

ASSESSMENT

Accuracy of a Peer Process Assessment Performed by Elementary Physical Education Students

Andrew E. Alstot

Abstract

The practice of peer assessment has been widely recommended for use in physical education (PE) classes. However, it is unknown how accurately students in lower elementary grades can assess peers' performance. Therefore, this study examined the accuracy at which students in lower elementary PE perform peer assessments of a throwing skill. Students in first, second, and third grades conducted assessments of peers' overhand throw performance once per week for 4 weeks. Sessions were video recorded; a researcher viewed the videos and examined participants' assessments for accuracy. Results indicate that students in lower elementary grades can assess peers' performance with a relatively high degree of accuracy, with third graders performing significantly better than their younger counterparts. Results also show that students can accurately perform a peer process assessment immediately after a reasonably short assessment training session. It can be concluded that physical education teachers can implement peer process assessments with young students with the confidence that they will be conducted accurately.

Andrew E Alstot is an associate professor, Department of Kinesiology, Azusa Pacific University. Please send author correspondence to aalstot@apu.edu

Physical education (PE) students benefit greatly from receiving fast and accurate feedback on motor performance (Fredenburg, Lee, & Solmon, 2001; Silverman, Tyson, & Krampitz, 1992); this augmented feedback maximizes learning and student achievement within the PE environment. However, when class sizes grow exceedingly large, it becomes nearly impossible for teachers to provide this feedback consistently to all students directly. One suggested solution to this problem is the use of peer assessments (Kniffin & Baert, 2015; Lund & Veal, 2013; Veal, 1995). Peer assessments where students observe a partner's performance of a motor skill and make judgments on the quality of that performance can allow for all students in a class to receive augmented feedback on their skill behavior, thus maximizing the opportunity for motor learning within the PE context. However, despite the recommendations for the use of peer assessments, little evidence indicates the ability of students in elementary PE to correctly complete an assessment of their peers' performance of a motor skill and thereby provide accurate augmented feedback. Therefore, this study examined the accuracy at which lower elementary students can perform a peer process assessment in PE.

Benefits of and Recommendations for Using Assessments in Physical Education

Using assessments within PE has a variety of potential benefits. Formalizing the assessment process demonstrates to students that the content within PE has inherent value (Kniffin & Baert, 2015; Reeves, 1986), thereby improving the likelihood that students will understand the importance of the activities being assessed (Kniffin & Baert, 2015) and possibly increasing their motivation for engagement in PE activities (Tan & Wright, 2004). This process also allows students to receive important feedback about their performance (Reeves, 1986) in the short and long term (Veal, 1988), aiding in the facilitation of learning within a PE class (DeJong, Kokinakis, & Kuntzleman, 2002; Veal, 1992). Assessments can also provide evidence to school administrators and parents regarding the educational process that occurs in PE (Reeves, 1986). Finally, teachers can evaluate the data gathered from the assessment process to make informed judgments about teaching effectiveness (Fencl, 2014); the main role of a teacher is to facilitate learning, and the only way to

know if learning has occurred in a PE class is for assessments to be successfully administered (Kniffin & Baert, 2015). The use of assessments within PE has many potential benefits for students, teachers, and administrators.

The most fundamental purpose of using assessments within PE, however, should be to help students learn (DeJong et al., 2002; Veal, 1992). Teachers can link assessments to their instruction (Veal, 1992) with the purposes of helping students achieve the national standards for PE (DeJong et al., 2002); if this is done effectively, ultimately students will be learning and achieving within the PE environment. Fencil (2014) suggested using multiple forms of assessment within a class; as students have different learning styles, using various forms of assessments can help teachers to target this variety of learning styles, thereby increasing the potential for learning and performance of PE students.

To accomplish this, teachers can implement published assessments in their classes, such as PE Metrics (Fisette & Franck, 2012), or create their own (Lund & Veal, 2013). Veal (1995) suggested a variety of ways that teachers can implement the assessment process into PE to measure an assortment of psychomotor skills: (a) assessments can be performed by the teacher, by the students themselves, or by student peers; (b) assessments can measure the product of performing motor skills, such as the number of shots made in basketball, long-jump distance, or mile run time; and/or (c) assessments can evaluate the process of performing a skill (Veal, 1995), such as the technique the participant uses to perform a golf swing or an overhand throw in softball.

