Australian Journal of Educational & Developmental Psychology. Vol 14, 2014, pp. 29-44

Using Rasch analysis to examine the dimensionality structure and differential item functioning of the Arabic version of the Perceived Physical Ability Scale for Children

Sabry M. Abd-El-Fattah[1], Yousra AL-Sinani, Sahar El Shourbagi
Sultan Qaboos University

&

Hessa A. Fakhroo
Qatar University, Qatar

ABSTRACT

This study uses the Rasch model technique to examine the dimensionality structure and differential item functioning of the Arabic version of the Perceived Physical Ability Scale for Children (PPASC). A sample of 220 Omani fourth graders (120 males and 100 females) responded to an Arabic translated version of the PPASC. Data on students' participation in physical activity were also collected using the Participation in Physical Activity Scale (PPAS). The analyses supported the unidimensionality of the factorial structure of the PPASC. The PPASC items were found to function equivalently across male and female groups. There were significant gender differences in perceived physical ability favoring males. The PPASC correlated positively and significantly with the PPAS.

Key words: Perceived physical activity, Rasch analysis, differential item functioning, participation in physical activity, Omani children

INTRODUCTION

Self-efficacy is described as ‘‘people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances’’ (Bandura, 1986, p. 391). The basic principle behind self-efficacy is that individuals are more likely to engage in activities for which they have high self-efficacy and less likely to engage in those for which they do not (Stajkovic & Luthans, 1998), and thus self-efficacy functions as a self-fulfilling prophecy. Bandura (1997) proposed that self-efficacy beliefs are shaped by cognitive processing and integration of four main sources of information: (1) performance attainments and failures, that is, what we try to do and how well we succeed; (2) vicarious performances, that is, what we see other people do; (3) verbal persuasion, that is, what people tell us about what we are able or not able to do; and (4) imagined performances, that is, what we imagine ourselves doing and how well or poorly we imagine ourselves doing it.

Self-efficacy beliefs are presumed to have actual ability to do the task as an underlying determinant. In other words, someone who typically does well on a task knows that he or she

[1] Contact
Sabry M. Abd-El-Fattah,
Department of Psychology,
Faculty of Education,
Sultan Qaboos University, Al-Khoud,
PO Box 32 PC 123, Muscat, Oman
Email: sabryrahma@hotmail.com

does well and shows this knowledge in his or her self-efficacy ratings. However, beliefs in one's self-efficacy are not based solely on knowledge of one's ability. Self-beliefs go beyond actual capability, being ''instrumental in determining what individuals do with the knowledge and skills they have'' (Pajares & Miller, 1995, p.190). According to Bandura (1997), self-efficacy has ''effects on thought, affect, action, and motivation'' (p. 46). Thus, a person high in self-efficacy might do better because that person approaches a task with a different mindset from that of a person low in self-efficacy, even though both of them might have the same level of ability.

Self-efficacy and domain specificity

Bandura (1997, p. 42) maintained that self-efficacy ''is not a contextless global disposition [to be] assayed by an omnibus test.'' Instead, proper self-efficacy measures ''must be tailored to domains of functioning.'' Such domains can refer to any activity, or class of activities, where individuals can differ in their success rates and, more importantly, in their beliefs about their success rates. The domain might be mathematics, biology, or language. The domain might include tasks involving physical strength, eye-hand coordination, or memory. The domain could be personal relationships, being a good parent, or sticking to a diet.

Within a domain, self-efficacy beliefs can be measured with respect to a diverse array of accomplishments. Consider the domain of physical fitness. At a narrow level, one could measure self-efficacy for performing a specific physical exercise. At a broader level, one could measure self-efficacy for passing Standing Stork Tests (tests of an athlete's ability to maintain a state of equilibrium in a static position). At a broader level, one could measure self- efficacy for physical fitness.

There are some measures of self-efficacy that are so broad in scope that they do not refer to any specific performance domain. Such global measures refer to general competence and life coping skills, with items related to accomplishing goals in general and performing effectively on different tasks (e.g., Chen, Gully, & Eden, 2001). However, Bandura (1997) has maintained that global self-efficacy measures ''violate the basic assumption of the multidimensionalities of self-efficacy beliefs'' (p. 48) and that ''undifferentiated, contextless measures of personal efficacy have weak predictive value'' (p. 49). Ideally, according to Bandura, a self-efficacy measure should match, in level of generality, the performance criterion of interest. For example, if the criterion is a particular score on a Standing Stork Test, then the self-efficacy measure should assess a person's beliefs about his or her performance on that narrow task. If the criterion is overall physical fitness, then the self-efficacy measure should be broader, referring to a person's expectations about his or her performance on a physical fitness aptitude test.

Perceived physical ability

An important domain for examination of a person's self-efficacy beliefs is physical activity. Colella, Morano, Bortoli, and Robazza (2008) defined perceived physical ability as one's confidence to participate in a particular physical activity, to overcome barriers to physical activity, and to organize times and responsibilities for physical activity. This is an important area to investigate because physical activity rates for most children are insufficient for health benefits and inactivity-related diseases are on the rise in many countries, including Canada (Hills, King, & Armstrong, 2007), the United States (Kimm & Obarzanek, 2002), England (Information Centre, 2006), Australia (Sanigorski, Bell, Kremer, & Swinburn, 2007), and Oman (Al-Saidi, 2010). For example, the Omani guidelines suggest that children and adolescents engage in 90 minutes of moderate to vigorous physical activity per day. However, over 90% of Omani youth aged 6 to 20 years are not meeting these guidelines (Ministry of Health, 2010).

