Psychometric Properties of the Arabic Version of the Perceived Physical Ability Scale for Children in Oman

Sabry M. Abd-El-Fattah*
Sultan Qaboos University, Oman

This study aimed at examining the psychometric properties of the Perceived Physical Ability Scale for Children (PPASC) within an Omani context. A sample of 250 fourth graders (130 males and 120 females) from six public primary schools in four governorates in Oman responded to Arabic translated version of the PPASC. Data on students' in-class physical activity, social desirability, and teachers' nominations of students' physical ability were also collected. Confirmatory factor analysis revealed that the unidimensional model of the PPASC had adequate fit to the data. Multigroup CFA revealed gender invariance for PPASC unidimensional model when it was compared across male and female groups. A latent mean analysis showed that there were significant gender differences in perceived physical ability favoring males. The PPASC correlated positively with in-class physical activity and teachers' nominations of students' physical ability but not with social desirability.

Keywords: children, physical activity, physical ability, teachers' nominations, confirmatory factor analysis.

الخصائص السيكومترية للنسخة العربية من مقياس إدراك القدرة البدنية للأطفال في سلطنة عمان

صبري محمود عبد الفتاح*

جامعة السلطان قابوس، سلطنة عمان

هدفت الدراسة الحالية إلى فحص الخصائص السيكومترية لمقياس إدراك القدرة البدنية للأطفال في سلطنة عمان، حيث أجاب على مفردات المقياس عينة مكونة من ٢٥٠ طالبا من طلبة الصف الرابع (١٣٠ ذكور و ١٢٠ إناث) موزعين على ست مدارس إبتدائية حكومية في أربع محافظات. و كذلك تم جمع بيانات متعلقة بالنشاط البدني لأفراد العينة في الفصل ومستوى مرغوبيتهم الإجتماعية وترشيحات المعلمين لقدرتهم البدنية. وأظهر التحليل العاملي التوكيدي أن نموذجا أحادي البعد يمثل مقياس إدراك القدرة البدنية للأطفال قد تطابق مع بيانات الدراسة بصورة جيدة. وأظهر التحليل العاملي التوكيدي متعدد المجموعات أن هذا النموذج متكافئ بين الذكور والإناث. كما أظهر تحليل المتوسطات الكامنة أن الذكور يتفوقون على الإناث في إدراكهم لقدرتهم البدنية. وقد أظهر معامل ارتباط بيرسون أن مقياس إدراك الأطفال لقدرتهم البدنية يرتبط ايجابيا وبصورة دالة إحصائيا بالنشاط البدني للأطفال في الفصل وبترشيحات المعلمين لقدرتهم البدنية بينما لم يترتبط بصورة دالة إحصائيا بمستوى مرغوبيتهم الإجتماعية.

الكلمات المفتاحية: الأطفال، النشاط البدني، القدرة البدنية، ترشيحات المعلمين، التحليل العاملي التوكيدي.

*sabryamin@squ.edu.om

Inactivity, or sedentary behavior, is of particular concern among children because physical activity is necessary for optimal growth and development (Hills, King, & Armstrong, 2007). Presently, physical activity rates for children are insufficient for health benefits, whereas inactivity-related diseases like obesity are on the rise in many countries, including Canada (Hills et al., 2007), 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 youth accumulate 90 minutes of moderate- to vigorous physical activity per day. However, over 90% of Omani youth aged 6 to 20 years do not meet these guidelines (Ministry of Health, 2010). To better understand the patterns of activity and inactivity among children, numerous psychological factors must be considered. One of these factors is children's perceived physical ability; also known as physical activity self-efficacy.

Self-efficacy describes ''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 they do not (Stajkovic & Luthans, 1998), thus, self-efficacy functions as a self-fulfilling prophecy. Bandura (1997) proposed that self-efficacy beliefs are shaped by the cognitive processing and integration of four main sources of information: (1) performance attainments and failures-what we try to do and how well we succeed or not; (2) vicarious performances- what we see other people do; (3) verbal persuasion- what people tell us about what we are able or not able to do; and (4) imaginal performances-what we imagine ourselves doing and how well or poorly we imagine ourselves doing it.

Self-efficacy beliefs are presumed to have actual task ability as an important underlying determinant. In other words, someone who typically does well on a task knows that he or she does well and shows this knowledge in his or her self-efficacy ratings. However, beliefs in one's self-efficacy are not just based on a simple knowledge of one's level of ability. Those self-beliefs go beyond actual capability, being ''instrumental in determining what individu-

als 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, someone high in self-efficacy might do better because that person approaches a task with a different mindset than does someone low in self-efficacy, even though both might reach the same level of ability.

