perceptions of PE assessment. A simple analysis of students' answers, as shown in Table 3, provides the feedback needed to highlight the points of the assessment process that students consider important, those that need better explanation or modification, to connect teaching with assessment.

It was observed that students in the 7th grade had a more positive attitude toward assessment procedures than students in other grades. Moreover, there were no statistically significant differences for students in the 5th and 6th or 10th grades, while their attitude toward assessment was not as positive as that of the 7th-grade students. This result

may depend on the summative function of assessment, the role of grades, and different assessment practices used by teachers (Borghouts et al., 2017). Furthermore, while students' attitudes became more positive from the 5th and 6th grades to the 7th grade, these attitudes seemed to decline sharply in the 10th grade. This decline may result from the decline in attitude toward PE lessons that occurs as the school level increases (Evangelou & Digelidis, 2018).

Concerning students' perceptions of assessment, this research reveals that the students considered the assessment process to be significant in their learning and helped them improve in PE lessons (skills, knowledge, behavior). Furthermore, students felt that the PE teacher assessed what had been taught, graded them according to whether they had achieved their goals, and helped them improve. Moreover, students thought that having a good grade in the PE lesson was crucial, while being informed about the grading criteria for the lesson was helpful. Finally, they appreciated having the opportunity to test their abilities through fitness tests.

A better balance between the curriculum and the assessment procedures, accurate grading decisions, and an understanding of student assessment consequences will strengthen the construct validity of the assessment processes (Redelius & Hay, 2009). Teachers should align grades with official syllabus expectations through task-specific criteria and clearly defined student implications (Redelius & Hay, 2009). According to the research, students who thought the teacher graded fairly were more assessment literate, as they were the ones to identify the elements that would reward them with better grades (Leirhaug, 2016).

As for gender, there were no significant differences between boys and girls, and gender did not influence students' attitudes toward assessment in PE. Consequently, the scale is not influenced by gender, and both genders can use it equally (Robbins et al., 2008). In other research, gender had minor statistical importance in predicting students' attitudes toward assessment (Leirhaug, 2016).

Regarding the strengths and limitations of the study, the scale was consistent with the theory of planned behavior, with a single factor to account for much of the variance in attitudinal responses (Ajzen, 2005). Furthermore, concurrent validity was examined through a question that was answered using a semantic differential format (Osgood et al., 1967) on five bipolar adjectives, which achieved satisfactory internal reliability and a significant relationship between the two scales, with a large effect size ($r^2 = .37$). Furthermore, using a sample from several municipal districts with different social and economic characteristics strengthens the generalization of the findings. In addition, the sample was large, with an equal number of male and female participants.

However, the total number of elementary students was smaller than the number of students in other grades, and the conduct of an elicitation study could emerge or confirm the items. Additionally, we cannot know whether the anonymity of the questionnaire ensured the truthfulness of the participants' responses.

As assessment is one crucial aspect of teaching and learning procedures, students' attitudes toward assessment form another critical aspect when considering students as essential factors in the educational process. For this reason, students' perspectives should be considered for any assessment procedure (Tolgfors, 2018). To make PE a meaningful and enjoyable experience, PE teachers should mobilize all the means provided.

Conclusion

This research contributes insight into students' attitudes toward assessment procedures in PE while investigating grade and gender differences. Furthermore, it provides educators and researchers with a reliable, valid, easy-to-use quantitative instrument to assess students' attitudes toward assessment in PE. Using this instrument, they can better understand what students think and feel about assessment procedures. With this understanding, teachers can modify their teaching and curriculum to make assessment a meaningful experience that enhances students' learning and enjoyment of the lesson. Additional research can further confirm the validation and reliability of the scale. Finally, the SATA-PE scale can be used in other science disciplines to collect information about students' attitudes toward assessment procedures by replacing the word "Physical Education (PE)" in the items presented.

Ethics approval and informed consent

The research received approval from the Institute for Educational Policy of Greece. Participants were fully informed of the study's nature, and their participation was voluntary, with written consent being obtained from all participants and their parents.