The Perceived Physical Ability Scale for Children

To quantify self-efficacy beliefs about physical activity in school children, Colella et al. (2008) developed the Perceived Physical Ability Scale for Children (PPASC) that measures children's beliefs about their ability to engage in and perform physical activities. The scale consists of six items that cover strength, speed, and coordination related to performing physical activities. They are rated on a scale from 1 to 4. A label is assigned to each point of the response scale to help children understand the items (e.g., I run very slowly; I run slowly; I run fast; I run very fast). Items 1, 3, and 5 are scored on a scale from 1 to 4 while the scores of items 2, 4, and 6 are reversed. The total test score can range from 6 to 24. High scores would indicate a high perceived physical ability while low scores would reflect low perceived physical ability (See appendix 1).

In developing the PPASC, Colella et al. (2008) administered the six items to a sample of 1914 children (997 girls and 917 boys) aged between 8 to 10 years, drawn from fifteen elementary schools. An exploratory factor analysis (EFA) of data from a sub-sample (n=300) retained one factor, perceived physical ability, which explained 40% of the total variance extracted. A series of confirmatory factor analyses (CFA) of a single factor model using data from 1614 students subdivided into six categories of sex by age (that is, girls 8 years, boys 8 years; girls 9 years, boys 9 years; girls 10 years, boys 10 years) showed that the model fitted the data adequately in all instances. The PPASC showed a split-half reliability coefficient of .70 and a Cronbach's alpha of .72.

Several studies have examined the validity of the factorial structure of the original English version of the PPASC. For example, Draun and Stevens (2009) reported that an EFA of responses from a sample of 204 British children aged between 12 and 13 years retained a single factor that accounted for 54% of the total variance extracted. Rabi and Swanson (2010) found that a CFA of responses from 311 Canadian children aged between 10 and 12 years demonstrated that a single factor model fitted the data adequately after correlating the error terms of Item 1 and Item 5. Other studies have examined the validity of the factorial structure of translated versions of the PPASC. For example, Carmen and Shineder (2011), using CFA, reported that a single-factor structure of the French version of the PPASC fitted the data from 280 children aged between 9 and 12 years only after correlating the error terms of Items 1, 3 and 6. They concluded that the “structure of the PPASC should be examined with more meticulous procedures that can explore its items functions" (p.11).

Perceived physical ability and participation in physical activity

Self-efficacy beliefs in the domain of physical activity are of interest because much research has shown that self-efficacy is a critical antecedent to physical activity. High self-efficacy has been linked to better performance on physical activity tasks, expending more effort on physical activity tasks, and persevering when difficulties arise (Gao, Lee, Kosma, & Solmon, 2010; Gao, Lodewyk, & Zhang, 2009). For example, Gao et al. (2010) found that self-efficacy predicted 54% of the variance in physical activity among 207 middle school students in physical education classes. Gao, Lochbaum, and Podlog (2011) found that self-efficacy predicted 27% of the variance in physical activity among 194 middle school students in physical education classes when it was set as the sole predictor, and it predicted 28% of the variance in physical activity when it was set as a predictor along with a mastery-approach goal and mastery motivational climate variables.

Gender differences in perceived physical ability

With respect to gender differences in perceived physical ability using the PPASC, Colella et al. (2008) reported that males had higher levels of perceived physical ability than females. The age main effect and age by sex interaction effect were not statistically significant. Likewise, Carmen and Shineder (2011), using the French version of the PPASC, found significant differences in perceived physical ability favoring males. In contrast, Draun and Stevens (2009) and also Rabi and Swanson (2010), using the original English version of the PPASC, reported significant gender differences in perceived physical ability favoring females.

Rationale for the present study

Cultural values and norms in relation to physical activity in different countries may account for the cross-cultural variations in levels of physical activity reported by children (Lee & Martinek, 2009). In fact, a number of studies have examined physical activity among specific cultural groups (Nakamura, 2002; Vertinsky, Batth, & Naidu, 1996) and found that social norms of ethno-cultural communities play a significant role in exposure to and attitudes toward physical activities and these in turn affect physical activity.

Culture values and norms may affect not only the type of information provided by the various sources of self-efficacy (see, Bandura, 1997), but also may affect what information is selected and how it is weighted and integrated into a person's self-efficacy judgments. For example, people in an individualist culture may focus their self-appraisals on information about their personal attainments. On the other hand, for people in collectivist cultures, evaluation of their performance by members of their cultural group may be the most important source of efficacy formation. Modeling of people within their cultural group may be important too.

In an individualist society, when approaching a new task, an individual's self-appraisal of efficacy is likely to be affected by his or her previous performance on similar tasks. In a collectivist society, an individual's self-appraisal of efficacy is likely to be affected by the beliefs of the cultural group. Does the group think the individual has the capability to perform the task? Would other members of the group be likely to do the task better (Bandura, 1986; Oettingen, 1995)? In addition to concerns about the effects of cultural values on both perceived physical ability and actual physical activity, the validity of the PPASC needs to be assessed in a non-Western context because it is possible that instruments developed in the West might not work in the same manner in non-Western settings (Maneesriwongul & Dixon, 2004).

Specifically, there are two areas that require more attention. First, it is not clear whether the items of the PPASC are consistent with the Rasch model (Rasch, 1960; Wright, & Stone, 1979). The Rasch model uses location parameters (item location) to model item characteristics. The location parameters specify the position of the item and the item's response categories on the continuum of the latent variable. Thus, item parameters and person measures are directly comparable because they are on the same metric (Bond & Fox, 2007). The Rasch model assesses the metric properties of unidimensionality and provides information about the behaviour of the items, the relative ease with which each item can be endorsed by respondents, and whether the difficulty levels of the items reflect the full range of respondents' trait characteristics. Furthermore, the Rasch model strengthens the measurement quality of a questionnaire by weighting individual items based on their contribution to the underlying trait, and allowing transformation of raw scores into continuous data. Items which are redundant for precise measurement can be identified and removed from the scale (that is, misfitting items) (Hambleton & Swaminathan, 2010; van der Linden & Hambleton, 2010).