Veal (1992) also distinguished between formative and summative assessment—formative assessments occur concurrent with instruction, are ongoing and recurring, and focus on the process of student learning, while summative assessments usually occur at the end of an instructional unit, assessing how much a student has learned, and are typically associated with grading and evaluation. Both formative and summative assessments have important roles within PE instruction, but formative assessments, because of their effective use as tools that aid in student learning, are recommended for more frequent use. Regardless of whether teachers implement teacher-, self-, or peer-directed assessments, whether the product or process of

engaging in motor skills is measured, and whether the tool is used for formative or summative purposes, the assessments used in PE classes should have a high degree of validity (Hay & Penney, 2009).

Validity of Psychomotor Assessments

Several published assessments have gone through laborious validation processes. For example, the Team Sport Assessment Procedure (Grehaigne, Godbout, & Bouthier, 1997; Nadeau, Richard, & Godbout, 2008) and the Game Performance Assessment Instrument (Oslin, Mitchell, & Griffin, 1998) have been shown to have a high degree of validity. The psychometric properties of the PE Metrics series of assessments (Society of Health and Physical Educators, 2010) have been rigorously examined as well (Fox et al., 2011). Additionally, the Fitnessgram test battery (Meredith & Welk, 2007) has been investigated and shown to hold quality psychometric properties (Morrow, Martin, & Jackson, 2010). Teachers are encouraged to introduce these and other published assessments into their PE classes, to provide valuable feedback and aid in the facilitation of learning.

Some studies have also examined the accuracy of assessments completed by student peers, a method suggested by Veal (1995). Hill and Miller (1997) examined the precision at which fifth grade PE students assessed their peers when performing the curl-ups, push-ups, back-saver sit and reach, and trunk lift tests of the Fitnessgram and found that the student peers performed the assessments accurately. Conversely, Patterson, Bennington, and De La Rosa (2001) examined the psychometric properties of the curl-up component of the Fitnessgram and found that students aged 10 to 12 tended to overestimate the amount of correct repetitions performed by peers.

Despite these minor discrepancies in the ability of peers to assess curl-up performance, other researchers have discovered that students can accurately perform a variety of peer assessments of additional skills. Kolovelonis and Goudas (2012) found that fifth and sixth grade PE students accurately conducted peer assessments of the product and process of performing a basketball chest pass. Similarly, Nadeau et al. (2008) examined the accuracy at which 14- to 17-year-old ice-hockey players assessed peers during game play and found they did so with a high degree of accuracy. Ward, Crouch, and colleagues (Crouch, Ward, & Patrick, 1997; Ward, Crouch, &

Patrick, 1998; Ward, Smith, Makasci, & Crouch, 1998) conducted a series of studies examining the effects of a peer-mediated accountability intervention in PE and its effects on student performance; one of the components of this tool included students conducting a peer assessment. It was discovered that children in Grades 4 and 5 can accurately assess the process and product of performing volleyball (Crouch et al., 1997; Ward, Crouch, & Patrick, 1998) and basketball skills (Ward, Smith, et al., 1998).

Finally, only one study examined the accuracy of peer assessments with younger elementary PE students. Alstot (2015) looked at the use of a token economy for reinforcing overhand throw performance; a part of this intervention required second grade participants to peer-assess throwing technique. These students performed peer process assessments of two components of the overhand throw with a high degree of accuracy. These studies provide evidence that elementary students can accurately provide feedback to peers via an assessment on not only the product of performing a motor skill, but also the process.

Despite the recorded levels of validity of several psychomotor assessment instruments (Grehaigne et al., 1997; Morrow et al., 2010; Nadeau et al., 2008; Oslin et al., 1998), and despite the evidence supporting the ability of upper elementary students to perform forms of assessments within PE accurately (Hill & Miller, 1997; Kolovelonis & Goudas, 2012; Patterson et al., 2001), little is known regarding how accurately younger elementary students can engage in the practice of peer assessments and provide precise feedback on the process of performing a motor skill. Veal (1995) suggested that students of any age can be taught to perform peer assessments properly, but it is unknown how accurately they can do so. Alstot (2015) initially investigated the accuracy of second graders' peer assessments of overhand throw performance, but it was a minor variable in that study and necessitated further inquiry. Therefore, this study expands on these findings and examines Veal's suggestions, by investigating the accuracy at which first, second, and third grade PE students can perform a peer process assessment of overhand throw technique, an elementary skill appropriate and recommended for lower elementary-aged children (Pangrazi & Beighle, 2010).