Self-efficacy 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 important, in their beliefs about their success rates. The domain might be related to good performance in a course in mathematics, biology, or language. The domain might concern doing well at tasks involving physical strength, eye-hand coordination, or memory. The domain could even represent maintaining successful relationships, being a good parent, or sticking to a diet. Within any one domain of performance, self-efficacy beliefs can be measured with respect to diverse arrays of accomplishments differing in breadth. Consider the domain of physical fitness. At a very narrow level, one could measure self-efficacy for performing a specific physical exercise. At a broader level, self-efficacy could be assessed with regard to passing Standing Stork Tests (tests of an athlete's ability to maintain a state of equilibrium in a static position). Even broader yet would be to evaluate beliefs about one's physical fitness aptitude.

There are even some measures of self-efficacy that are so broad in scope that they do not refer to any specific performance domain. Instead, such global measures refer to general competence and life coping skills, with items related to ac-accomplishing goals in general and performing effectively on different tasks (e.g., Chen, Gully, & Eden, 2001). Bandura (1997) has maintained that global self-efficacy measures ''violate the basic assumption of the multidimensionality 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. So if the criterion is, for example, an obtained score on a Standing Stork Test, then the self-efficacy measure should represent a person's beliefs about his or her performance on that narrow task. On the other hand, if the criterion is overall physical fitness, then the self-efficacy measure should be broader, referring perhaps to a person's expectations about his or her performance on a physical fitness aptitude test.

Perceived physical ability

One important domain to examine individual's self-efficacy beliefs is the domain of physical activity. Colella, Morano, Bortoli, and Robazza (2008) defined perceived physical ability as one's confidence to participate in physical activity (i.e., task), overcome physical activity related barriers (i.e., barriers), and organize time and responsibilities around physical activity (i.e., scheduling). Self-efficacy beliefs applied to physical activity because research has repeatedly shown that self-efficacy is a critical antecedent to motivated behaviors in physical activity settings. Specifically, these studies have shown that self-efficacy is linked to better performance in physical activity tasks, expending of more effort on mastery tasks, and persevering when encounter challenges (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.

The perceived physical ability scale for children (PPASC)

To further quantify self-efficacy beliefs in physical activity among school children, Colella et al. (2008) developed the Perceived Physical Ability Scale for Children (PPASC) that measures children's beliefs about their physical capabilities to successfully engage in and perform physical activities. The scale consists of 6 items that represent strength, speed, and coordinative abilities 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 re-

versed. The total test score can range from 1 to 24. High scores would indicate a high perceived physical ability, whereas low scores would reflect a low perceived physical ability.

In developing the PPASC, Colella et al., (2008) administered the 6 items to 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; the 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 (i.e., 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 the 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 well the data from 280 children aged between 9 and 12 years only after correlating the error terms of Items 1, 3 and 6.

Gender differences on the PPASC

With respect to gender differences on the PPASC, Colella et al. (2008) demonstrated that females reported lower perceived physical ability scores than males. Age main effect and age by sex interaction effect were not statistically significant. In contrast, Draun and Stevens (2009) and also Rabi and Swanson (2010) reported gender differences in perceived physical ability favoring females. Carmen and Shineder (2011) found nonsignificant differences

in perceived physical ability scores using the French version of the PPASC.

Rationale and aim of the present study

Cultural values and norms in relation to physical activity and the unique features of socialization practices 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 determinants 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 and attitudes toward physical activities and in turn, actual physical activity behaviors. Culture values and norms may also affect not only the type of information provided by the various sources of self-efficacy (see Bandura, 1997), but also which information is selected and how it is weighted and integrated in individuals' self-efficacy judgments. For example, individuals in an individualist culture focus their self-appraisals of efficacy on information concerning their personal performance attainments (e.g., improvement or decline), whereas individuals in collectivist cultures consider the evaluation by in-group members as the most important source of efficacy information, with modeling by other in-group members also being influential. Furthermore, in an individualist society, when approaching a new task, an individual's self-appraisal of efficacy would focus on past personal successful performances on similar tasks, whereas in a collectivist society, an individual's self-appraisal of efficacy would center on the in-group's beliefs that he or she has the capabilities to perform the task, and whether other members of the in-group might have higher capabilities to perform better on that task (Bandura, 1986; Oettingen, 1995). In line with the argument concerns the effects of cultural values on both physical activity and self-efficacy, the validity of the PPASC needs to be assessed in a non-Western context because it is possible that some instruments developed in the West might not work properly in non-Western settings due to cultural differences (Maneesriwongul & Dixon, 2004).