Second, it is not clear whether gender differences in perceived physical ability, as measured by the PPASC, are due to that fact that male and female children differ in the underlying latent trait of perceived physical ability or whether these differences represent an artifact in methodology because the items of the PPASC function differently across gender. This is an important methodological issue because unless there is reasonable support for the invariance of the PPASC items across gender, it may not be appropriate to pool data across male and female participants (Abd-El-Fattah, 2013).

Tittle (1994) noted that examination of test items for bias towards groups is an important part in the evaluation of the overall instrument because it influences not only testing decisions but also use of test results. Under these circumstances, it is necessary to apply differential item functioning (DIF) detection procedures (Osterlind & Everson, 2009; Zumbo, 2007) to determine whether the individual items on the PPASC function in the same way for male and female children. Thus, the present study uses the Rasch model item threshold approach to assess DIF of the PPASC items.

The aims of the present study are fourfold. First, are the items of the PPASC (designed to measure children's beliefs about their physical capabilities to successfully engage in and perform physical activities) consistent with the Rasch model? Second, do gender differences emerge when examining the DIF of the PPASC items across male and female children using the Rasch model item threshold approach? Third, what is the convergent validity of the PPASC? This question will be answered by investigating its relationship to a self-report measure of participation in physical activity. Fourth, are there gender differences in perceived physical ability? The results from this study can contribute to the body of literature on perceived physical ability by providing empirical evidence of construct and convergent validity, as well as validity of inferences regarding gender differences in perceived physical ability.

METHOD

Participants

Subjects of the study included 220 Omani children (120 males and 100 females) from three public primary schools in three governorates in Oman. All students were at Year Four. All schools were in metropolitan areas and had male and female students. The means and the standard deviations of age were 10.6 years (SD=0.68) for boys and 10.2 years (SD= 0.44) for girls. Only students with complete data were retained. The percentage of missing data was 2%. Those students left several items blank on the PPASC. Arabic was the native language of all participants students.

Measures

Perceived physical ability

The PPASC (Colella et al., 2008) is a self-report measure of children's beliefs about their physical capabilities to successfully engage in and perform physical activities. The scale consists of six items that represent strength, speed, and coordination related to performing physical activities and are rated on a scale from 1 to 4. A label is assigned to each point of the response scale to help children grasp the meaning of the items (e.g., I run very slowly; I run slowly; I run fast; I run very fast). Items 1, 3, and 5 are scored on a scale from 1 to 4, whereas the scores of items 2, 4, and 6 are reversed. The total test score can range from 6 to 24. High scores would indicate a high perceived physical ability, whereas low scores would reflect low perceived physical ability. Based on the dataset of the present study, the PPASC has a Cronbach alpha of .85.

The author translated the PPASC from English into Arabic using the back-translation method. Three other qualified translators, working without reference to the English version of the PPASC, independently translated the Arabic version back to English. Three other qualified translators independently compared the original English version of the PPASC with the new English version that was translated back from Arabic. Any discrepancies were noted. This iterative process of translation and back-translation continued until no semantic differences were noticed between both questionnaires (Brislin, 1980). Within the dataset of the present study, Cronbach's alpha of the PPASC was 0.85.

Participation in physical activity

Al-Saharti and Al-Mahroky (2008) developed the Participation in Physical Activity Scale (PPAS) using a sample of 450 fourth and fifth graders in Oman. The PPAS measures children's participation in physical activity outside school physical education classes. The scale consists of three items. The first item assesses frequency by asking participants how many times per week, for the past two months, they exercised or played sport (four categories: 'hardly ever/not at all'; 1 to 2 times per week; 3 to 4 times per week; more than 4 times per week). The second item assesses the duration of time participants engaged in these activities per week (four categories: less than one 1 hour; 1 to 3 hours; 3 to 5 hours; more than 5 hours). The third item assesses recent physical activity by asking participants about their physical activity for the last week (four categories: hardly ever/not at all; 1 to 2 times in the week; 3 to 4 times in the week; more than 4 times in the week). Within the dataset of the present study, Cronbach's alpha of the PPAS was 0.78.

Procedures

Approval was obtained to conduct the research. Students were recruited to participate during their normal physical education classes at their schools. The PPASC and the PPAS were administered by trained researchers using standardized instructions. To minimize students' tendency to give socially desirable responses, students were encouraged to answer truthfully and were assured that confidentiality of their answers would prevail at all times. The participant classes were chosen depending on students' schedules on the day and time of the administration of the measures. Students first responded to the PPASC and then to the PPAS. The items of the Arabic version of the PPASC were apparently within the age-equivalent reading level of the Omani children because they did not indicate any difficulty understanding their content. The students completed the PPASC and the PPAS in 10 to 15 minutes.

Overview of the analysis

Data on PPASC items were analyzed using the Rasch modeling measurement procedure (Rasch, 1960), which allowed both students' performance and item difficulties to be measured using the same metric and placed on the same scale. The basic Rasch model is a dichotomous response model (Rasch, 1960; Wright & Stone, 1979) that represents the conditional probability of a binary outcome as a function of a person's trait level (B) and an item's difficulty (D). Andrich (1978, 1988) is credited with extending the Rasch dichotomous response model to the rating scale by the addition of an additional difficulty parameter; either a second δ parameter or a τ parameter. The rating scale model is an additive linear model that describes the probability that a specific person (n) will respond to a specific Likert-type item (i) with a specific rating scale step (x) (Abd-El-Fattah, 2007). It is important to note that the Likert scale can be modelled with either the rating scale or the partial credit model (Masters, 1982; Wright & Masters, 1982). In essence, the rating scale models are a subset of the partial credit models (Andrich, 1978).

