

4-2022

Rasch Analysis of the Physical Activity Enjoyment Scale Among 9th- and 10th-Grade Students

Jane Shimon

Boise State University, jshimon@boisestate.edu

Yong Gao

Boise State University, yonggao@boisestate.edu

Follow this and additional works at: <https://scholarworks.boisestate.edu/ijpah>



Part of the [Exercise Science Commons](#), [Health and Physical Education Commons](#), and the [Public Health Commons](#)

Recommended Citation

Shimon, Jane and Gao, Yong (2022) "Rasch Analysis of the Physical Activity Enjoyment Scale Among 9th- and 10th-Grade Students," *International Journal of Physical Activity and Health*: Vol. 1: Iss. 1, Article 9.

DOI: <https://doi.org/10.18122/ijpah1.1.9.boisestate>

Available at: <https://scholarworks.boisestate.edu/ijpah/vol1/iss1/9>

Rasch Analysis of the Physical Activity Enjoyment Scale Among 9th- and 10th-Grade Students

Abstract

Rasch Model analysis was used to evaluate construct validity and reliability of the Physical Activity Enjoyment Scale (PACES) in an adolescent population. A total of 349 9th and 10th graders (180 male; 169 female), aged 13 to 15 years ($M = 14.13$ $SD = .793$), completed the 18-item PACES scale at the conclusion of physical education classes. Responses to PACES items were analyzed using Rasch partial credit model. Of the 18 PACES items, five were identified as misfit items and were removed. The remaining 13 good-fit items (difficulty level ranged from -0.81 for Item #10 to 0.66 logits for Item #18) demonstrated one dimension which accounted for 62.4% of the variance, with the person separation index was 3.01; however, the 6-point Likert response format was not appropriate for some remaining items. No items presented differential item functioning (DIF) between genders. Rasch analysis supported the construct validity and reliability of PACES to assess PA enjoyment among 9th and 10th grade high school students after removing five misfit items.

Rasch Analysis of the Physical Activity Enjoyment Scale Among 9th- and 10th-Grade Students

Jane Shimon and Yong Gao^a

^aBoise State University

Abstract

Rasch Model analysis was used to evaluate construct validity and reliability of the Physical Activity Enjoyment Scale (PACES) in an adolescent population. A total of 349 9th and 10th graders (180 male; 169 female), aged 13 to 15 years ($M = 14.13$ $SD = .793$), completed the 18-item PACES scale at the conclusion of physical education classes. Responses to PACES items were analyzed using Rasch partial credit model. Of the 18 PACES items, five were identified as misfit items and were removed. The remaining 13 good-fit items (difficulty level ranged from -0.81 for Item #10 to 0.66 logits for Item #18) demonstrated one dimension which accounted for 62.4% of the variance, with the person separation index was 3.01; however, the 6-point Likert response format was not appropriate for some remaining items. No items presented differential item functioning (DIF) between genders. Rasch analysis supported the construct validity and reliability of PACES to assess PA enjoyment among 9th and 10th grade high school students after removing five misfit items.

Key words: validity, measurement, scale, evaluation

Introduction

It has been widely accepted that participating in physical activity throughout a lifetime is an essential component of a healthy lifestyle. When focusing on adolescents, positive outcomes found in participating in physical education during the school day may play an influential role in motivating students to engage in physical activity outside of school and into adulthood (Dishman et al., 2005; Hagger, Chatzisarantis, Culverhouse, & Biddle, 2003; Telama et al., 2005; Wang & Chen, 2020). Enjoyment is a positive outcome of participating in physical activities. Enjoyment can be described as obtaining pleasure, satisfaction, and/or fun from participation in the activity. These affective facets of enjoyment have been linked to intrinsic motivation and have been positively associated with continued physical activity in youth (Chung & Phillips, 2002; DiLorenzo, Stucky-Ropp, VanderWal, & Gotham, 1998; Dishman et al., 2005; Dishman et al., 2018; Motl et al., 2001) and enjoyment in physical education (Bibik, Goodwin, & Orsega-Smith, 2007; Prochaska, Sallis, Slymen, & McKenzie, 2003; Yli-Piipari, Watt, Jakkola, Liukkonen, & Nurmi, 2009).

In order to assess enjoyment of physical activity, sound measurement tools are essential. The Physical Activity Enjoyment Scale (PACES) is a valid and commonly used tool to measure enjoyment of physical activity. The original PACES was an 18-item bipolar scale developed to measure exercise enjoyment in college-aged populations, but has since been modified to assess enjoyment in physical activity (Kendzierski & DeCarlo, 1991). A seven-point semantic differential scale was used to rate the feelings perceived when doing physical activities (e.g., I enjoy it versus I hate it; I find it energizing versus I find it tiring; I dislike it versus I like it).

