

The Lieberman–Brian Inclusion Rating Scale for Physical Education

European Physical Education Review
1–14

© The Author(s) 2017

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1356336X17733595

journals.sagepub.com/home/epe



Lauren Lieberman

The College at Brockport, State University of New York, USA

Ali Brian

University of South Carolina, USA

Michelle Grenier

University of New Hampshire, USA

Abstract

Children with disabilities have mixed feelings about their inclusion experiences. The purpose of this study was to explore the validity and reliability of results from the Lieberman-Brian Inclusion Rating Scale for Physical Education. Experts in adapted physical education ($n = 10$) established content and face validity (mean = 4.68 ± 0.56) on a five-point scale through three rounds of evaluation through the Delphi method. Next, elementary physical education teachers ($n = 15$) and two independent raters established test–retest reliability ($r = 0.87$, $p < 0.001$; intra-class correlations (ICC) = 0.93 , $p < 0.001$) and inter-rater reliability ($r = 0.69$, $p < 0.001$; ICC = 0.82 , $p < 0.001$) respectively. Physical education teachers among others can use this valid and reliable scale to determine the extent to which teachers attempt to make an environment inclusive. Practical uses for this instrument are program evaluation, intervention research, and as a teaching tool.

Keywords

Adapted physical education, children with disabilities, professional preparation, teacher assessment, special education

Introduction

United Nations (UN) policies affirm the rights of all children, including those with a disability, to be valued equally, treated with respect, provided with equal opportunities, and experience full and effective participation and inclusion in society. UN policies for all children include the UN

Corresponding author:

Ali Brian, Department of Physical Education, University of South Carolina, 1300 Wheat Street, Columbia, SC 29208, USA.

Email: brianalib@mailbox.sc.edu

Convention on the Rights of the Child (1989), the UN Rules for the Equalization of Opportunities for Persons with Disabilities (1993), the UNESCO Salamanca Statement (1994), and the UN Convention on the Rights of Persons with Disabilities, in which Article 30 specifically refers to physical activity as a right for all children within educational institutions (United Nations International Children's Emergency Fund, 2007).

Although there is no single recommended form of inclusive practice (Dyson, 1999), practices should follow an "ecology of inclusion" which promotes responsibility of schools to be accountable regarding environmental factors and classroom practices (Dyson et al., 2004). Curricular and teaching adaptations, instructional supports, and environmental conditions are considered by teachers and planners so that students can reach their maximum potential in cognitive, affective, and social domains.

In many cases, inclusion is guided by policies and practices that address educational standards aligned with cultural values (Meegan and MacPhail, 2006). These policies are prioritized internationally based on the individual country's values, concepts, and approaches to education (Norwich, 2008). Within European countries for example, inclusion can be expressed in several ways (Fitzgerald, 2006). The United Kingdom (UK) requires that all students (with and without disabilities) have a balanced curriculum, which includes physical education through the National Curriculum (Department for Education and Employment/Qualifications and Curriculum Authority, 1999). However, the way in which inclusive practices occur can vary greatly. Heiman (2004) found that placements for students with disabilities in the UK and Israel consisted of a number of placement options including partial instruction in special education classrooms, full inclusion within general education classrooms, or total separation from mainstream classes. This suggests that inclusionary practices may be aligned with the concept of integration rather than inclusion (Avramidis and Norwich, 2002; Bredahl, 2013). However, policies for inclusion are distinctly different in developing countries, where inclusive education may have variable outcomes (Lloyd, 2013).

In the United States, general physical education (GPE) teachers are called upon to provide physical education for all students, kindergarten-12, within inclusive environments (United States Department of Education, 2003, 2005). When inclusion is viewed as a school-wide practice, students with disabilities who attend their neighborhood schools should be receiving educational services in the general education setting (Hunt and McDonnell, 2007). Unfortunately, most pre-service physical education teacher education programs provide minimal coursework regarding evidence-based supports for GPE teachers to create inclusive environments for all students (Beamer and Yun, 2014; Haegele and Sutherland, 2015). As a result, many GPE teachers feel ill-prepared to teach high quality physical education that includes all students with and without disabilities (Block and Obrusikova, 2007; Hodge et al., 2004; LaMaster et al., 1998).

Many teachers feel ill-prepared despite an emergent literature base supporting several recommendations to foster inclusion and maximization of student learning for all students within a GPE setting (Doulkeridou et al., 2010; Qi and Ha, 2012). These recommendations include a collaborative approach between peer tutors, paraprofessional support, and general and special educators, featuring cooperative learning, and differentiated instruction (Grenier, 2006; Klavina and Block, 2008; Valentini and Rudisill, 2004). Despite various definitions and positions regarding inclusion, children with disabilities in many countries are often included within GPE in a manner that does not promote maximization of student learning (Asbjørnslett and Hemmingsson, 2008; Block and Obrusnikova, 2007; Coates and Vickerman, 2008; De Schipper et al., 2017; Hutzler et al., 2002; Perkins et al., 2013). As a result, adapting activities for students with disabilities to feel

as though they contribute to the success of the class is also essential in creating an inclusive environment (Spencer-Cavaliere and Watkinson, 2010).