Using the Rasch rating scale, the items of the PPASC were calibrated in terms of the extent to which students agreed with the items which corresponded to the item difficulty for the scale. The response to an item was governed by referring the latent variable score to the threshold(s) and the item response was determined from this comparison. A threshold parameter (τij) for an item indicates the transition point between two adjacent response categories j and j + 1 within item i, where two adjacent categories are equally likely. The average of the threshold parameters represents the overall location (δi) of the item (Andrich, 1978, 1988). As the PPASC had four response categories, there were three thresholds estimated for each item. An item with a higher threshold is more difficult than other items.

INFIT and OUTFIT mean square fit statistics were used to examine how well individual items fitted the Rasch model. These information weighted index statistics assessed the extent to which unpredicted responses to an item were given by students whose position in the hierarchy, as determined by their perceived physical ability, is either close to the item's position (INFIT statistic) or far from the item's position (OUTFIT statistic) in the hierarchy of items (Yim, Abd-El-Fattah, & Lee, 2007). For the data to fit the model adequately, it is generally recommended that the two fit statistics ranged from 0.72 to 1.30 logits (Bond & Fox, 2007). Fit statistics higher than 1.30 and below 0.72, respectively indicated that these items did not discriminate well or provided redundant information. Items with poor fit statistics should be considered for removal from the PPASC (Wright & Stone1979).

Another way of investigating the fit of the data to the Rasch scale was to examine the estimates for each case which indicate the performance level of each student on the PPASC. The case Outfit mean square statistic (OUTFIT MNSQ) measured the consistency of the fit of the person to the student characteristic curve for each student, with special consideration given to extreme items while INFIT MNSQ was more sensitive to the pattern of responses to items targeted on the person, and vice-versa. In the present study, the general guideline used for interpreting a suitable value for the OUTFIT MNSQ value was whether the |t- value| was greater than 5 (Wright & Stone 1979). If the obtained value was greater than 5, the person did not fit the scale and should thus be deleted from the analysis.

A test item is labeled with differential item functioning (DIF) when examinees with equal ability, but from different groups, have an unequal probability of item success (Osterlind & Everson, 2009; Walker, 2011; Zumbo, 2007). DIF examination usually involves two or more subgroups. The reference group provides a baseline for comparing the responses (e.g. males), and the focal group is the focus of equivalence concerns (e.g. females). To illustrate, imagine there are two groups of participants, one males and the other female, who both have the same level of latent perceived physical ability. If, when responding to item X, the males are more likely to choose ‘strongly agree’, but the female participants are more likely to choose ‘disagree’ it is said that item X exhibits DIF.

Scheuneman and Bleistein (1994) proposed that the item threshold approach is a common procedure used for evaluating DIF within the context of the IRT. This approach focuses on the difference between the threshold values (difficulty levels) of the item in the two groups of interest. If the difference in the item threshold values is noticeably large, it implies that the item is particularly difficult for members of one of the groups being compared, not because of their different levels of the underlying latent trait, but due to other factors probably related to being members of that group. With the item threshold approach, an item found to be more difficult for a group than the other items in a test is biased against that group.

Through use of the item threshold approach, items that are unexpectedly hard, as well as those unexpectedly easy for a particular subgroup, can be identified. Thorndike (1982) proposed that in order to compare the difficulty of the items in a pool of items for two (or more) groups, it is necessary first to convert the raw percentage of correct answers for each item to a difficulty scale in which the units are approximately equal. The simplest procedure is probably to calculate the Rasch difficulty scale values separately for each group. If the set of items is the same for each group, the Rasch procedure has the effect of setting the mean scale value at zero within each group, and then differences in scale value for any item become immediately apparent. Those items with largest differences in a scale value are the suspect items. Adams and Khoo (1993) proposed that an item whose difference in standardized item threshold between any of the groups fall outside the predefined range of -2.00 to + 2.00 (i.e., st (d1 - d2)> ± 2.00) should be considered a biased item.

RESULTS

Table 1 summarizes means, standard deviations, inter-item correlations, and item-total correlations of the PPASC items for male and female groups. The means of items for males were generally higher than those for females; however, the standard deviations of the items in the male group were lower than those in the female group, indicating that female participants had a greater variation in their responses to the items than their male counterparts. An eye-ball comparison showed that the inter-item and the item-total correlations were generally greater in the male group than those in the female group.

Table 1: Means, standard deviations, inter-item correlations, and item-total correlations of the PPASC items for male and female groups

	Items	**Inter-item correlations**						**M**	**SD**	**Item –total correlations**
		1	**2**	**3**	**4**	**5**	**6**			
Males	1							3.33	.44	.66
(n= 120)	2	.62	1.0					3.37	.39	.69
	3	.69	.67	1.0				3.44	.55	.68
	4	.72	.76	.66	1.0			3.40	.67	.73
	5	.65	.63	.59	.74	1.0		3.30	.60	.69
	6	.60	.68	.74	.70	.67	1.0	3.27	.52	.72
Females	1	1.0						2.91	.75	.62
(n = 100)	2	.54	1.0					2.87	.80	.64
	3	.55	.47	1.0				3.19	.89	.58
	4	.44	.52	.49	1.0			3.16	.94	.60
	5	.46	.56	.53	.48	1.0		2.85	.77	.55
	6	.49	.55	.44	.46	.56	1.0	2.79	.70	.53