The scale has been validated and modified to assess enjoyment of physical activity (PA) in various age-level settings such as youth sports (Crocker, Bouffard, & Gessaroli, 1995), adolescent girls and high school students (Molt et al., 2001; Dunton, Tscherne, & Rodrigues, 2009), children (Moore et al., 2009; Paxton et al., 2008), older adults (Mullen et al., 2011), and adults with physical limitations (Murrock, Bekhet, & Zauszniewski, 2016). Molt et al. (2001) shortened the original

PACES to a 16-item survey (S-PACES) and rephrased certain items to better fit the comprehension of adolescent girls. The rating scale was also changed from a 7-point bipolar scale to a 5-point Likert-type scale (Disagree a lot to Agree a lot). In a study with children, Moore et al. (2009) used the 16-item survey and changed the word “depressed” with “sad” to accommodate 3rd grade comprehension. As with the Molt et al. study, both studies showed a methodological effect (similar direction of positively- or negatively-worded items). Due to the methodological effect of PACES, Paxton and colleagues (2008) shortened the S-PACES by removing positively worded items and used the seven remaining negatively-worded items on the 5-point rating scale. Dunton et al. (2009) applied the 18-item PACES and modified the responses to a 6-point Likert (Disagree a lot to Agree a lot) for high school students. The best fit included a one-factor model, with a correlated error among positively worded items.

In the previously mentioned studies, factorial validity of PACES was tested mostly through confirmatory factor analyses (CFA). Based on CFA, modifications to PACES have included removing survey items, rephrasing items, and changing the scale to determine the best fit of the enjoyment scale. However, the prevailing practice in the PA enjoyment studies had been to assign numerical scores to the ordinal response choices (e.g., 1=Disagree a lot to 6=Agree a lot; 1=Totally true to 5=Totally false) and to add the scores across items together to have a total score to represent a participant’s PA enjoyment level. This may not be appropriate because ordinal responses are not aggregates, and the use of total score (addition of scores across items) to represent a participant’s magnitude of a latent variable (e.g., PA enjoyment) is not justifiable when the scores do not represent an interval scale (Bond & Fox, 2007; Rasch, 1980; Wright & Stone, 1999). For this reason, the full 18-item PACES has been re-evaluated with adults using Rasch modeling analyses (Heesch, Måsse, & Dunn, 2006). Rasch models describe the interaction between a respondent’s ability (e.g., level of physical function) and item difficulty (e.g., difficulty of a physical task) that leads to his/her response to an individual item. Respondents with high ability will have a higher probability to endorse a given item or conduct a particular task than respondents with lower ability levels (Bond & Fox, 2007; Rasch, 1980; Tennant, McKenna, & Hagell, 2004). Rasch modeling is often used in the areas of social science, education, psychology and health to examine construct validity of a measure by determining unidimensionality and item-model fit. When an instrument measures one dominant latent trait of interest (unidimensionality), Rasch models transform the routinely used ordinal scores in questionnaire data into interval scores (in logits; log (odds)), and arithmetic operations can be conducted (Bond & Fox, 2007; Rasch, 1980). Rasch analyses can also establish reliability through person and item separation indexes, and the formatting of appropriate responses can be revealed (Bond & Fox, 2007; Rasch, 1980).

Results of Rasch model analyses on PACES from an adult population (Heesch et al., 2006) demonstrated that a one-factor solution was the best fit using a partial credit model. However, two recommendations were made to item fit and response options of PACES. Item 5 (“I am very absorbed”) did not contribute to the measurement of enjoyment and items 9 (“It’s very pleasant”) and 11 (“It’s very invigorating”) were redundant. It was recommended to remove item 5 and item 9 or 11 from the survey. Second, the 7-point bipolar scale was found to not target all respondents. It was recommended to use a 5- or 6-point response scale to help increase the probability of selecting each response category of an item along the Rasch scale continuum (i.e., logit). In addition, adding descriptors to each response was suggested to help provide a better meaning of each option choice.

As alluded to previously, a Rasch model analysis of PACES has only been studied with an adult sample and has not been conducted with adolescent boys and girls. For the PACES to

accurately evaluate PA enjoyment and make meaningful comparisons between genders, it is important that the items in PACES work in the same way for respondents regardless of their gender. That is, the probability of participants with same levels of PA enjoyment responding to a PACES item should not be different simply because participants are in different gender groups. Otherwise, the item is considered function differently (Holland & Thayer, 1986) between genders, and the aggregated scores from such items may lead to biased conclusions regarding gender difference on PA enjoyment. This aspect of PACES has not been examined in the previous studies using Rasch model analysis.