Further examples of quality GPE strategies that can maximize student learning for all students include: reduced wait time and few lines; enough equipment for all students; using smaller sided games in shorter dimension; and avoiding activities that place students “on display” (Brian et al., 2014; Rink, 2014; Weaver et al., 2013). Other recommendations such as not placing numbers of repetitions but using time as a criterion (e.g. not running laps, but completing as many laps as one can within three minutes), using various students with and without disabilities and regardless of sex to provide demonstrations, providing ample and wide varieties of equipment choices made available to all students, and making sure all students are present for the set induction and closure can also help foster an inclusive environment that does not exclude anyone (Grenier, 2013; Lieberman and Houston-Wilson, 2018). If considered, the above recommendations may lead to an inclusive environment promoting maximization of student learning outcomes for all children in GPE (Haegele and Sutherland, 2015).

While there is a growing body of literature on inclusion and best practices associated with inclusive education, research is needed that explores effective inclusive practices within the general education setting, which may illuminate behaviors and practices that support inclusion and the overall experiences of students with disabilities (Qi and Ha, 2012). However, to date no observation tool exists to provide GPE teachers with feedback evaluating the extent to which GPE classes are inclusive for all children. Therefore, the purpose of this study was to explore the validity and reliability of results from the Lieberman–Brian Inclusion Rating Scale for Physical Education (LIRSPE) which was developed from the literature providing evidence-based best practices in inclusion in GPE.

Methods

The methods for this study occurred across two phases. Phase I included item construction and face and content validity. Phase II secured the reliability for the LIRSPE.

Phase I

Item construction and alignment with inclusion literature. All the items within the LIRSPE are in alignment with evidence-based best practices from within the global literature regarding inclusion (Figure 1). For example, using the paraeducator to support instruction (items 10–13; Figure 2), providing students with choices/autonomy supportive climates (items 7–8; Figure 2), and basing instruction upon one’s individual needs (e.g. items 3–6; Figure 2) all align with evidence-based best practices within the inclusion literature. Along with constructing items that align with the evidence-based best practices it is also important to show that experts outside of the research team provide the extent to which they agree that each item possesses content and face validity. Content validity refers to the extent to which a test accurately represents all facets of what it is purported to measure (Ary et al., 2010). Face validity is a form of content validity assessing the suitability of an instrument for the particular subject matter (Ary et al., 2010). Expert opinion is regularly operationalized via Delphi techniques (Barnett et al., 2015).

Phase I participants. To secure the items within the LIRSPE and assess their face validity a panel of experts in adapted physical education (APE) participated in a Delphi process (Figure 1). Experts

| Column one: These are sample items sent to Delphi raters | Column two: This is the actual rating scale | Column three was for general feedback | Column four represents where Delphi raters were asked to provide a judgement for the extent to which the item in column one aligned with the concept presented in the column below (1=no alignment; 5=excellent) |
|---|---|---------------------------------------|--|
| <p>Introduction</p> <ul style="list-style-type: none"> The children with disabilities are sitting/standing with their peers and included in the instructions of the introduction | <p>1 2 3 4 5 N/A</p> | | <p>“promoting a sense of belonging / not excluded” 1 2 3 4 5</p> |
| <p>Warm-up</p> <ul style="list-style-type: none"> The class does the warm-up together at the same time at their individualized pace thus not excluding children with lesser skills (The idea here is about duration not frequency such as children run as many laps as they can in X mins vs. requiring X laps in X mins. When a class is held to a standard number such as five laps or 20 push-ups some children may take much longer than others so everyone starts and finishes together.) | <p>1 2 3 4 5 N/A</p> | | <p>“promoting a sense of belonging / not excluded” 1 2 3 4 5</p> |
| <p>Speed of Play Within the Lesson</p> <ul style="list-style-type: none"> Speed of play is based upon present level of performance of all children including children with disabilities so as not to leave anyone behind (Such as during a dance unit the line dance can be executed in two counts, four counts or six counts for each move) | <p>1 2 3 4 5 N/A</p> | | <p>“meeting needs/interests of students” 1 2 3 4 5</p> |
| <p>Differentiated Instruction</p> <ul style="list-style-type: none"> Instruction is provided that allows for all students to succeed and benefit within the general program capturing different learning styles (Audio, visual, kinesthetic, and approaches specific to the needs of the child such as tactile modeling) The lesson provides a variety of choices to execute skills including all levels of the class. (This may be done in stations, within the task, and task-to-task) | <p>1 2 3 4 5 N/A</p> | | <p>“meets needs/interests of students” 1 2 3 4 5</p> |