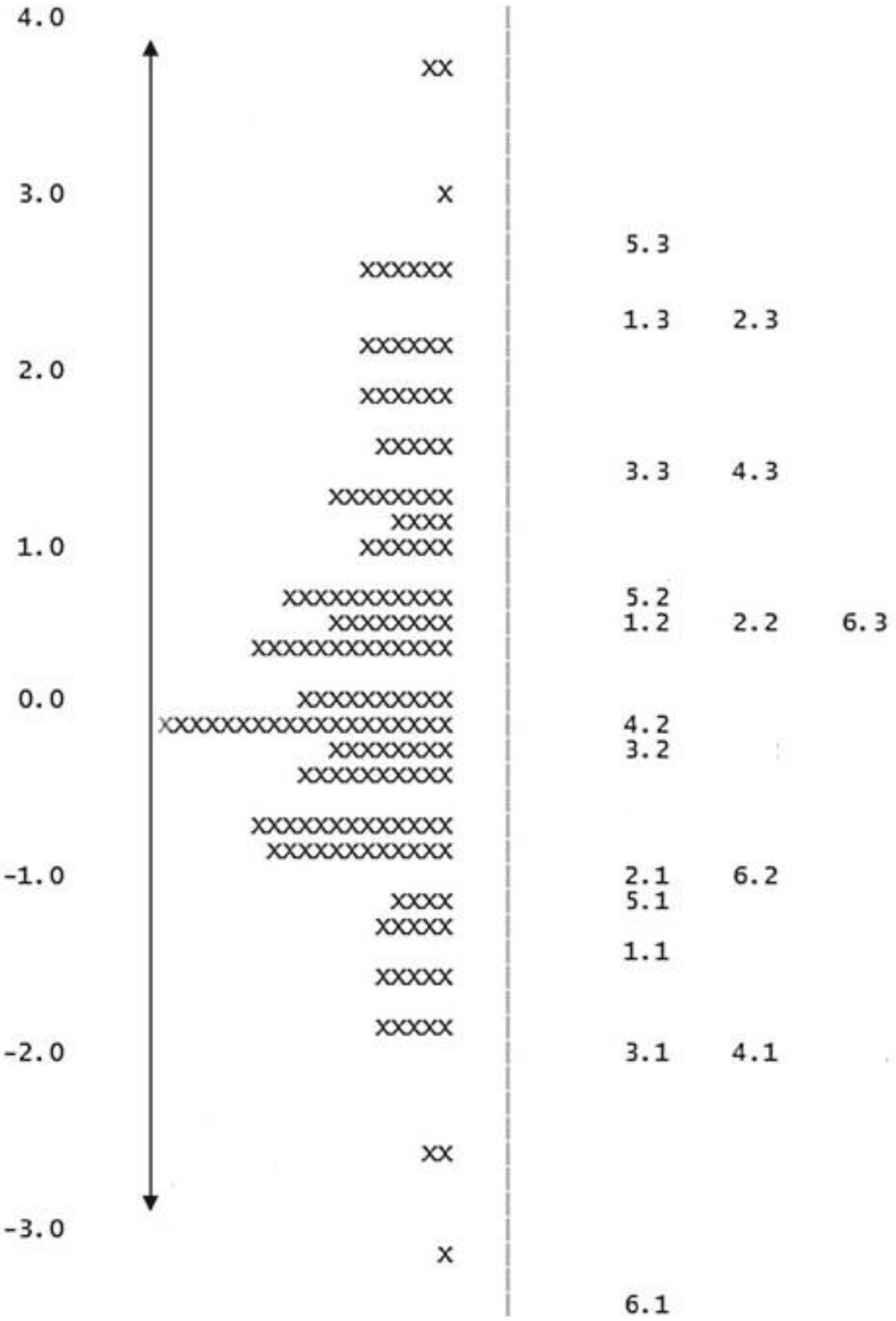
Note: N = 220, p < .001 for all instances

Rasch analysis

The Quest program (Adams & Khoo, 1993) was used to assess the PPASC to obtain the Rasch person-item map presented in Figure 1. Self-reported ratings of perceived physical ability in response to the PPASC items are shown on the left hand side of the map, while the thresholds of the items of the overall PPASC are on the right hand side. Numerical values on the extreme left hand side of the map which range from −3 to +4 are expressed as a log odd unit interval or logit which is the natural unit of the Rasch scale.

The Rasch person-item map is used to compare the range and position of the person measure distribution on the left hand side of Figure 1 to the item measure on the right hand side. Persons represented in the map as an X appear in ascending order of PPASC from the bottom of the figure to the top. Items on the right are represented by item numbers, with a decimal representing the response scale boundary or threshold of each of the ratings (Adams & Khoo, 1993). Items at the top of the scale on the right hand side are harder for children to rate as 4 (e.g., I run very fast), while items become easier (e.g., I run very slowly) for children further down the scale. Children with higher perceived physical ability at the top of the scale are more likely to rate the PPASC items as being almost always true (e.g., I move very rapidly); students with lower perceived physical ability at the bottom of the scale are more likely to rate the items as being not true or almost never true for them (e.g., I move very slowly).

The vertical scale produced by Quest program is an interval scale. Spaces between items, between persons, and between items and persons have substantive meaning in terms of the underlying variable (Callingham & Bond, 2006). The perceived physical ability of each child while rating the statements is referred to as the ‘person measure’ and the level of perceived physical ability while performing each item with a criterion level of difficulty is called an ‘item measure’ (Adams & Khoo, 1993). Items should be located at each point on the scale to measure meaningful differences and must cover all the areas on the scale to measure the perceived physical ability of all children. Rasch rating scale structure parameters, the step calibrations or Tau’s, are related directly to category probabilities. These probabilities relate to the probability of a category being observed, not to the substantive order of achievement of the categories (Linacre, 1999).



Note. Each x represents 1 participant

Figure 1: PPASC item estimates (thresholds)

In Figure 1, both students and items appear along the same scale with the six items of the PPASC forming a unidimensional scale. The range of item difficulties approximately matches the range of students' scores, implying that the scale is appropriate for this group of students. Item 5 (response 4) is seen as the most difficult item in the PPASC while item 6 (response 1) is the easiest item. Three students at the higher end of the scale do not have any corresponding items, implying that while they have high levels of perceived physical ability, the actual level cannot be estimated accurately because of the paucity of item thresholds. Likewise, three students at the lower end of the scale who do not have any corresponding items from the PPASC have a low level of perceived physical ability, which has not been estimated with corresponding items.