Therefore, the purpose of this study was to re-examine the psychometric properties of PACES with adolescents using Rasch modeling analyses. Based on the recommendations of Heesch et al., (2006), a 6-point Likert scale response was used to assess the 18-item PACES. A secondary aim was to also examine whether PACES item responses functioned differently between boys and girls.

Materials and Methods

Participants

This study included 9th- and 10th-grade physical education (PE) students from one area high school located in the western United States. The school was part of the largest school district in the state, which comprised approximately 34,581 K-12 students. The school was located in a higher-class area of the district where only 19.2% of its students were eligible to receive free and reduced lunch. Overall, 80% of the students were Caucasian. The investigation was approved by the University's IRB, the school district's Board of Trustees, and school principal.

Measures and Procedures

At the start of fall term, parental consent forms were initially dispensed to all 9th grade students enrolled in PE (n = 424). Students were given one week to return the signed forms to their PE teachers. Those students who returned the forms also indicated their willingness to participate (n= 349; 82.3%) by signing an assent form.

The 18-item PACES used for this study with adolescents was revised from a seven-point bipolar scale to a six-point Likert scale (*Totally true, True, Slightly true, Slightly false, False, Totally false*), based on recommendations by Heesch et al. (2006). The revised PACES scale was distributed to students in the gymnasium near the end of each physical education class. Instructions were read to students by the researcher as they followed along. As part of the instruction protocol, the first PASES statement was used as an example to help students become familiar with the 6-point Likert responses they would use when filling out the survey. Additional cues were added by the administrators that helped students reflect on the Likert responses (e.g., "All the time/none of the time" for me, or I completely agree/disagree" were cues used for *Totally true* or *Totally false* responses, while "Sometimes I agree/disagree" were cues used to help clarify *Slightly True* or *Slightly False*); After instructions were given and student questions were answered, students completed the survey in approximately 8-10 minutes. Some students sat in small groups with friends (no more than 4 were in a group) while filling out the survey while others sat alone. No students required individual assistance to complete the form, although on a few occasions, researchers interrupted an entire class to re-define certain terms that appeared to be in question by numerous students. Two consecutive days of data collection were needed to cover all the 9th and 10th -grade physical education classes that were taught on alternate days as part of the school's block scheduling scheme.

Analyses

The Rasch partial credit model was used to evaluate measurement properties of the PACES. The Rasch partial credit model was applied to be consistent with the analysis model used by Heesch et al. (2006) in their study. Rasch partial credit model assumes the distances between adjacent ordinal response options are not equal across items in an instrument (i.e., each item has its own response structure) while Rasch rating scale model assumes equal distance between ordinal response options across all items (i.e., all items share the same response structure; Wright, 1998). In Heesch et al.'s study on PACES, partial credit model showed better model fit and lower percentage of misfit items than Rasch rating scale model, and the item response category analysis also indicated that distances between item response options/categories were not the same across items, which suggest the use of a partial credit model for PACES (Heesch et al., 2006).

The Rasch principle components analysis (PCA) of the residuals was first run to examine construct validity dimensionalities of the PACES. The idea behind a Rasch PCA is that if the Rasch model can explain most of the variances in the data, then the unexplained part of the data, the residuals of the items, is random error and should have no relationship to each other. Therefore, no common factors can be extracted from the residuals (Linacre, 2009; Wright & Stone, 1999). If not, it indicates the data of an instrument contains more than one construct. The suggested cutoff score for a measure to be considered as unidimensional in Rasch PCA analysis is the unexplained variance by the residuals (i.e., the size of each contrast in the PCA output) should be less than 2.0 Eigenvalue units (Linacre, 2009).

When the PCA of the residuals identified one dimension in the PACES data, Rasch analysis was then conducted for that dimension. Simultaneously, misfit items in the PACES were removed, and Rasch analysis was rerun. The process was iterated until the PACES contained only good fit items. To determine model-data fit, Infit and Outfit indexes are the most commonly used fit statistics (Wright & Linacre, 1994). When data fit a Rasch model well, the Infit and Outfit statistics values for items are close to 1. Items with Infit and Outfit values greater than 1.33 or less than 0.75 are considered misfit items (Bond & Fox, 2007). Item difficulty level estimate was then reported for each item.

During Rasch analysis, additional information that can be used to evaluate the quality of PACES, such as the item-person map, was presented to display a clear picture of person ability and item difficulty difference by placing respondents' ability levels and item difficulties on the same continuous metric (Linacre, 2009; Wright & Stone, 1999). The item-person map shows whether the difficulty levels of a set of items in an instrument match the ability levels of the study sample (Wright & Stone, 1999). When every ability level is covered by difficulty-level matched items, it indicates the instrument is able to accurately assess ability of the respondents. When large gaps exist between items on the metric, it indicates that the instrument is less able to distinguish between ability levels of respondents at these points (Linacre, 2009; Wright & Stone, 1999).