Method

Participants and Setting

Thirty-eight PE students (19 male, 19 female) from in-tact first, second, and third grade classes at a private elementary school located in the Pacific Northwest were selected as participants. Upon approval from the university institutional review board, informed consent was obtained from participants' parent(s) or guardian(s) and assent was obtained from the participants. Data were collected during regularly scheduled PE classes in a covered outside area on the school grounds. Students in each class who were not subjects of the study performed the same overhand throw and assessment tasks as participants; these tasks were conducted as part of their regular PE classes, despite data not being collected on these students.

Data Collection and Equipment

For each grade level, sessions were video recorded; a digital video camera was set up on a tripod and arranged to capture video of each participant's overhand throw performance. Additional equipment included the assessment sheets peers used to analyze correct and incorrect performance of the first two critical components of the overhand throw performed by their partners (see Figure 1). Pencils were also provided for participants to complete the assessment sheets and bean bags were used for partners to throw during the overhand throw trials.

Procedures

The researchers met with each grade level during their regularly scheduled PE classes once per week for 5 weeks. During the first week, following assessment demonstration guidelines suggested by Johnson (2004), the students were trained to use the peer assessment sheet, which included analysis on the first two components of the overhand throw (see Figure 1): (a) "place the throwing-arm side of the body away from the target" (Pangrazi & Beighle, 2010, p. 674)—*side to target* and (b) step toward the target with the foot opposite the throwing hand" (Pangrazi & Beighle, 2010, p. 674)—*step toward target with opposite foot*. These aspects of overhand throwing skill are characteristic of more proficient performance (Haywood & Getchell,

Your Number _____

Thrower's Number _____

Your partner will throw the bean bag 5 times as far as he or she can. Every time your partner throws the bean bag, you will make sure he or she is throwing the right way.

Put an X next to each part of the skill your partner does correctly. Leave it blank if he or she does not do that part of the skill.


	CUES	Throw 1	Throw 2	Throw 3	Throw 4	Throw 5
	Side to target					
	Step toward target with opposite foot					

Figure 1. Peer process assessment participants used to analyze two critical components of the overhand throw skill: (a) “place the throwing-arm side of the body away from the target” (Pangrazi & Beighle, 2010, p. 674)—*side to target* and (b) step toward the target with the foot opposite the throwing hand” (Pangrazi & Beighle, 2010, p. 674)—*step toward target with opposite foot*.

2014) and are a typical focus of elementary PE. This assessment training took approximately 20 min and consisted of a researcher modeling correct and incorrect technique of the overhand throw and discussing how to appropriately complete the peer assessment based on the observed performance (i.e., students were to view their partner's performance of five trials of the overhand throw and, after each trial, write an X in the appropriate box if the corresponding skill

component was performed correctly, while leaving it blank if the component was not performed correctly). Students then observed the researcher perform several overhand throw trials and completed an assessment of his performance. The students' sheets were then examined by the researchers to ensure students were appropriately checking the boxes on the assessment.

During the second through fifth weeks, participants were placed in pairs. Partner A was given instructions to complete five trials of an overhand throw, attempting each time to throw a bean bag as far as he or she could while trying to maintain correct technique. During these five trials, Partner B completed the peer assessment. After the five trials and assessment sheet were completed, the partners switched roles; Partner B then completed five overhand throw trials while Partner A assessed the performance. This allowed each participant to complete five overhand throw trials and an assessment of a peer. For each session, a digital video camera was placed in the corner of the instructional area and recorded each participant's performance. Each session took approximately 10 min and was conducted at the beginning of the students' regularly scheduled PE class.

Data Analysis

After each session, the videos were returned to the university lab for analysis. Videos were viewed in slow motion while the participants' performances of the overhand throw were assessed with the same assessment sheets used by the peers. These researcher-conducted assessments served as the criterion for comparison; the researchers then compared the peer assessments to the criterion assessments to check for accuracy of each trial and then recorded them as correct or incorrect. Mean percentage of correctly assessed components and standard deviations were calculated for each grade for Weeks 1 to 4 and a combined score for Grades 1 to 3 across all 4 weeks. Further analyses included one-way ANOVAs that examined differences in assessment accuracy across grades, differences between Weeks 1 to 4 for all participants, and differences between Weeks 1 to 4 within each grade level. SPSS 21 was used for data analyses.