Competing interests

The authors of a paper have no competing interest or any commercial associations or financial interests to declare.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). Berkshire, England: Open University Press.
- Bernstein, B. (1975). *Class, codes and control: Towards a theory of educational transmission*. (Vol. 3). London: Routledge and Kegan Paul.
- Borghouts, B. L., Slingerland, M., & Haerens, L. (2017). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473-489. <https://doi.org/10.1080/17408989.2016.1241226>
- Cale, L., & Harris, J. (2009). Fitness testing in physical education – A misdirected effort in promoting healthy lifestyles and physical activity? *Physical Education and Sport Pedagogy*, 14(1), 89-108. <http://dx.doi.org/10.1080/17408980701345782>
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in physical education. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3-18. <https://doi.org/10.1080/18377122.2011.9730340>
- Chróinin, N. D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18(2), 219-233. <http://dx.doi.org/10.1080/17408989.2012.666787>
- Evangelou, E., & Digelidis, N. (2018). Students' attitudes and predispositions toward physical education in Greece. *Journal of Physical Education and Sport*, 18(3), 1624-1631. <http://doi.org/10.7752/jpes.2018.03238>
- Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389-405. <https://doi.org/10.1177/1356336X09364294>
- James, A., Griffin, L. L., & France, T. (2005). Perceptions of assessment in elementary physical education: A case study. *The Physical Educator*, 62(2), 85-95. <http://hdl.handle.net/20.500.12648/2455>
- Jansson, A., Sundblad, B. G., Lundvall, S., & Norberg, J. (2019). Assessing students' perceived learning and contentment in physical education: A scale development study and structural equation modeling analysis. *Measurement in Physical Education and Exercise Science*, 23(3), 280-290. <https://doi.org/10.1080/1091367X.2019.1617713>
- Keating, D. X., Guan, J., Ferguson, H. R., Chen, L., & Bridges, D. M. (2008). Physical education teacher attitudes toward fitness tests scale: Cross-revalidation and modification. *Measurement in Physical Education and Exercise Science*, 12(2), 72-87. <https://doi.org/10.1080/10913670801903969>
- Keating, D. X., & Silverman, S. (2004). Teachers' use of fitness tests in school-based physical education programs. *Measurement in Physical Education and Exercise Science*, 8(3), 145-165. https://doi.org/10.1207/s15327841mpe-e0803_2
- Leirhaug, E. P. (2016). Exploring the relationship between student grades and assessment for learning in Norwegian physical education. *European Physical Education Review*, 22(3), 298-314. <https://doi.org/10.1177/1356336X15606473>
- Leirhaug, E. P., & Annerstedt, C. (2016). Assessing with new eyes? Assessment for learning in Norwegian physical education. *Physical Education and Sport Pedagogy*, 2(6), 616-631. <https://doi.org/10.1080/17408989.2015.1095871>
- Leirhaug, E. P., & MacPhail, A. (2015). "It's the other assessment that is the key": Three Norwegian physical education teachers' engagement (or not) with assessment for learning. *Sport, Education and Society* 20(5), 624-640. <https://doi.org/10.1080/13573322.2014.975113>
- Leirhaug, E. P., MacPhail, A., & Annerstedt, C. (2016). "The grade alone provides no learning": Investigating assessment literacy among Norwegian physical education teachers. *Asia-Pacific Journal of Health, Sport and Physical Education*, 7(1), 21-36. <https://doi.org/10.1080/18377122.2016.1145429>
- López-Pastor, M. V., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society* 18(1), 57-76. <https://doi.org/10.1080/13573322.2012.713860>
- Lorente-Catalan, E., & Kirk, D. (2014). Making the case for democratic assessment practices within a critical pedagogy of physical education teacher education. *European Physical Education Review*, 20(1), 104-119. <https://doi.org/10.1177/1356336X13496004>
- MacPhail, A., & Halbert, J. (2010). "We had to do intelligent thinking during recent PE": Students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy & Practice*, 17(1), 23-39. <https://doi.org/10.1080/09695940903565412>

- Mercier, K., & Silverman, S. (2014). Validation of an instrument to measure high school students' attitudes toward fitness testing. *Research Quarterly for Exercise and Sport*, 85(1), 81-89. <https://doi.org/10.1080/02701367.2013.872221>
- Mintah, K. J. (2003). Authentic assessment in physical education: Prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science*, 7(3), 161-174. https://doi.org/10.1207/S15327841MPEE0703_03
- Osgood, C. E., Suci, G. J., & Tannebaum, P. H. (1967). *The measurement of meaning* (2nd ed.). Champaign, IL: University of Illinois Press.
- Phillips, R. S., & Silverman, S. (2012). Development of an instrument to assess fourth and fifth grade students' attitudes toward physical education. *Measurement in Physical Education and Exercise Science*, 16(4), 316-327. <http://dx.doi.org/10.1123/jtpe.2014-0022>
- Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15(3), 275-294. <https://doi.org/10.1177/1356336X09364719>
- Redelius, K., & Hay, J. P. (2012). Student views on criterion-referenced assessment and grading in Swedish physical education. *Physical Education and Sport Pedagogy*, 17(2), 211-225. <https://doi.org/10.1080/17408989.2010.548064>
- Robbins, B. L., Wu, Y. T., Sikorskii, A., & Morley, B. (2008). Psychometric assessment of the adolescent physical activity perceived benefits and barriers scales. *Journal of Nursing Measurement*, 16(2), 98-112. <https://doi.org/10.1891/1061-3749.16.2.98>
- Simonton, L. K., Mercier, K., & Garn, C. A. (2019). Do fitness test performances predict students' attitudes and emotions toward physical education? *Physical Education and Sport Pedagogy*, 24(6), 549-564. <https://doi.org/10.1080/17408989.2019.1628932>
- Subramaniam, P. R., & Silverman, S. (2000). Validation of scores from an instrument assessing student attitude toward physical education. *Measurement in Physical Education and Exercise Science*, 4(1), 29-43. https://doi.org/10.1207/S15327841Mpee0401_4
- Svennberg, L., Meckbach, J., & Redelius, K. (2018). Swedish PE teachers struggle with assessment in a criterion-referenced grading system. *Sport, Education and Society*, 23(4), 381-393. <https://doi.org/10.1080/13573322.2016.1200025>
- Tabachnick, G. B., & Fidell, S. L. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Tolgfors, B. (2018). Different versions of assessment for learning in the subject of physical education. *Physical Education and Sport Pedagogy*, 23(3), 311-327. <https://doi.org/10.1080/17408989.2018.1429589>
- Tolgfors, B. (2019). Transformative assessment in physical education. *European Physical Education Review*, 25(4), 1211-1225. <https://doi.org/10.1177/1356336X18814863>
- Tolgfors, B., & Öhman, M. (2016). The implications of assessment for learning in physical education and health. *European Physical Education Review*, 22(2), 150-166. <https://doi.org/10.1177/1356336X15595006>

This is Open Access article distributed under the terms of CC-BY-NC-ND 4.0 International License.