To assess reliability, Rasch analysis uses person (and item) separation and reliability statistics (Wright & Stone, 1999). The person separation index describes the range of ability estimates along the continuous metric where a value of 2.0 is considered minimally acceptable, and the higher index value is a more consistent measure. The person reliability statistic ranges from 0 to 1, with higher values (≥ 0.8) representing higher reproducibility (Linacre, 2009; Wright & Stone, 1999). Estimates of the item difficulty, spread along the same continuous metric, is represented by the item separation index, where again, a value of 2.0 is considered minimally acceptable. Item reliability shows how confident the same estimated difficulty order of items would be if another sample of participants respond to that measure (Duncan et al., 2003).

Differential Item Analysis (DIF) analysis was conducted to identify if items functioned differently due to respondents' gender. In Rasch modeling, DIF implies that the item difficulty is different for different groups who are in the same ability level (e.g., same PA enjoyment scores in this study). Items presenting significant DIF with moderate to large size (DIF size $>.64$ logits) were considered DIF items (Linacre, 2009).

Finally, the suitability of the 6-point Likert response format (1= "Totally true" to 6 = "Totally false") used in the PACES was assessed by examining the order of average category measures and Rasch-Andrich thresholds. An appropriate response format should demonstrate increasing levels of a response on the underlying trait being measured. In this study, for example, participants who selected 3 = *Slightly true* should have higher PA enjoyment than those who selected 4 = *Slightly false* for the items such as #1: "I enjoy it". Such increasing levels of a response are indicated by the increasing values of Rasch-Andrich thresholds (also called step difficulty) from a lower category to a higher category (Andrich, de Jong, & Sheridan, 1997). A response format is not considered appropriate if "disordered" thresholds occur, where for example, a higher category (e.g., Option 4) has lower Rasch-Andrich threshold scores than the lower category (e.g., Option 3) (Andrich, deJong, & Sheridan, 1997; Linacre, 2009).

Table 1*Item Difficulty*

Item		Difficulty	SE	Infit	Outfit
*1	I enjoy it	-0.18	0.1	0.85	0.86
3	I dislike it.	0.08	0.1	0.77	0.83
*4	I find it pleasurable.	0.48	0.1	0.83	0.81
*5	I am very absorbed in this activity.	0.36	0.1	1.02	1.18
6	It's no fun at all.	-0.11	0.1	1.06	1.12
*7	I find it energizing.	-0.28	0.1	0.99	0.94
*9	It's very pleasant.	0.35	0.1	1.03	1.09
*10	I feel good physically while doing it.	-0.81	0.1	0.98	1.13
*13	It's very gratifying.	0.01	0.1	1.19	1.33
*14	It's very exhilarating.	0.01	0.1	1.08	1.12
*16	It gives me a strong sense of accomplishment.	-0.68	0.1	1.12	1.28
*17	It's very refreshing.	0.11	0.1	1	1.02
18	I felt as though I would rather be doing something else.	0.66	0.1	0.93	1.05

Note. * Reversely scored in Rasch analysis; SE: Standard error

Rasch analysis was conducted using WINSTEPS 3.68 software. The item difficulty, fit statistics, item-person maps, DIF, Rasch reliability statistics, and Rasch-Andrich thresholds were reported.

origin and of combined ethnicities (12%). No additional data were collected to determine whether students were engaged in interscholastic sports or other physical activity participation after school.

Construct Validity

Dimensionality and Item-Model Fit

When all 18 items of the PACES were included in the initial Rasch analysis, the PCA of the residuals results revealed an eigenvalue of 1.4 unit for the first contrast, which is lower than 2.0 eigenvalue that is commonly used to suggest another dimension. The first contrast explained only 5.2% of the total variance in the data, which is much lower than the variance that was explained by the items (26%) or the persons (30%). Such results suggest that there is not a second dimension and the PACES holds a unidimensional structure. Of the 18 PACES items, five items (i.e., (2: I feel bored 8: It makes me depressed, 11: It's very invigorating, 12: I am very frustrated by it, and 15: It's not at all stimulating) were identified as misfit items (Infit and Outfit statistics < 0.75 or > 1.33 logits) and were removed. The remaining 13 good-fit items demonstrated one dimension and accounted for 62.4% of the variance in the measure. The most difficult item (difficult to select higher response options) was Item 18 (I felt as though I would rather be doing something else), which had a difficulty level of 0.66 logits. The easiest item (easier to select higher response options) was Item 10 (I feel good physically while doing it) with a difficulty level of -0.81. Item difficulty estimates with standard errors and fit statistics of the 13 good fit items are provided in Table 1.