Interobserver Agreement (IOA)

To ensure the criterion assessment data were analyzed appropriately, a trained second observer randomly selected and viewed 25% of the sessions. The second observer's scores were then compared to the criterion assessments and analyzed for agreement. Percentage agreement was calculated by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100. Overall agreement was 96.1%, while agreement across first, second, and third grades was 96.0%, 94.6%, and 97.3%, respectively, indicating the criterion assessments were initially analyzed appropriately.

Results

Across all 4 weeks of data collection, the accuracy with which participants assessed their peers' performance for first, second, and third grade students was 79%, 77%, and 92%, respectively. When comparing assessment accuracy across grade levels, significant differences were discovered, $F(2, 143) = 9.182, p < .001$ (Figure 2). Post hoc analyses found differences between third grade ($M = 91.77\%$, $SD = 13.97$) and first grade ($M = 78.95$, $SD = 22.76$; $p = .005$) and between third grade and second grade ($M = 77.17$, $SD = 22.28$; $p < .001$). These results indicate that the third grade students conducted the assessment with a significantly higher degree of accuracy than their younger counterparts.

However, one-way ANOVA results indicated no significant differences in the participants' accuracy across time, $F(3, 142) = 0.392, p = .759$, indicating the participants' accuracy did not change from week to week (Figure 3). There was also no difference in first grade, $F(3, 34) = 0.327, p = .806$, second grade, $F(3, 42) = 0.112, p = .953$, or third grade, $F(3, 58) = 1.814, p = .155$, across weeks, indicating each grade level remained consistent in its accuracy over time. Regardless of grade level, participants' assessment accuracy showed no improvement from week to week, which demonstrates the students were as accurate in their assessment practices during the first week after training as they were during the fourth week.

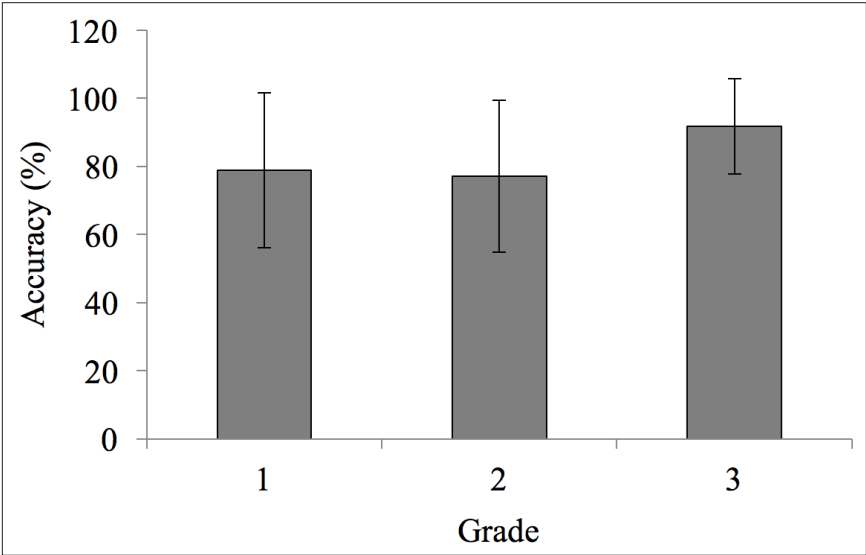


Figure 2. Accuracy of completed assessments by grade level across all 4 weeks. Grade 3 participants assessed peers' performance with a significantly higher degree of accuracy than the first and second grade participants. Error bars indicate \pm standard deviation.