(continued)

(continued)

| Column one: These are sample items sent to Delphi raters | Column two: This is the actual rating scale | Column three was for general feedback | Column four represents where Delphi raters were asked to provide a judgement for the extent to which the item in column one aligned with the concept presented in the column below (1=no alignment; 5=excellent) | | | | | | | | | | | | |
|--|---|---------------------------------------|--|---|---|---|-----|---|---|---|---|---|-----|--|--|
| Autonomy Supported Instruction <ul style="list-style-type: none"> ● Student has opportunities to make some choices driving his/her own learning ● Proper accommodations and supports are available for her or his choices ● (This may require some pre-teaching so children know what choices are available and comfortable for them) | <table border="0"> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> <tr> <td>4</td> <td>5</td> <td>N/A</td> </tr> <tr> <td>1</td> <td>2</td> <td>3</td> </tr> <tr> <td>4</td> <td>5</td> <td>N/A</td> </tr> </table> | 1 | 2 | 3 | 4 | 5 | N/A | 1 | 2 | 3 | 4 | 5 | N/A | | “promoting a sense of autonomy” 1 2 3 4 5 |
| 1 | 2 | 3 | | | | | | | | | | | | | |
| 4 | 5 | N/A | | | | | | | | | | | | | |
| 1 | 2 | 3 | | | | | | | | | | | | | |
| 4 | 5 | N/A | | | | | | | | | | | | | |

Figure 1. Sample items from the Lieberman–Brian Inclusion Rating Scale for Physical Education (LIRSPE) sent to Delphi raters.

Note: not all items are present due to space constraints. Column three was for raters to provide feedback. Column four was for raters to provide their rating for each item. Columns three and four are not a part of the actual LIRSPE.

($n = 10$) consisted of two groups of participants (in-service APE/GPE teachers ($n = 2$) and professors of APE in teacher preparation programs ($n = 8$)). All in-service APE/GPE teachers possessed Master’s degrees and over 10 years of teaching experience. APE and GPE teachers provided a practical opinion and a viewpoint speaking to the ecological validity of the items (e.g., does this really happen and is this realistic). Professors in APE teacher education programs are all faculty who have published research regarding inclusion, possess a doctorate degree in physical education with emphasis on APE, and possess over five years of experience.

Phase I design and procedures. Phase I featured a descriptive–analytic methodological design. The Institutional Review Board (IRB) of the co-lead author’s (author two) institution approved all procedures. Experts (10 respondents out of approximately 25 original contacts), solicited via email, provided consent and agreed to provide qualitative feedback on any portion of the LIRSPE. Additionally, experts provided ratings for the extent to which each LIRSPE item aligned with inclusive practices suggested from the inclusion literature (scale = 1–5; 1 = no alignment, 5 = high alignment; Figure 1). Upon receipt, the research team implemented the feedback from all experts and then sent back the revised instrument to the same experts for additional feedback and clarity on each item. This process continued for a total of three cycles (approximately three months, one month per cycle depending upon expert response time and the extent of detail within each comment) with the same experts to ensure all comments were addressed and reviewers were highly satisfied with all adjustments.

| Descriptor | Score | Comments |
|--|------------------|----------|
| Start of Class | | |
| 1. When the general physical education teacher welcomes the children into the gymnasium all of the children in the class are together including the children with disabilities (Children with disabilities do not walk into the gymnasium late). | 1 2 3 4 5 N/A | |
| Introduction | | |
| 2. Children with disabilities are sitting/standing with their peers and included in the instructions of the introduction. | 1 2 3 4 5 N/A | |
| Warm-up | | |
| 3. The class does the warm-up together with children performing at their own pace (For example, children run as many laps as they can in X mins vs. requiring X laps in X mins). | 1 2 3 4 5 N/A | |
| Speed of Play Within the Lesson | | |
| 4. Speed of play is varied based upon present level of performance of all children including children with disabilities so as not to leave anyone behind (Examples include: volleyball - players use a beach ball to slow down the speed of the game; floor hockey - players use a Frisbee instead of a ball or puck; softball - hit ball off a tee; or basketball - eliminating the five second rule). | 1 2 3 4 5 N/A | |
| Differentiated Instruction | | |
| 5. Instruction is provided that allows for all students to succeed and benefit within the general program by accommodating different learning styles (Audio, visual, kinesthetic, and approaches specific to the needs of the child such as tactile modeling). | 1 2 3 4 5 N/A | |
| 6. The lesson provides a variety of choices to execute skills. (This may be done in stations, within the task, and task-to-task.) | 1 2 3 4 5 N/A | |
| Autonomy Supported Instruction | | |
| 7. Student has opportunities to make some choices driving his/her own learning. | 1 2 3 4 5 N/A | |
| 8. Proper accommodations and supports are available for her or his choices (This may require some pre-teaching so children know what choices are available and may be comfortable for them to use.) | 1 2 3 4 5 N/A | |
| Demonstrations | | |
| 9. Use various members of the class including children with disabilities to demonstrate skills to the class (Only when you know they can demonstrate the skill being taught and that they would enjoy doing so). | 1 2 3 4 5 N/A | |
| Use of Paraeducator | | |
| 10. Support staff assists the child in learning as needed. | 1 2 3 4 5 N/A | |
| 11. Lesson is provided to the paraeducator before the class and explains their role throughout the lesson. | 1 2 3 4 5 N/A | |