Item-Person Maps

Figure 1 represents the item-person map for the 13 best-fit PACES items. On the map, the continuous numeric scale is shown on the left; next to the numeric scale are the person ability estimates, followed by item difficulty estimates with difficulty thresholds on the right. On the top of each map are placed the persons with higher physical activity enjoyment levels and most difficult items. To reflect the item difficulty spread on the continuous numerical scale, each item is shown five times: the left bottom item section shows each item with a difficulty threshold that separates the likelihood of scoring category 1 from the likelihood of scoring categories 2 or higher; and items in the right upper section are presented with difficulty threshold that separates the likelihood of scoring category 6 from the likelihood of scoring categories 5 or below.

For Figure 1, item difficulty thresholds ranged from -1.83 to 3.11 logits while person ability spread from below -2.43 to around 6.02 logits. Large gaps were observed between person enjoyment levels and the item difficulty levels at the higher end of the Rasch person-item map, where PA enjoyment levels of 11.7% of students was not covered by item difficulty levels.

DIF Attributed to Gender

No item was identified as a DIF item due to respondents' gender (DIF size < 0.64 logits, $p > 0.05$). This result was also confirmed simultaneously by a generalized *Mantel-Haenszel* DIF analysis approach ($p > 0.05$) provided in the Winsteps software (Linacre, 2009; Holland & Thayer, 1986). At the same levels of PA enjoyment, male and female participants responded to each item in a similar way.

Reliability

The separation index for person measures was 3.01, with a reliability value of 0.90, indicating a high degree of confidence that person PA enjoyment would be in the same order in another sample of items. The separation index for item measures was 6.12 with a reliability value of 0.97, indicating a high degree of confidence that item difficulty estimations in the PACES would be in the same order in another sample of participants.

Response Format Assessment

The appropriateness of the 6-point Likert response format utilized in the PACES was examined using Rasch analysis. Like the result showed in Table 2 for Item 1, the step difficult levels did not increase as supposed from the lower category (3 = *Slightly true*) threshold of -0.8 logits to next higher category (4 = *Slightly false*) threshold. Instead, the step difficult level decreased to -1.39 logits. This is an indication of “disordered” thresholds. Similarly, the study results revealed that the threshold estimates (i.e., step difficulty) were also disordered for Items #5, 6, 7, 16 & 18) in the PACES. See Table 2 for details.

Table 2

Disordered Step Difficulty Levels of Six PACES Items

Category	*Item 1	*Item 5	Item 6	*Item 7	*Item 16	Item 18
Totally True (=1)	NONE	NONE	NONE	NONE	NONE	NONE
True (=2)	-1.02	-0.94	-0.82	-0.82	-2.83	0.13
Slightly True (=3)	-0.8	-0.33	-1.54	-1.39	-0.54	-0.49
Slightly False (=4)	-1.39	-0.35	0.39	-1.41	-1.21	1.06
False (=5)	0.31	0.75	0.19	0.19	-0.12	0.78
Totally False (=6)	1.98	2.67	1.25	2.02	1.31	1.83

Note. *: Reversely scored in Rasch analysis; *Step Difficulty* is also called Rasch Andrich threshold

Discussion

The current study is the first known study to analyze the Physical Activity Enjoyment Scale (PACES) utilizing Rasch modeling with 9th and 10th grade boys and girls. The objective was to use Rasch modeling to re-evaluate psychometric properties of PACES with adolescents and determine if the scale contained DIF items attributed to gender. Rasch analysis supported the construct validity and reliability of PACES to assess PA enjoyment among 9th and 10th grade high school students after removing five misfit items. Of the remaining 13 items, the 6-point Likert response format was not appropriate for some items. There was no gender DIF item in the scale.

The most revealing outcome of this study was directed to misfit items. Based on the one-dimension structure, five items were deemed as misfits. In other words, those items did not contribute to the construct validity of the instrument. Removing five misfit items (2: I feel bored, 8: It makes me depressed, 11: It's very invigorating, 12: I am very frustrated by it, and 15: It's not at all stimulating) led to better assessment of enjoyment of PA with male and female 9th and 10th grade high school students.