Table 2 shows the PPASC items and their INFIT and OUTFIT statistics. Curtis and Boman (2007) consider that OUTFIT statistics scores are more sensitive to outliers. A close investigation for case outliers should occur if an item shows an acceptable fit on one index but marginal or poor fit on the other. Item fit statistics, person fit statistics, and a detailed analysis of item thresholds and function provide evidence to accept or reject a misfitting item (Curtis & Boman 2007). The analysis showed that the PPASC fitted the Rasch model with the six items falling within the expected values of .72 to 1.30 logits (Bond & Fox, 2007).

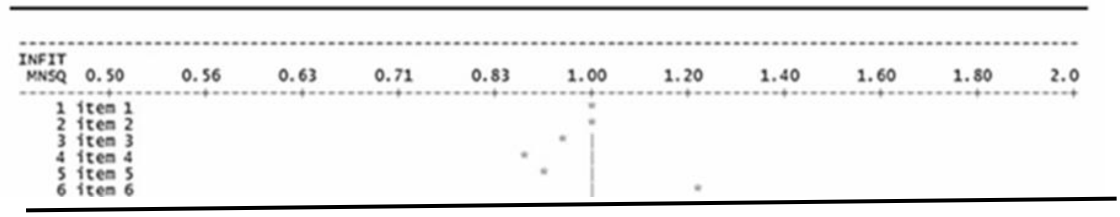
On the other hand, the difficulty error of an item depends on the position of the distribution of students providing the data relative to the location of the item. In Table 2, Item 2 is considered as the most difficult item in the questionnaire, but this is not reflected in the item fit diagram (Figure 1). A possible explanation could be that there was not enough information gathered on the number of students being able to provide a probable answer for Item 2. Figure 2 provides a visual diagram showing item fits, with figures corresponding to the INFIT mean square values presented in Table 2.

Rasch modeling has the advantage of applying the same analytical logic, and therefore the same logic of interpretation, to persons as it does to items. For a well-matched test, the mean person estimate (that is, the group average) would be closer to 0. The mean obtained for the PPASC was +0.39, which is an indicator that the sample found this subscale comparatively easy. This is also compounded by a high standard deviation of 1.26, signifying that there is a greater spread of person measures than item measures.

Table 2: Rasch fit statistics for the PPASC items

Item	Difficulty (error)	Tau 1 (error)	Tau 2 (error)	Tau 3 (error)	INFIT MNSQ	OUTFIT MNSQ
Item 1	.48 (.11)	- 1.61 (.23)	- .03 (.19)	1.64 (.29)	1.00	.97
Item 2	.77 (.11)	- 1.31 (.21)	- .24 (.19)	1.55 (.29)	1.00	.95
Item 3	-.24 (.11)	- 1.52 (.29)	.06 (.19)	1.46 (.22)	.95	.95
Item 4	-.19 (.11)	- 1.66 (.29)	.28 (.19)	1.38 (.23)	.87	.83
Item 5	.73 (.11)	- 1.61 (.21)	- .16 (.19)	1.77 (.33)	.92	.89
Item 6	- .19 (.12)	- 2.04 (.62)	.42 (.21)	1.62 (.19)	1.23	1.12

Note. *N* = 220



```
------------------------------------------------------------------------------------------------
INFIT
MNSQ  0.50      0.56      0.63      0.71      0.83      1.00      1.20      1.40      1.60      1.80      2.0
------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+
 1 item 1                                               *
 2 item 2                                               *
 3 item 3                                           *   |
 4 item 4                                     *         |
 5 item 5                                         *     |
 6 item 6                                               |          *
```

Figure 2: Item fit for the PPASC

The separation reliability coefficient

The separation reliability is the Rasch analogue to the Cronbach alpha (Wright & Masters, 1982). The person separation reliability differentiates persons on the measured variable and replicates placement of persons across other items measuring the same construct, while the item separation reliability identifies a distinct hierarchy of items across other samples (Wright & Masters 1982). The index ranges from 0 to 1, with values equal to or greater than .80 being regarded as acceptable (Fox & Jones, 1998). Although in the present analysis, the person and the item separation reliability were above the criterion of .80, the person separation reliability score was higher (.89) than the item separation reliability (.84), indicating that the behavior of persons people was consistent but the behavior of items was less consistent.

Differential item functioning

Table 3 shows the results of the comparison analyses carried out using the QUEST computer program (Adams & Khoo, 1993) for males and females over the PPASC items. The items threshold values for males (d1) ranged from – 1.49 to +.97 whereas the items threshold values for females (d2) ranged from – 1.35 to + 87. The difference between the threshold values of the PPASC items for males and females (d1-d2) ranged from - .14 to + .10. The standardized difference between the threshold values of the PPASC items for males and females {st(d1-d2)} ranged from – 1.26 to + 1.75, suggesting that the PPASC items function equivalently across males and female groups. A pictorial presentation of the information presented in the Tables 3 is provided in Figure 3. The figure is a plot of the standardized differences generated by the QUEST computer program for comparison of the performance of males and females over the PPASC items.

```
                          Easier for males   Easier for females
-6       -5       -4       -3     -2       -1         0       1          +2          +3        +4          +5         +6
-+---------+----------+---------+--------+---------+--------- +---------+-----------+-------------+------------+-----------+-----------
+                                |                 :              |
Item1                            |          *      :              |
                                 |                 :              |
Item2                            |                 :        *     |
                                 |                 :              |
Item3                            |        *        :              |
                                 |                 :              |
Item4                            |    *            :              |
                                 |                 :              |
Item5                            |                 :          *   |
                                 |                 :              |
Item 6                           |                 :            * |
```

Figure 3: Plot of the standardized differences for male and female groups for all items of the PPASC

Table 3: The standardized threshold values of the PPASC items for male and female groups

Items	Males (d1)	Females (d2)	d1-d2	st(d1-d2)
Item 1	.52	.55	- .03	- .74
Item 2	.81	.74	.07	1.20
Item 3	- .37	- .41	- .04	- .86
Item 4	- .38	- .30	- .08	- 1.26
Item 5	.97	.87	.10	1.49
Item 6	- 1.49	- 1.35	- .14	1.75

A negative value of difference in item threshold (or difference in standardized item threshold) indicates that the item was relatively easier for the male group than for the female group, while a positive value indicates the opposite. Using this criterion, Table 3 shows that items 1, 3, and 4 were easier for males, whereas items 2, 5 and 6 were easier for females. However, it is important to remember that a difference between threshold values of an item for males and females may not be sufficient evidence to imply bias for or against a gender.