(continued)

(continued)

| Descriptor | Score | Comments |
|---|------------------|----------|
| Peer-partner (when possible) | | |
| 12. When using partners the student with a disability has opportunities to partner with a same-aged peer (if appropriate) and not only the paraeducator when possible. | 1 2 3 4 5 N/A | |
| 13. Paraeducator encourages social interactions with peers in the class when possible. | 1 2 3 4 5 N/A | |
| Skill/Activity-partner Activity | | |
| 14. Teacher plans ahead to organize and manage partners effectively. | 1 2 3 4 5 N/A | |
| 15. Teacher ensures that children with disabilities have a partner. | 1 2 3 4 5 N/A | |
| Game/Activity-team Sport | | |
| 16. Students do not pick teams. | 1 2 3 4 5 N/A | |
| 17. Teacher avoids elimination games. | 1 2 3 4 5 N/A | |
| 18. Teacher avoids students waiting in line. | 1 2 3 4 5 N/A | |
| 19. Teacher distributes as much equipment as possible to maximize opportunities to respond. | 1 2 3 4 5 N/A | |
| 20. Teacher maximizes opportunities to respond and engagement time for all students by modifying the organization and rules of the game. (For example: using smaller sided games like 3v3 instead of 11 v 11 or allowing two bounces a side for volleyball.) | 1 2 3 4 5 N/A | |
| Equipment | | |
| 21. There is a range of equipment to meet the learning needs of all the students in the class. (For example: in a striking unit, the child could use a foam paddle, badminton racquet, flat bat, or tennis racquet.) | 1 2 3 4 5 N/A | |
| Environment | | |
| 22. Noise and distractions are reduced to maximize success. | 1 2 3 4 5 N/A | |
| Assessment | | |
| 23. When assessing the class, children with disabilities are assessed alongside their peers and modifications are provided as needed. (For example: a child with a disability may do wall push-ups or sit-ups on a wedge mat yet they are still being assessed along with his/her peers.) | 1 2 3 4 5 N/A | |
| Assessment Scores | | |
| 24. When children with disabilities are assessed, the scores count at least for their baseline of performance. (For example, when the teacher is collecting scores from the class she will always record the performance of the child with the disabilities at the same time to ensure inclusion and show that their performance and scores matter. If a child who uses a wheelchair is batting using the Test of Gross Motor Development their performance will be recorded and counted to measure present level of performance). | 1 2 3 4 5 N/A | |

(continued)

(continued)

| Descriptor | Score | Comments |
|---|------------------|----------|
| Skill-related Feedback | | |
| 25. Feedback on skill performance is given throughout the class to all children when possible. | 1 2 3 4 5 N/A | |
| 26. Feedback in regard to skill performance is positive general and/or positive specific feedback with the use of first names. (Children are held to a high standard and not just going through the motions of the performance. The teacher shows that they care about achievement and learning and not just participation.) | 1 2 3 4 5 N/A | |
| Closure | | |
| 27. The whole class is together and present when the teacher presents the closure/warm down of the class. | 1 2 3 4 5 N/A | |
| 28. Teacher checks for understanding of all children during closure. | 1 2 3 4 5 N/A | |
| Mean Overall Score: | | |
| Sum from each item used above / Total number of items used (excluding n/a) = inclusion rating related to effort the teacher makes to include all children | | |
| For example: 11 items received a total score of 47 / 11 items (excluding 17 that coded as n/a) = 4.27. | | |

Figure 2. The final version of the Lieberman–Brian Inclusion Rating Scale for Physical Education.

Phase II

Phase II represented the reliability portion of this study and included test–retest reliability, and inter-rater reliability. Test–retest reliability is a measure of stability across multiple time points (Nevill and Atkinson, 1997). Conducting test–retest reliability within 5–7 days is a common standard of practice within psychometric evaluation studies (Nevill and Atkinson, 1997). Intra-rater reliability refers to the extent to which ratings are consistent across raters (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). Intra-rater reliability as opposed to inter-observer agreement is the preferred method when raters are providing observation across a scale (e.g. 1–5 Likert scale; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). Inter-observer agreement is preferred when measurements are categorical and the observer needs to decide in which each category the observation fits (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). Given our raters decided the extent to which the teachers performed each portion of the LIRSPE via rating scale (e.g. 1–5 Likert scale), inter-rater reliability was the preferred method and was assessed via correlational coefficients.