Item fit results in this study were in partial agreement with previous PACES studies. Item 11 (It's very invigorating) was one of the misfit items in our study with 9th and 10th graders. A Rasch model analysis on PACES with adults also suggested to remove item 11 due to its similar meaning to item 9 (It's very pleasant) (Heesch et al., 2006). In addition, based on PACES analyses on female adolescents (Motl et al., 2001), item 11 was one of the items removed, and item 11 was also absent in the revised 7-item PACES with children (Paxton et al., 2008). Finally, item 11 was also found to have a low factor load in PACES results on high school students (Dunton et al., 2009), although the item contributed to the overall fit of the scale. In the current study, some high

school students questioned the meaning of “invigorating”. Confusion of the term may have contributed to the misfit item.

The other four misfit items found in this study (2: I feel bored; 8: It makes me depressed; 12: I am very frustrated by it; 15: It’s not at all stimulating) were not found to be an issue in the overall item-fit of PACES in the Rasch model analysis on adults (Heesch et al., 2006) or the CFA conducted on PACES in other studies (Brown, 2006; Dunton et al., 2009; Motl et al., 2001; Mullen et al., 2011). It is plausible that the 9th and 10th grade students in this study may have interpreted some of the words differently than their peers or adults. For example, the word “bored” found in item 2 could have meant “uninteresting” for some students, while others could have taken the meaning to “be tired”. The word “depressed” found in item 8 may have been confusing to students. Some students may have thought the word associated with the signs and symptoms of depression, while others may have assumed the word related to more of a global meaning, such as being blue, unhappy, or sad. Previous modifications of PACES with children and adolescents have included word changes or re-phrasing of items to help students comprehend the meaning of the items (Moore et al., 2009; Motl et al., 2001). For example, in a previous study with female adolescents, item 8 was replaced with the word “sad” instead of “depressed” and item 15 was replaced with “interesting” instead of “stimulating”. Many students in this study questioned the meaning of “stimulating” as they were completing the survey.

The potential exists for an additional explanation for the five misfit items in this study. Four of the five items (80%) were phrased in a negative direction. While previous studies reported findings in regards to positively and negatively worded items, the results were found to be irrelevant methodological effects and did not impact the overall fit of PACSE items in a one-factor model (Dunton et al., 2009; Motl et al., 2001). Nevertheless, when looking at the 13 good fit items in this study, only three negatively-worded items remain (3: I dislike it, 6: It’s no fun at all, and 18: I felt as though I would rather be doing something else).

One item (5: I am very absorbed in this activity) contributed to the overall fit of PACES in this study, yet was found to be a misfit item in the Rasch model analysis on adults (Heesch et al., 2006) and other studies using CFA with adolescent girls and older adults (Motl, 2001; Mullen et al., 2011). This was a surprise, for numerous students had questions regarding the meaning of the word “absorbed”. During the administration of PACES, researchers provided a definition or clarification of the word “absorbed” to those students who had questions. This action may have helped them record a better choice to that item.

The item-person map in this study provides useful information to determine whether the item difficulties matched the person abilities, and how to improve the PACES. Although it can be seen in the map that the item difficulty estimates and the person ability levels were relatively well-matched, large gaps were observed between person enjoyment levels and the item difficulty levels at the higher end (i.e., on the top) of the map. This is an indication of a ceiling effect, which reveals that the 13-item PACES cannot distinguish among students with high levels of PA enjoyment, most likely those who always enjoy physical activities they have been doing. This also suggests new items that measure high levels of PA enjoyment should be developed for the PACES to better assess PA enjoyment among high school students.

The reliability estimates for both persons and items were high (>0.9), indicating the observed responses to the PACES items are highly reproducible. Previous PACES studies have also supported that PACES is a reliable measure of PA enjoyment (Crocker et al., 1995; Kendzierski & DeCarlo, 1991; Moore et al., 2009).

Rasch modeling also examined the response format of PACES. Responses for each item should show an increase in difficulty levels among the choices. Based on the participants' responses, the 6-point scale used in this study did not discriminate scale-item difficulty on six items, which can be seen in Table 2. This finding suggests that there was not a clear distinction among the order of choices for some of the items (disordered thresholds). Participants in this study had a hard time discerning whether to mark items 5 (I am very absorbed in this activity) and 16 (It give me a strong sense of accomplishment) “*Slightly true* or *Slightly false*”. Items 6 (It’s no fun at all) and 18 (I felt as though I would rather be doing something else) posed similar discrepancies for both the true and false sides of the scale. If adolescents were inclined to agree that items 6 and 18 were true or false, they struggled to determine the level of True (*True* or *Slightly true*) or the level of False (*Slightly false* or *False*). Finally, two items (1-I enjoy it and 7 - I find it energizing) demonstrated a disordered level among three response choices: *True*, *Slightly true*, or *Slightly false*. Authors of the Rasch model analysis conducted with adults suggested using a 5- or 6-point scale (Heesch et al., 2006). Based on our study, future investigations are needed to determine if a 5-point Likert-type scale would further delineate item choice responses for 9th and 10th grade students.