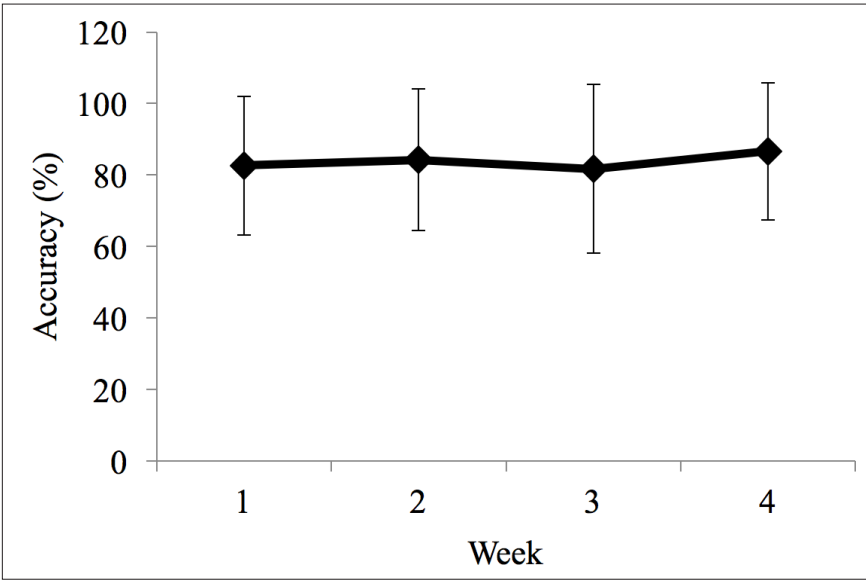


Figure 3. Accuracy of completed assessments for all participants (Grades 1 to 3, $N = 38$) across 4 weeks. Participants' completed assessment accuracy did not significantly change across time. Error bars indicate \pm standard deviation.

Discussion

The main objective of this study was to examine the accuracy at which younger elementary-aged PE students perform a process assessment of their peers' overhand throw performance. These younger students assessed their peers with a relatively high degree of precision, with each grade level exceeding 75% accuracy. The third grade students, however, performed the assessments with a significantly higher degree of accuracy (i.e., greater than 90%) compared to their younger counterparts. These results indicate that, although not as accurate as older children, students as young as first grade can peer-assess the process of performing a motor skill; PE teachers can implement these types of assessments and be confident that the feedback partners are receiving is relatively accurate. Another finding of the study showed that students neither improved nor regressed in their assessment accuracy across time. This indicates that students could immediately and accurately implement the peer process assessment after only one short training session. Elementary PE teachers can introduce assessments to their students and be assured that they can immediately assess their peers with a relatively high degree of accuracy.

Veal (1995) suggested that students of any age can be trained to engage in the assessment process, including peer assessments. Although some studies have shown this was correct for some elementary grades (Hill & Miller, 1997; Kolovelonis & Goudas, 2012; Patterson et al., 2001), little was known regarding how true Veal's assertion held for younger elementary PE students. The results of this study, though, support Veal's claim and Alstot's (2015) findings with second grade PE students; first and second grade students performed the assessment with a relatively high degree of accuracy. However, it is not known if, for participants in this study, the precision of the assessments was high enough; despite students' assessments reaching greater than 75% accuracy in Grades 1 and 2 and greater than 90% in Grade 3, a suggested minimum criterion for acceptable assessment accuracy in PE was not found in the literature, so the results cannot be compared to a standard. The author of this study suggests an agreement criterion similar to that accepted for measuring behavioral data: 80% (Cooper, Heron, & Heward, 2007). If this is used as the standard for comparison, third grade participants easily

surpassed the minimum criterion, but the first and second grade students did not; however, the younger students were within a few percentage points of reaching the benchmark. Third grade students assessed their peers with a high degree of accuracy, while first and second graders assessed their peers with a relatively high degree of accuracy, nearly reaching the 80% criterion.

Johnson (2004) suggested several steps for teachers to demonstrate to students in PE how to conduct an assessment, including demonstrating how to perform the skill, how to observe peers' performance, how to record observations, and how to provide appropriate feedback to peers based on the assessment. During Week 1 of this study, the researchers followed similar steps in training participants how to properly complete the peer assessment. After only one 20-min training session, the students could fully engage in the assessment process. These specific overhand throw peer assessments were included in the students' regularly scheduled PE classes once per week for 4 weeks following the assessment training session. In the week immediately following the training session, students in all three grades could immediately conduct the peer assessment with a relatively high degree of accuracy and maintained consistent performance throughout the remaining 3 weeks. The assessment accuracy did not change over time; students engaged in the process with the same level of accuracy during the first week as they did the last. These results indicate that there may be no lag in younger students' learning of how to perform simple process assessments. First, second, and third grade students could immediately understand how to conduct the assessment properly and maintained similar engagement throughout the study.

This study had two main limitations, and suggestions for future research are provided. First, the assessment used with the students was relatively simple, only assessing two components of the overhand throw, *side to target* and *step toward target with opposite foot*, while ignoring the other critical components of a proficient throw. It was designed intentionally with younger students' cognitive development in mind. Based on the results of the study though, the simplistic nature of the assessment may have been appropriate for the younger students (i.e., the first and second graders whose accuracy was 79% and 77%, respectively), while the third graders,

whose accuracy exceeded 90%, could have successfully completed a more complex assessment, inclusive of the remaining components of proficient overhand throw. Future research should examine how more complex process assessments are performed by younger elementary PE students, investigating various levels of complexity in assessments.