Phase II participants. Elementary physical education teachers (K-5; $n = 15$; ages 22–26 years = 10; ages 27–40 years = 5) comprised the participants within phase II. All participants were GPE teachers (years of experience ranging from 1–10; African American women = 3, African American men = 4, Caucasian men = 8) whose class make ups included students (Grades

kindergarten-5; ages 6–12 years, 55% boys, 45% girls, Caucasian = 60%, African American = 20%, Hispanic = 15%, other = 5%) with (25%) and without disabilities (75%). Overall, the participants in Phase II represent a diverse sample that matches census data for urban locations within the United States (United States Census Bureau, 2017).

Phase II design and procedures. Phase II featured a descriptive correlational design. The IRB of the co-lead author's (author two) institution approved all procedures within Phase II. Parents of all students enrolled within each class provided informed consent while students provided assent. All teachers provided written consent. Each physical education teacher taught three lessons of an educational games unit for a total of 45 lessons (15 teachers \times three lessons each = 45 total lessons). Members of the research team digitally recorded all 45 lessons and each teacher wore a wireless microphone. First, we asked each teacher to rate his/her lesson by filling out the LIRSPE the same night of the actual teaching performance from his/her own video. Next, each teacher filled out the LIRSPE again for the same initial lesson five days later to perform test–retest reliability (from the original video recorded from his/her lesson). Concurrently, two trained independent raters systematically observed each lesson (100%) to provide a measure of inter-rater reliability during the live teach. Inter-rater reliability occurred from the original video rating by the teacher with the two independent raters. The independent raters were experts on inclusion and APE teachers, recruited from pre-existing contacts with the lead authors.

Data analyses

Phases I and II included descriptive and correlational analyses. The descriptive analyses within phase I include reporting results of the Delphi method via means and standard deviations. For phase II, three separate Pearson product–moment (PPM) correlations and intra-class correlations (ICC) examined the test–retest reliability and inter-rater reliability for the LIRSPE. The interpretations for PPM correlations include 0.1–0.29 = small effect, 0.30–0.49 = moderate effect, and 0.50 and above = a large effect (Field, 2009). Measures of strength and magnitude for each ICC include < 0.40 = poor reliability, $0.40 - < 0.74$ = adequate reliability, and $ICC > 0.75$ = excellent reliability (Field, 2009). The ICC calculations included two-way mixed effects models for consistency. The level of significance was set *a priori* at $p \leq 0.05$ and all analyses were conducted via SPSS (version 22) and Microsoft Excel for Macintosh.

Results

Phase I: Content and face validity

All items were rated above a four (out of five) except for set induction (mean (M) = 3.84; standard deviation (SD) = 1.41) and teacher avoids elimination games (M = 3.86; SD = 1.68). The preliminary version of the LIRSPE (seen in Figure 1) was highly rated by the experts (M = 4.41; SD = 0.49). Major critiques included issues of clarity (providing concrete examples) and parsimony (keeping wording short), while minor critiques centered on alignment. Despite receiving high ratings, feedback was addressed and a revised version of the LIRSPE was re-sent to the experts with the same instructions as the first round.

During round two, ratings and feedback received from experts (n = 10) were mostly positive with only one item scoring below a four (set induction, M = 3.89; SD = 1.36). However, the overall mean (M = 4.38; SD = 0.41) lowered from the first round (-0.03) due to six reviewers

providing ratings lower than the first review. During round two, feedback and major critique centered on clarity and parsimony, with minor feedback around alignment. For example, the item referring to warm up now includes a concrete scenario (e.g. numbers of laps completed in X time versus doing X laps) beneath the description of the actual warm-up item for clarity. As a result of the feedback and ratings compiled from the experts, changes were made to the LIRSPE. Subsequently, experts were then asked to again rate and provide feedback.

Results of the ratings from round three were very positive. Experts required minimal changes centered primarily on clarity. No individual item rated below a mean of 4.37, and the overall mean of the LIRSPE was very high ($M = 4.68$; $SD = 0.56$). Despite the high rating, the authors made the final changes recommended in round three to create the current version of the LIRSPE. The current version of the LIRSPE is located in Figure 2.