Given this argument, there are two areas in which the validation attempts of the PPASC are lacking. First, it is not clear whether the factorial structure of the PPASC as originally developed by Colella et al. (2008) within a Western context will replicate within a non-Western context. Second, it is not clear whether gender differences in perceived physical ability as measured by the PPASC are due to the fact that male and female children differ in the underlying latent trait of perceived physical ability or whether these differences represents an artifact methodology because the items of the PPASC functions 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 justified to pool data across male and female participants. Tittle (1994) noted that the examination of test items for bias towards groups is an important part in the evaluation of the overall instrument as it influences not only testing decisions, but also the use of the test results. Under these circumstances it is clearly 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. 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).

Overall, the present study has five aims: (a) to examine the factorial structure of the PPASC within an Omani context using confirmatory factor analysis (CFA), (b) to examine the DIF of the PPASC items across male and female groups using multigroup confirmatory factor analysis (MCFA), (c) to examine the convergent validity of the PPASC using correlational analysis with a measure of in-class physical activity and teacher nominations, (d) to examine the discriminant validity of the PPASC using correlation analysis with a measure of social desirability, (e) to examine gender differences in perceived physical ability using latent mean analysis. The results from this study can contribute to the body of literature on perceived physical ability by providing strong and necessary empirical evidence of construct, convergent, and discriminant validity, as well as validity of inferences regarding gender differences for the PPASC.

The context of the present study

The present study is conducted within the context of Oman because we know little about the levels of the perceived physical ability amongst native Omani children. It is possible that the paucity of research on perceived physical ability in Omani children has, in part, been due to the lack of Arabic language measures with acceptable psychometric properties and also to the fact that many Omani children do not have an adequate command of the English language for the use of English language questionnaires. Thus, there is a lack of an even more rapidly applicable and reliable measures of perceived physical ability in the Arabic-speaking context; a tool not only valid but also useful for providing information about children's beliefs about their physical capabilities to successfully engage in and perform physical activities. Doing so, it is also possible to facilitate and promote perceived physical ability research with Arabic-speaking children.

Although the PPASC was originally developed in the West, there are several meaningful characteristics of Oman that may strengthen the cross-cultural utility of the measurement model of the perceived physical ability as described in the PPASC. Oman is a developing, Muslim, Arab country. The Islamic code is an enthusiastic advocate of the health and care of the body. The Prophet Muhammad, peace be upon him, was particularly keen on the physical education of the children primarily to improve the children's health and to develop their mental and physical faculties. He therefore urged parents to teach their children reading and writing, swimming and archery. He commanded that the right of the child is one of obliging his or her parents to teach him or her writing and physical activities. While reading and writing were considered important for learning the holy Qur'an, it was hoped that physical activities such as swimming would promote the health of the children (Benn, Dagkas, & Jawad, 2011; Dagkas & Benn, 2006). Furthermore, from an academic perspective, Omani schools offer at least two physical education classes every week, each is 45 minutes long. These classes focus mainly on moderate and vigorous physical activities.

METHODS

Participants

Subjects of the present study included 250 Omani children (130 males and 120 females) from six public primary schools in four governorates in Oman (Muscat, Al Batinah South; Ad Dakhiliyah, and Dhofar). Three schools were from metropolitan areas and 3 schools were from rural areas, and all schools had mixed-gender populations. The public education system in Oman consists of primary school (6 years; 6 to 12 years old), preparatory school (3 years, 12-15 years old), and secondary schools (3 years; 15-18 years old), and the tertiary level. All participant students were at Year Four. The means and the standard deviations of the sample ages were 10.4 and .63 for boys and 10.1 and .39 for girls. The percentage of missing data was 2%. Those students left several item blank on the PPASC forms. Only students with complete data were retained for the present study. Arabic was the native language of all participants.

Measures

The PPASC: The PPASC (Colella et al., 2008) is a self-report measure which consists of 6 items that represent strength, speed, and coordinative abilities 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 1 to 24. High scores would indicate a high perceived physical ability, whereas low scores would reflect a low perceived physical ability. Based on the dataset of the present study, the PPASC has a Cronbach alpha of .86.