The second limitation was associated with the sample used in this study. Students were selected from a private elementary school, and although no screening was conducted for participants' socio-economic status (SES), reading comprehension, or other cognitive or social components, it was assumed that the students who participated were likely at a middle or high SES with appropriate levels of reading comprehension, representing a relatively homogeneous subset of the population. Future studies should focus on implementing peer process assessments with a wider sample of first, second, and third graders, particularly concentrating on public schools in lower SES areas. A future study could examine the influence that moderating variables, such as SES, reading comprehension levels, and cognitive development, have on assessment accuracy in younger elementary PE students.

Conclusion

Unfortunately, the use of assessments in PE lacks regularity (Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013). Perhaps the lack of widespread use is due to, in part, an unawareness of the ability of students to conduct these assessments properly. Regardless, assessments are “. . . an integral and necessary aspect of education across all subject areas of the school curriculum, and physical educators can no longer afford to be ambivalent about this practice . . .” (Lopez-Pastor et al., 2013, p. 73). PE teachers should embrace the many benefits of the assessment process, not the least of which is that assessments help facilitate student feedback, of which the importance for student learning is well documented (Fredenburg et al., 2001; Silverman et al., 1992). When teachers are significantly outnumbered by students, the feedback process becomes more complicated; as suggested by Kniffin and Baert (2015), Lund and Veal (2013), and Veal (1995), having students engage in peer assessments can help to alleviate some of these complications. Students can, through the peer assessment process, provide immediate and

accurate feedback; all students will be engaged cognitively and receive augmented feedback on their performance, which thus maximizes the learning opportunities within PE. Teachers can implement peer process assessments with students as young as first grade with the confidence the assessments will be performed accurately, providing immediate feedback to all students in the class and maximizing the opportunity for learning and student achievement in PE.

References

- Alstot, A. E. (2015). Effects of a peer-administered token economy on second grade physical education students' overhand throw performance. *Physical Educator*, *72*, 245–265.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Crouch, D. W., Ward, P., & Patrick, C. A. (1997). The effects of peer-mediated accountability on task accomplishment during volleyball drills in elementary physical education. *Journal of Teaching in Physical Education*, *17*, 26–39. <https://doi.org/10.1123/jtpe.17.1.26>
- DeJong, G., Kokinakis, C. L., & Kuntzleman, C. (2002). The role of assessment in meeting the NASPE physical education content standards. *Journal of Physical Education, Recreation, and Dance*, *73*, 22–25. <https://doi.org/10.1080/07303084.2002.10607842>
- Fencl, M. J. (2014). Fun and creative unit assessment ideas for all students in physical education. *Journal of Physical Education, Recreation, and Dance*, *85*, 16–21. <https://doi.org/10.1080/07303084.2014.855589>
- Fisette, J. L., & Franck, D. M. (2012). How teachers can use PE metrics for formative assessment. *Journal of Physical Education, Recreation, and Dance*, *83*, 23–34. <https://doi.org/10.1080/07303084.2012.10598775>
- Fox, C., Zhu, W., Park, Y., Fisette, J. L., Graber, K. C., Dyson, D., . . . Raynes, D. (2011). Related critical psychometric issues and their resolutions during development of PE Metrics. *Measurement in Physical Education and Exercise Science*, *15*, 138–154. <https://doi.org/10.1080/1091367X.2011.568367>