Finally, the results of this study showed there were no DIF items attributed to gender. While boys tended to report higher overall PACES scores (higher levels of enjoyment) than girls, the 13 PACES items appeared to have similar meanings to both boys and girls who were at the similar levels of PA enjoyment. As a result, when using the 13-item PACES to assess PA enjoyment among 9th and 10th grade students, interpretations of PA enjoyment score difference between genders can be made with confidence. In previous studies, modified versions of PACES have also demonstrated invariance between boys and girls (Dunton et al., 2009; Paxton et al., 2008).

Limitations

Although every attempt was made to give precise directions and examples to 9th- and 10th-grade students before they filled out PACES, some participants may have been confused with some of the words used in the survey. Some students may not have raised their hand to ask questions regarding the meaning of certain words, while others who sat in small groups sometimes talked about certain words within the items, which may have resulted in a shared response instead of an individual choice. It is unknown whether small-group discussion regarding terminology played a factor in marked responses for various survey items. For future studies, it is recommended that adolescents sit in their own personal space while filling out the form to assure item responses are based solely on personal choices.

Different researchers administered PACES over the course of the 2 recording days. While every attempt was made to answer questions by students as they arose (i.e., explanation of “absorbed” (item 5), “exhilarating” (item 14), and “stimulating” (item 15), the answers they received by the researchers may have slightly varied which could have affected the choices students recorded. This was an unanticipated event during the study. Although being helpful is an automatic default when working with students, perhaps in this instance, it would be best for investigators to tell students to fill out the survey as best they can, without the added clarifications to certain words or phrases.

Finally, the PE classes in this high school consisted exclusively of repetitive and rigorous fitness and conditioning exercises. While daily PE lessons were different within the realm of fitness and conditioning activities, students did not experience a variety of traditional sports or other game-like activities compared to other studies using PACES in a PE setting. It is unknown whether the PE curriculum students experienced had any impact on PACES outcomes in this study.

Future studies should examine the potential impact of different types of physical activities on PACES outcomes. Future investigation using PACES may also need to clarify students' understandings on physical activity and/or provide examples of physical activities that are outside of PE lessons to eliminate any confounding effect that various understanding of PA may have on PACES outcomes.

Conclusion

Rasch analysis supported the construct validity and reliability of the PACES to assess PA enjoyment among both male and female 9th- and 10th-grade high school students after removing 5 misfit items. Further Rasch analyses using PACES with other adolescent age levels are warranted. More empirical evidence is needed to determine if a 5-point response scale would demonstrate an ordered step of item difficulty on the revised 13-item scale, and it is recommended to develop additional items to help distinguish between higher levels of physical activity enjoyment. Finally, it is unknown whether varied interpretations or a lack of understanding of certain words had an effect on the 5 misfit PACES items in this study. Based on previous word modifications made to PACES, (i.e., used "sad" in lieu of "depressed" in Item 8 and replaced "stimulating" with "interesting" in item 15; Dunton et al., 2009; Motl et al., 2001), it may be warranted to revisit the five misfit items in this study to determine if more appropriate or recognizable terms with this age level of adolescents would change Rasch modeling analyses of PACES.

References

- Andrich, D. J., de Jong, A. L. & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost and R. Langeheine (Eds.), *Application of Latent Trait and Latent Class Models in the Social Sciences* (59-72). Waxmann Publisher.
- Bibik, J. M., Goodwin, S. C., & Orsega-Smith, E. M. (2007). High school students' attitudes towards physical education in Delaware. *Physical Educator*, 64(4), 192-204.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model - Fundamental Measurement in the Human Sciences*, 2nd ed. New York, NY: Routledge.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*, 1st ed. New York, NY: The Guilford Press.
- Chung, M., & Phillips, D. A. (2002). The relationship between attitude towards physical education and leisure-time exercise in high school students. *Physical Educator*, 59(3), 126-138.
- Crocker, P. R. E., Bouffard, M., & Gessaroli, M. E. (1995). Measuring enjoyment in youth sport settings: A confirmatory factor analysis of the physical activity enjoyment scale. *Journal of Sport and Exercise Psychology*, 17: 200-205.
- DiLorenzo, T. M., Stucky-Ropp, R. C., VanderWal, J. S., & Gotham, H.J. (1998). Determinants of exercise among children: A longitudinal analysis. *Preventive Medicine*, 27(3), 470-477.
- Dishman, R. K., Molt, R.W., Saunders, R., Felton, G., Ward, D. S., Dowda, M., & Pate, R. R. (2005). Enjoyment mediates effects of a school-based physical-activity intervention. *Medicine Science in Sports and Exercise*, 37(3), 478-487.
- Dishman, R. K., McIver, K. L., Dowda, M., & Pate, R. R. (2018) Declining physical activity and motivation from middle school to high school. *Medicine & Science in Sports & Exercise* 50(6): 1206-1215.
- Duncan, P. W., Bode, R. K., MinLai, S., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation*, 84: 950-963.