Phase II: Reliability

A total of 40 out of the originally planned 45 lessons were included. Five individual lessons were dropped due to a lack of completion of the data. Lack of completion of data varied from a failure to perform the evaluation at five days ($n = 2$), camera failure ($n = 1$), and failure to submit results ($n = 2$). Results of the PPM correlations confirmed both test–retest reliability (ratings range = 3.25–4.92; $r = 0.87$, $p < 0.001$) and inter-rater reliability (ratings range = 2.80–4.89; $r = 0.69$, $p < 0.001$) for the LIRSPE. ICCs corroborated the PPM correlations for test–retest reliability (ICC = 0.93, $F(39) = 13.74$, $p < 0.001$, 95 confidence interval (CI) = 0.86–0.96) and for inter-rater reliability (ICC = 0.82, $F(39) = 5.46$, $p < 0.001$, 95 CI = 0.65–0.90). See Figure 2 for the complete rating scale.

Discussion

The purpose of this study was to examine the validity and reliability of the results from a preliminary investigation of the LIRSPE. To the researchers' knowledge, this is the first tool aimed at measuring the extent to which inclusion is promoted within a GPE environment. Therefore, the development of the LIRSPE fills a much-needed gap within the literature.

The validity results were not surprising given that each item was based upon best practices recommended from experts within the GPE and APE literature base (e.g. the experts within the Delphi portion rated each item very high (greater than four) and all agreed that each item represents best practices contributing towards an inclusive environment). Unfortunately, there was no pre-existing theory from which to develop a model to test as a form of internal validity. However, given the strength of the results regarding face/content validity and low variance among reviewers, future research should explore whether an existing factor structure is present.

Not only did the LIRSPE results reveal face and content validity, the results also were reliable. Reliability was tested through test–retest and inter-rater reliability. Measuring reliability in multiple ways was a strength to this study. Multiple raters viewed each lesson and demonstrated strong inter-rater reliability. This finding is important because observation tools are rarely self-completed as outside raters typically use observation tools. Therefore, it was important to show that the LIRSPE was reliable across multiple outside raters so that when evaluated, teachers can be assured of reliable results.

Given that outside raters typically complete observation tools, one may be enticed to use LIRSPE results in a manner which may be viewed as misuse. The purpose of the LIRSPE was not

to punish teachers who demonstrate lower scores. Rather, the LIRSPE was initially developed for the purpose of promoting professional development through increased awareness and reflection on one's inclusive practices.

In measuring the ecological validity of the LIRSPE as a reflection tool, the teachers completed the LIRSPE directly after their lessons and five days later via video. Using five days as the time period between each assessment of the same lesson could be viewed as a limitation. Memory recall could interfere with accuracy. However, five to seven days between observations is a common practice for test-retest reliability (Nevill and Atkinson, 1997). In addition, the teachers' self-ratings across five days were as strong as the inter-rater reliability coefficients for the outside raters. These multiple reliability observations are a strength to this study.

Overall, the LIRSPE was originally developed with items that represent evidence-based best practices from the inclusion literature. Experts in APE and GPE provided content and face validity for the LIRSPE corroborating that the items represent best practices. Many teachers feel ill-prepared to include students with disabilities in their classrooms (Block and Obrusikova, 2007; Hodge et al., 2004; LaMaster et al., 1998). If teachers wanted to improve those feelings and assess the extent to which they deliver inclusive GPE content to diverse students they can feel confident that the LIRSPE is capable of generating valid results. For example, if teachers include all children during the warm-up, differentiate instruction, and support autonomy within their classes then their classes would align with recommendations from the literature (e.g. Grenier, 2006; Valentini and Rudisill, 2004) and scores from the LIRSPE would be high. However, if teachers do not sometimes use students with disabilities in demonstrations, include them with assessment scores, or try to maximize the use of peer and paraeducator support then scores from the LIRSPE would be low and in alignment with the literature inferring that more effort is needed to foster inclusion (e.g. Grenier, 2006; Klavina and Block, 2008). Thus, the results from this study revealed internal validity, and did so consistently and in a reliable manner.

Future research and implications

Despite the strengths inherent within the results (e.g. strong validity and reliability scores), the extent to which the LIRSPE scores correlated with students' ratings of inclusion remains unknown. Examining students' ratings along with teachers' ratings of inclusion could occur through a future research project but can also be a practical application for the LIRSPE. If teachers provide GPE opportunities in a manner which yields high scores on the LIRSPE it is inferred that students could experience feeling included in the class. If both teachers and students show high or low marks on the same items then the LIRSPE would demonstrate construct validity. Future research should address this next step in the validation process of the LIRSPE.

Along with construct validity, teachers may consider the use of video and an outside rater to conduct their preliminary LIRSPE evaluation. The outside rater needs to be an individual that the teacher believes is trusted and supportive and can provide feedback in a fair but unbiased manner. If the teacher trusts the outside rater then the results should not be used in a manner that could be interpreted as misuse. Afterwards, it is recommended that the rater and teacher debrief, discuss the results, and develop an action plan to improve items that teachers may potentially score low. After the teacher makes changes, perhaps students who are cognitively capable of accurate estimations can complete the LIRSPE after class so that his/her results could then be compared with self and observer (rater) evaluations of the new lesson. Potentially, the resulting discussion after the

improved lesson would yield continued improvements with inclusive practices in GPE environments.