Gender differences in perceived physical ability

An independent-samples t test (t = 5.97, df = 218, p < .001) showed that males (n = 120, M = 3.35, SD = .33) reported higher levels of perceived physical ability than females (n = 110, M = 2.96, SD = .39). This mean difference had an effect size of .80 using Cohen’s d effect size. Values of .2, .5, and .8 are considered small, medium, and large effects, respectively (Cohen, 1988). Note that the mean of six items, on a 4- point scale, was taken to represent the mean perceived physical ability for male and female groups.

Correlational analysis

Pearson’s correlation showed that PPASC and PPAS correlated significantly at .61 (p < .001).

DISCUSSION

One important finding of this study is that the Rasch analysis supported the unidimensionality of the factorial structure of the PPASC as a measure of a single underlying latent trait of perceived physical ability. This pattern of results is consistent with Colella et al.'s (2008) original findings of the PPASC, and also replicates the findings from research conducted internationally on the factorial structure of the PPASC in England (Draun & Stevens, 2009), Canada (Rabi & Swanson, 2010), and France (Carmen & Shineder, 2011).

The Rasch person-item map presented in Figure 1 showed that half of the 220 participant students were located above the item mean of 0.00 which was set by default as the mean of the items, with 44 students located above +1.0 logits showing high levels of perceived physical ability, and 15 students above +2.0 logits being extremely high. The perceived physical ability levels of the three students at the top of the scale were estimated with greater error because of a lack of corresponding items. Similarly, the scores of the three students with low levels of perceived physical ability at the bottom of the scale were likely to be estimated with greater error as there was only a single item (item 6) at this level. The higher person separation index of the PPASC indicates that students responded to the rating scale consistently. Figure 1 also shows that students found it easiest to endorse item 6 (I feel very insecure when I move) and hardest to endorse item 5 (I don’t feel tired at all when I move).

The Rasch analysis also demonstrated that the PPASC was equivalent across gender; all items did not display significant DIF between males and female groups. This finding indicates that the PPASC items are not influenced by external variables such as gender and that students with the same level of perceived physical ability have equal probability of item success. This is an important finding because unless there is reasonable support for the invariance of the PPASC items across gender, it may not be justified to pool data across male and female children. This finding supports the validity of inferences regarding gender differences for the PPASC.

There were significant gender differences in perceived physical ability favoring males. This finding is consistent with the findings of Colella et al. (2008). However, this finding is at odds with Draun and Stevens (2009) and Rabi and Swanson (2010) who reported gender differences in perceived physical ability favoring females. This finding can be interpreted within the cultural values of Oman which is seen as a masculine, conservative society. Perceived greater physical ability in males may be a means to maintain dominance and express adherence to masculine gender norms. As such, physical activity may be seen as a training ground for manhood. By expressing high levels of perceived physical ability, male students are able to demonstrate components of masculinity including vigour, hard work, competition, aggression, toughness, dominance, and physicality.

The PPASC correlated strongly with PPAS, suggesting convergent validity of the PPASC. This finding fits with the principle behind self-efficacy, that individuals are more likely to engage in activities for which they have high self-efficacy and less likely to engage in those they do not (Stajkovic & Luthans, 1998). This finding is consistent with the findings of several researchers in the field of physical education (e.g., Xiang, Lee, & Williamson, 2001; Parish & Treasure, 2003) who have posited that the strength and quality of students' outcomes (e.g., effort, persistence, performance) are closely linked to their beliefs about their own competence, with self-efficacy being conceptualized as a determining factor for behavior. In physical education, those with higher self-efficacy were found to be more likely to perform better, expend more effort on mastery tasks, and persevere longer when they encounter challenges than those with lower self-efficacy (Gao et al., 2009, 2010; Gao, Newton, & Carson, 2008). For example, Gao et al. (2011) reported that self-efficacy predicted 27% of the variance in physical activity among 194 middle school students in physical education classes when it was set as the sole predictor, and predicted 28% of the variance in physical activity when it was set as a predictor along with mastery-approach goal and mastery motivational climate variables.

In summary, the development of the Arabic version of the PPASC is one of the strengths of this study. The Rasch analysis demonstrated that the PPASC had acceptable psychometric properties as a unidimensional measure of children's perceived physical ability. The PPASC functioned well and equivalently across gender. Although future studies are needed to replicate these results in additional settings, our findings suggest that researchers and practitioners can be confident in their interpretation of the PPASC scores when used with samples containing males and females.

REFERENCES

Abd-El-Fattah, S. M. (2007). Is the Aggression Questionnaire bias free? A Rasch analysis. *International Education Journal*, 8, 237-248.

Abd-El-Fattah, S. M. (2013). A Cross-cultural examination of the Aggression Questionnaire–Short Form among Egyptian and Omani Adolescents. *Journal of Personality Assessment*, 95, 539-548.

Adams, R. J., & Khoo, S. T. (1993). QUEST: *The Interactive Test Analysis System*. Hawthorn, Victoria: Australian Council for Education Research.

Al-Saharti, H. & Al-Mahroky, M. (2008). The relationship between family structure and participation in physical activity. *Psychological Research*, 2, 78-93.

Al-Saidi, S. (2010). The effect of physical fitness on the activity of heart among Omani youth. *Health and Fitness*, 12, 73-87.

Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Andrich, D. (1988). *Rasch models for measurement.* Beverly Hills: Sage Publications.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. NewYork: Freeman.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences.* New Jersey: Lawrence Erlbaum.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* Vol. 2, pp. 389-444. Boston: Allyn and Bacon.

Callingham, R., & Bond, T. (2006). Research in mathematics education and Rasch measurement. *Mathematics Education Research Journal*, 18, 1-10.

Carmen, D., & Shineder, M. (2011, August). *Validation of the French version of the Perceived Physical Ability Scale for Children.* Paper presented at the First International Conference on Sport Psychology. Madrid, Spain.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, 4, 62-83.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale: Lawrence Erlbaum Associates.

Colella, D., Morano, M., Bortoli, L., & Robazza, C. (2008). A Physical Self-Efficacy Scale for Children. *Social Behavior and Personality*, 36, 841-848.

Curtis, D. D., & Boman, P. (2007). X-ray your data with Rasch. *International Education Journal: Comparative Perspectives*, 8, 249-259.

Draun, D., & Stevens, A. (2009). Factors affecting children's physical self-efficacy and attitudes towards diet. *Journal of Health, Physical Education, Recreation*, 3, 58-70.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counselling

psychology research. *Journal of Counseling Psychology*, 45, 30–45.

Gao, Z., Lee, A. M., Kosma, M., & Solmon, M. A. (2010). Understanding students' motivation in middle school physical education: Examining the mediating role of self-efficacy on physical activity. *International Journal of Sport Psychology*, 41, 199-215.

Gao, Z., Lochbaum, M., & Podlog, L. (2011). Self-efficacy as a mediator of children's achievement motivation and in-class physical activity. *Perceptual and Motor Skills*, 113, 969-981.

Gao, Z., Lodewyk, K., & Zhang, T. (2009). The role of ability beliefs and incentives in middle school students' intentions, cardiovascular fitness, and effort. *Journal of Teaching in Physical Education*, 28, 3-20.

Gao, Z., Newton, M., & Carson, R. L. (2008). Students' motivation, physical activity levels, and health-related physical fitness in fitness class. *Middle Grades Research Journal*, 3, 21-39.

Hambleton, R. K., & Swaminathan, H. (2010). *Item response theory: Principles and applications*. Boston, MA: Kluwer Nijhoff Publishing.

Hills, A. P., King, N. A., & Armstrong, T. P. (2007). The contribution of physical activity and sedentary behaviors to the growth and development of children and adolescents. *Sports Medicine*, 37, 533-545.

Information Centre. (2006). *Statistics on obesity, physical activity and diet: England, 2006*. Retrieved April 20, 2013, from http://www.ic.nhs.uk/webfiles/publications/opan06/OPAN%20bulletin%20finalv2.pdf

Kimm, S. Y. S., & Obarzanek, E. (2002). Childhood obesity: A new pandemic of the new millennium. *Pediatrics*, 110, 1003-1007.

Lee, O., & Martinek, T. (2009). Navigating two cultures: An investigation of cultures of a responsibility based physical activity program and school. *Research Quarterly for Exercise and Sport*, 80, 230-240.

Linacre, J. M. (1999). Category disordering vs. step (threshold) disordering. *Rasch Measurement Transactions*, 13, 675.

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing Research*, 48, 175-186.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Ministry of Health (2010). *Annual report on obesity and physical activity in Oman 2010*. Retrieved January 27, 2013, from http://www.moh.gov.om/en/nv_menu.php?o=hr/majorprojects.htm&SP=1.pdf.

Nakamura, Y. (2002). Beyond the hijab: Female Muslims and physical activity. *Women in Sport Physical Activity Journal*, 11, 21-48.

Oettingen, G. (1995). Cross-cultural perspective on self-efficacy. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 149-176). New York: Cambridge University Press.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning*. Thousand Oaks, CA: Sage Publishing.

Pajares, F., & Miller, D. M. (1995). Mathematics self-efficacy and mathematics performance: The need for specificity of assessment. *Journal of Counseling Psychology*, 42, 190-198.

Parish, L. E., & Treasure, D. C. (2003). Physical activity and situational motivation in physical education: Influence of the motivational climate and perceived ability. *Research Quarterly for Exercise and Sport*, 74, 173-182.

Rabi, H., & Swanson, T. (2010). Psychometric evaluation of the Perceived Physical Ability Scale for Children in Canada. *Journal of Sport Behavior,* 5, 13-28.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogiske Insitut.

Sanigorski, A. M., Bell, A. C., Kremer, P. J., & Swinburn, B. A. (2007). High childhood obesity population in an Australian population. *Obesity,* 15, 1908-1912.

Scheuneman, J. D., & Bleistein, C. A. (1994). Item bias. In T. Husén & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (2 ed., pp. 3043-3051). Oxford: Pergamon Press.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124, 240-261.

Thorndike, R. L. (1982). *Applied Psychometrics*. Boston, MA: Houghton-Mifflin.

Tittle, C. K. (1994). Test bias. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2 ed., pp. 6315-6321). Oxford: Pergamon Press.

van der Linden, W. J., & Hambleton, R. K. (2010). *Handbook of modern item response theory*.New York: Springer.

Vertinsky, P., Batth, I., & Naidu, M. (1996). Racism in motion: Sport, physical activity and the Indo-Canadian female. *Avante*, 2, 1-23.

Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment,* 29, 364-376.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.

Xiang, P., & Lee, A. M. (2002). Achievement goals, perceived motivational climate, and students' self-reported mastery behaviors. *Research Quarterly for Exercise and Sport*, 73, 58-65.

Yim, H. Y. B., Abd-El-Fattah, S. M., & Lee, L. W. M. (2007). A Rasch analysis of the Teachers Music Confidence Scale. *International Education Journal*, 8, 260-269.

Zumbo, B. D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly,* 4, 223-233.