- Dunton, G. F., Tscherne, J., & Rodrigues, D. (2009). Factorial validity and gender invariance of the physical activity enjoyment scale (PACES) in older adolescents. *Research Quarterly for Exercise and Sport*, 80(1): 117-121.
- Hagger, M. S., Chatzisarantis, N. L. D., Culverhouse, T., & Biddle, S. J. H. (2003). The processes by which perceived autonomy support in physical education promotes leisure-time physical activity intentions and behavior: A trans-contextual model. *Journal of Educational Psychology*, 95(4), 784-795.
- Heesch, K. C., Mâsse, L. C., & Dunn, A. L. (2006). Using Rasch modeling to re-evaluate three scales related to physical activity: Enjoyment, perceived benefits and perceived barriers. *Health Education Research*, 21 (supplement 1), i58-i72.
- Holland, P. W., & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure (Technical Report No. 86-69). Princeton, NJ: Educational Testing Service.
- Kendzierski, D., & DeCarlo, K. J. (1991). Physical activity enjoyment scale: Two validation studies. *Journal of Sport and Exercise Psychology*, 13:50-64.
- Linacre, J. M. (2009). Winsteps® (Version 3.68) [Rasch measurement computer software]. Beaverton, Oregon: Winsteps.
- Moore, J., Yin, Z., Hanes, J., Duda, J, Gutin, B., & Barbeau, P. (2009). Measuring enjoyment of physical activity in children: validation of the Physical Activity Enjoyment Scale. *Journal of Applied Sport Psychology*, 21, 116-129.
- Motl, R. W., Dishman, R. K., Saunders, R., Dowda, M., Felton, G., & Pate, R. R. (2001). Measuring enjoyment of physical activity in adolescent girls. *American Journal of Preventive Medicine*, 21(2), 110-117.
- Mullen, S. P., Olson, E. A., Phillips, S.M., Szabo, A. N., Wójcicki, T. R., Mailey, E. L., ... McAuley, E. (2011). Measuring enjoyment of physical activity in older adults: Invariance of the physical activity enjoyment scale (paces) across groups and time. *International Journal of Behavioral Nutrition and Physical Activity*, 8, 103-111.
- Murrock, C. J., Bekhet, A., & Zauszniewski, J. A. (2016). Psychometric evaluation of the physical activity enjoyment scale in adults with functional limitations. *Issues in Mental Health Nursing* 37(3): 164-171.
- Paxton, R.J., Nigg, C., Motl, R.W., Yamashita, M., Chung, R., Battista, J., & Chang, J. (2008). Physical activity enjoyment scale short form – Does it fit for children? *Research Quarterly for Exercise and Sport*, 79(3): 423-427.
- Prochaska, J. J., Sallis, J. F., Slymen, D. J., & McKenzie, T. L. (2003). A longitudinal study of children's enjoyment of physical education. *Pediatric Exercise Science*, 15,170-178.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (revised and expanded ed.). Chicago: The University of Chicago Press (Original work published 1960).
- Telama, R., Yang, X., Viikari, J., Valimaki, I., Wanne, O., & Raitakari, O. T. (2005). Physical activity from childhood to adulthood: A 21-year follow-up study. *American Journal of Preventive Medicine*, 28(3), 267-273.
- Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value Health*, 7 (Supple 1), S22-6.
- Wright, B. D. (1998). Model selection: Rating scale model (RSM) or partial credit model (PCM)? *Rasch Measurement Transactions*, 12(3), 641-642. Retrieved at <https://www.rasch.org/rmt/rmt1231.htm>.

- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wright, B. D., & Stone, M. H. (1999). *Measurement Essentials*, 2nd ed. Wide Range, Inc., Wilmington, Delaware: Wide Range, Inc. Retrieved at <http://www.rasch.org/rasch.htm#measess>.
- Yli-Piipari, S., Watt, A., Jaakkola, T., Liukkonen, J., & Nurmi, J. (2009). Relationships between physical education students' motivational profiles, enjoyment, state anxiety, and self-reported physical activity. *Journal of Sports Science and Medicine*, 8: 327-336.
- Wang, Y., & Chen, A. (2020). Two pathways underlying the effects of physical education on out-of-school physical activity. *Research Quarterly for Exercise and Sport* 91(2): 197-208.