Finally, although the LIRSPE tool may appear to focus upon games and sports in physical education, it is suggested that the LIRSPE can be used in any GPE context regardless of content. For example, in a dance unit, teachers can allow students to choose their own pace or speed of beat and feature several groups (see speed of play and differentiated instruction, Figure 2). Regarding health and fitness, students should be allowed to complete “task oriented or criterion-based” (e.g. how many can you do in a minute) tasks as opposed to “ego oriented or normative-based” (get 30 in a minute or beat the class record) tasks (see warm-up, autonomy supported instruction, Figure 2). Thus, it is recommended that all GPE lessons should be inclusive and the LIRSPE can produce valid and reliable data assisting teachers with reflection on their ability to create GPE contexts that support all children.

Conclusion

The results of this study reveal valid and reliable results from this preliminary investigation assessing the psychometric properties of the LIRSPE. The LIRSPE is the first tool, to the researchers’ knowledge, to provide an assessment for the extent to which GPE teachers provide an inclusive environment for all children. Future research may consider concurrently assessing teachers’ practices with the LIRSPE and examining the extent to which children felt included in the GPE class. Construct validity for the LIRSPE may also be examined. Practitioners may choose to use the LIRSPE for reflection, self-evaluation, with an outsider rater, or as their own intervention tool to improve inclusive practices within a GPE setting.

Authors’ note

Lieberman and Brian are co-lead authors and equally contributed to the completion of this manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014) *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Ary D, Jacobs LC, Sorensen CK, et al. (2013) *Introduction to Research in Education*. Belmont, CA: Cengage Learning.
- Asbjørnslett M and Hemmingsson H (2008) Participation at school as experienced by teenagers with physical disabilities. *Scandinavian Journal of Occupational Therapy* 15(3): 153–161.
- Avramidis E and Norwich B (2002) Teachers’ attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education* 17(2): 129–147.
- Barnett LM, Hardy LL, Brian A, et al. (2015) The development and validation of a golf swing and putt skill assessment for children. *Journal of Sports Science and Medicine* 14(1): 147–154.

- Beamer J and Yun JK (2014) Physical educators' beliefs and self-reported behaviors toward including students with autism spectrum disorder. *Adapted Physical Activity Quarterly* 31(4): 362–376.
- Block ME and Obrusnikova I (2007) Inclusion in physical education: A review of the literature from 1995–2005. *Adapted Physical Activity Quarterly* 24(2): 103–124.
- Bredahl AM (2013) Sitting and watching the others being active: The experienced difficulties in PE when having a disability. *Adapted Physical Activity Quarterly* 30(1): 40–58.
- Brian A, Ward P, Goodway JD, et al. (2014) Modifying softball for maximizing learning outcomes in physical education. *Journal of Physical Education, Recreation & Dance* 85(2): 32–37.
- Coates J and Vickerman P (2008) Let the children have their say: Children with special educational needs and their experiences of physical education – a review. *Support for Learning* 23(4): 168–175.
- Department for Education and Employment/Qualifications and Curriculum Authority (1999) *Physical Education: The National Curriculum for England*. London, UK: Her Majesty's Stationery Office.
- De Schipper T, Lieberman LJ and Moody B (2017) Kids like me, we go lightly on the head: Experiences of children with a visual impairment on the physical self-concept. *British Journal of Visual Impairment* 35(1): 55–68.
- Doulkeridou A, Evaggelina C and Kudlacek M (2010) Components of attitudes toward the inclusion of students with disabilities in physical education in the ATIPDPE-GR instrument for Greek physical educators. *Adapted Physical Education Laboratory, Serres, Aristotle University of Thessaloniki, Greece* 40(4): 63–68.
- Dyson A (1999) Inclusion and inclusions: theories and discourses in inclusive education. In: Daniels H and Garner P (eds) *Inclusive Education: Supporting Inclusion in Education Systems*. London, UK: Kogan Page, pp.36–53.
- Dyson A, Farrell P, Polat F, et al. (2004) *Inclusion and Pupil Achievement* (Research Rep. No. RR578). London, UK: Department for Education and Skills.
- Field A (2009) *Discovering Statistics using SPSS*. Thousand Oaks, CA: SAGE Publications.
- Fitzgerald H (2006) Disability and physical education. In Kirk D, Macdonald D and O'Sullivan M (eds) *The Handbook of Physical Education* (pp. 752–766). London: Sage. doi:10.4135/978-1-84860-800-9
- Grenier M (2006) A social constructionist perspective of inclusion. *Adapted Physical Activity Quarterly* 23(3): 245–260.
- Grenier M (ed.) (2013) *Physical Education for Students with Autism Spectrum Disorders: A Comprehensive Approach*. Champaign, IL: Human Kinetics.
- Haegele JA and Sutherland S (2015) Perspectives of students with disabilities toward physical education: A qualitative inquiry review. *Quest* 67(3): 255–273.
- Heiman T (2004) Teachers coping with changes: Including students with disabilities in mainstream classes: An international view. *International Journal of Special Education* 19(2): 91–103.
- Hodge SR, Ammah JO, Casebolt K, et al. (2004) High school general physical education teachers' behaviors and beliefs associated with inclusion. *Sport, Education & Society* 9(3): 395–419.
- Hunt P and McDonnell J (2007) Inclusive education. In: Odom SL, Horner HR, Snell M, et al. (eds) *Handbook of Developmental Disabilities*. New York: The Guilford Press, pp.269–291.
- Hutzler Y, Fliess O, Chacham A, et al. (2002) Perspectives of children with physical disabilities on inclusion and empowerment: Supporting and limiting factors. *Adapted Physical Activity Quarterly* 19(3): 300–317.
- Klavina A and Block M (2008) The effect of peer tutoring on interaction behaviors in inclusive physical education. *Adapted Physical Activity Quarterly* 25(2): 132–158.
- LaMaster K, Gall K, Kinchin G, et al. (1998) Inclusion practices of effective elementary specialists. *Adapted Physical Activity Quarterly* 15(1): 64–81.
- Lieberman LJ and Houston-Wilson C (2018) *Strategies for Inclusion: Physical Education for All*. 3rd edition. Champaign, IL: Human Kinetics.