- Fredenburg, K. B., Lee, A. M., & Solmon, M. (2001). The effects of augmented feedback on students' perceptions and performance. *Research Quarterly for Exercise and Sport*, 72, 232–242. <https://doi.org/10.1080/02701367.2001.10608956>
- Grehaigne, J. F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education*, 16, 500–516. <https://doi.org/10.1123/jtpe.16.4.500>
- Hay, P., & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15, 389–405. <https://doi.org/10.1177/1356336X09364294>
- Haywood, K. M., & Getchell, N. (2014). *Life span motor development* (6th ed.). Champaign, IL: Human Kinetics.
- Hill, G. M., & Miller, T. A. (1997). A comparison of peer and teacher assessment of students' physical fitness performance. *Physical Educator*, 54, 40–46.
- Johnson, R. (2004). Assessments in physical education. *Journal of Physical Education, Recreation, and Dance*, 75, 33–40. <https://doi.org/10.1080/07303084.2004.10607287>
- Kniffin, K. M., & Baert, H. (2015). Maximizing learning through assessment in middle and high school physical education. *Journal of Physical Education, Recreation, and Dance*, 86, 7–16. <https://doi.org/10.1080/07303084.2015.1009203>
- Kolovelonis, A., & Goudas, M. (2012). Students' recording accuracy in the reciprocal and the self-check teaching styles in physical education. *Educational Research and Evaluation*, 18, 733–747. <https://doi.org/10.1080/13803611.2012.724938>
- Lopez-Pastor, V. M., Kirk, D., Lorente-Catalan, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education, and Society*, 18, 57–76. <https://doi.org/10.1080/13573322.2012.713860>
- Lund, J., & Veal, M. L. (2013). *Assessment-driven instruction in physical education: A standards-based approach to promoting and documenting learning*. Champaign, IL: Human Kinetics.
- Meredith, M. D., & Welk, G. J. (Eds.). (2007). *Fitnessgram/Activitygram test administration manual* (4th ed.). Champaign, IL: Human Kinetics.

- Morrow, J. R., Martin, S. B., & Jackson, A. W. (2010). Reliability and validity of the FITNESSGRAM®: Quality of teacher-collected health-related fitness surveillance data. *Research Quarterly for Exercise and Sport*, 81, S24–S30. <https://doi.org/10.1080/02701367.2010.10599691>
- Nadeau, L., Richard, J. F., & Godbout, P. (2008). The validity and reliability of a performance assessment procedure in ice hockey. *Physical Education and Sport Pedagogy*, 13, 65–83. <https://doi.org/10.1080/17408980701444718>
- Oslin, J. L., Mitchell, S. A., & Griffin, L. L. (1998). The Game Performance Assessment Instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education*, 17, 231–243. <https://doi.org/10.1123/jtpe.17.2.231>
- Pangrazi, R. P., & Beighle, A. (2010). *Dynamic physical education for elementary school children* (16th ed.). San Francisco, CA: Benjamin Cummings.
- Patterson, P., Bennington, J., & De La Rosa, T. (2001). Psychometric properties of child- and teacher-reported curl-up scores in children ages 10–12 years. *Research Quarterly for Exercise and Sport*, 72, 117–124. <https://doi.org/10.1080/02701367.2001.10608941>
- Reeves, B. (1986). Assessment in physical education. *Bulletin of Physical Education*, 22, 29–31.
- Silverman, S., Tyson, L., & Krampitz, J. (1992). Teacher feedback and achievement in physical education: Interaction with student practice. *Teaching and Teacher Education*, 8, 333–344. [https://doi.org/10.1016/0742-051X\(92\)90060-G](https://doi.org/10.1016/0742-051X(92)90060-G)
- Society of Health and Physical Educators. (2010). *PE Metrics: Assessing national standards 1–6 in elementary school* (2nd ed.). Champaign, IL: Human Kinetics.
- Tan, S. K., & Wright, S. (2004). Authentic assessment to enhance student learning in primary school physical education. *Journal of Physical Education and Recreation*, 10, 36–44.
- Veal, M. L. (1992). The role of assessment in secondary physical education – A pedagogical view. *Journal of Physical Education, Recreation, and Dance*, 63, 88–92. <https://doi.org/10.1080/07303084.1992.10609932>
- Veal, M. L. (1995). Assessment as an instructional tool. *Strategies*, 8, 10–15. <https://doi.org/10.1080/08924562.1995.10592045>

- Veal, M. L. (1988). Pupil assessment issues: A teacher educator's perspective. *Quest*, 40, 151–161. <https://doi.org/10.1080/00336297.1988.10483896>
- Ward, P., Crouch, D. W., & Patrick, C. A. (1998). Effects of peer-mediated accountability on opportunities to respond and correct skill performance by elementary school children in physical education. *Journal of Behavioral Education*, 8, 103–114. <https://doi.org/10.1023/A:1022868808642>
- Ward, P., Smith, S. L., Makasci, K., & Crouch, D. W. (1998). Differential effects of peer-mediated accountability on task accomplishment in elementary physical education. *Journal of Teaching in Physical Education*, 17, 442–452. <https://doi.org/10.1123/jtpe.17.4.442>