- Lloyd C (2013) The Erasmus Mundus programme: Providing opportunities to develop a better understanding about inclusion and inclusive practice through an international collaborative programme of study. *International Journal of Inclusive Education* 17(4): 329–335.
- Meegan S and MacPhail A (2006) Irish physical educators' attitude toward teaching students with special educational needs. *European Physical Education Review* 12(1): 75–97.
- Mitchell D (ed.) (2005) *Contextualizing Inclusive Education: Evaluating Old and New International Perspectives*. London, UK: Routledge.
- Nevill AM and Atkinson G (1997) Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *British Journal of Sports Medicine* 31(4): 314–318.
- Norwich B (2008) Dilemmas of difference, inclusion and disability: international perspectives on placement. *European Journal of Special Needs Education* 23(4): 287–304.
- Perkins K, Columna L, Lieberman LJ, et al. (2013) Parental perceptions toward physical activity for their children with visual impairments and blindness. *Journal of Visual Impairments and Blindness* 107(2): 131–142.
- Qi J and Ha AS (2012) Inclusion in physical education: A review of literature. *International Journal of Disability, Development and Education* 59(3): 257–281.
- Rink J (2014) Teacher effectiveness in physical education – consensus? *Research Quarterly for Exercise and Sport* 85(3): 282–286.
- Spencer-Cavaliere N and Watkinson EJ (2010) Inclusion understood from the perspectives of children with disability. *Adapted Physical Activity Quarterly* 27(4): 275–293.
- United Nations International Children's Emergency Fund (UNICEF) (2007) Promoting the Rights of Children with Disabilities. Innocenti Digest No. 13. UNICEF. Available at: http://www.un.org/esa/socdev/unyin/documents/children_disability_rights.pdf (accessed 10 October 2016).
- United States Census Bureau (2017) Community Facts. Available at: <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml> (accessed 25 February 2017).
- United States Department of Education (2003) Twenty-fifth annual report to Congress on the implementation of the Individuals with Disabilities Education Act. Available at: <http://www.ed.gov/about/reports/annual/osep/2003/index.html> (accessed 8 March 2017).
- United States Department of Education (2005). Twenty-seventh annual report to Congress on the implementation of the Individuals with Disabilities Education Act. Available at: <http://www.ed.gov/about/reports/annual/osep/2003/index.html> (accessed 12 June 2016).
- Valentini NC and Rudisill ME (2004) An inclusive mastery climate intervention and the motor skill development of children with and without disabilities. *Adapted Physical Activity Quarterly* 21(4): 330–347.
- Weaver RG, Webster C and Beets MW (2013) Let us play: Maximizing physical activity in physical education. *Strategies* 26(6): 33–37.

Author biographies

Lauren Lieberman is a Distinguished Service Professor in the Department of Kinesiology, Sport Studies, and Physical Education at the State University of New York's College at Brockport.

Ali Brian is an Assistant Professor in the Department of Physical Education at the University of South Carolina.

Michelle Grenier is an Associate Professor in the Department of Kinesiology at the University of New Hampshire.