The author translated the PPASC from English into Arabic using the back-translation method. Three other qualified translators, working without referencing 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 to the new English version that was translated back from Arabic, and any discrepancies were noted. This iterative process of translation and back-translation continued for 2 rounds until

no semantic differences were noticed between both scale forms (Brislin, 1980).

Teachers' nominations: Three physical education teachers at each school were asked to independently nominate the participant students on an Arabic translated version of the PPASC during their normal physical education classes at their schools. On average, the teachers had more than 5 years of teaching experience (M = 5.7 year, SD = .65) and they were familiar with the participant students because they shared the responsibility of teaching three physical education classes for the participant students on every other day. The teachers were allowed to nominate as many students as they wished on each item of the PPASC and cross-sex nominations were allowed. For any given item of the PPASC, a student would be given a score of 1 if he or she got nominated by the teacher and a score of zero if not. Thus, the possible score for any student on any item of the PPASC ranged from zero (i.e., the student did not get nominated by any teacher) to 3 (i.e., the student get nominated by all three teachers). The possible score on the PPASC for any student ranged from 0 (i.e., the student did not get nominated by any teacher on any item of the PPASC) to 18 (the student got nominated by all three teachers on all items of the PPASC). Based on the dataset of the present study, teachers' nominations have a Cronbach alpha of .85 and intra-class correlation coefficients that ranged from .85 to .89.

Social desirability: Al-Moseny and Al-Rabie (2008) developed an Arabic short form of the Children's Social Desirability Scale (CSDS, Crandall, Crandall, & Katkovsky, 1991) using two samples of primary and middle school students from Oman and Kuwait. This is a 20-item single-factor measure. An example of the scale items is "I always finish all of my homework on time." Students responded to each item on dichotomous yes–no scale. A yes response was assigned a score of 1, indicating social desirability. A no response was assigned a score of 0, indicating a socially undesirable response. The possible scores for the scale ranged from 0 to 20. Based on the dataset of the present study, the CSDS has a Cronbach alpha of .87.

In-class physical activity: Physical activity monitor devices were utilized for three regularly scheduled physical education classes to measure students' on average in-class physical activity. Each class had approximately 40 minutes of physical activity time. The devices were worn at the students' arms. Activity counts were measured in 15-second epochs to better capture the activity patterns of children, and in-class activity levels were quantified as average activity count per minute (average count/min) for the moderate to vigorous intensity activities. Average activity levels were calculated by averaging the mean activity counts accumulated during moderate and vigorous intensity activities and dividing by the duration of three physical education class (i.e., approximately 120 minutes).

PROCEDURES

Approval was obtained to conduct the research investigation at the schools prior to data collection. Students were recruited to participate in the present study during their normal physical education classes at their schools. All students gave assent by getting their parents signing a consent form prior to their participation in the present study that they are willing to respond to the PPASC and the CSDS, that they will be nominated by their teachers on the PPASC, and that they will wear Physical activity monitor devices in their arms to measure their in-class physical activity during their normal physical education classes at their schools. The PPASC and the CSDS were administered by trained experimenters according to 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 the CSDS and then teachers independently nominated their students on the PPASC. Apparently, the items of the Arabic version of the PPASC were within the age-equivalent reading level of the Omani children because they did not indicate any difficulty understanding their content. The average time required to complete the PPASC and the CSDS was about 15 minutes for students and 8 minutes for teachers to nominate students on the PPASC. Students' in-class physical activity was measured using physical activity monitor devices

for three physical education classes. The devices, along with waistbands, were distributed to students while the teachers were taking roll. Each student was assigned an identification number which matched the number on the waistband. The research assistants helped students attach the waistbands and made sure the waistbands were placed in the right location.

Overview of the analysis

Factor structure

CFA was used to test a global one-factor model that assumes all six items reflect a single general factor of perceived physical activity as suggested by Colella et al. (2008). Several absolute and relative goodness-of-fit indexes were used to evaluate the model goodness-of-fit to the data. Absolute fit indices included Chi-square ($\chi2$), Standardized Root Mean-Square Residual (SRMR), and Root-Mean-Square Error of Approximation (RMSEA). Relative fit indices included Comparative Fit Index (CFI) and Nonnormed Fit Index (NNFI). When modeling normally distributed data, SRMR values of approximately .08 or below, RMSEA values of approximately .06 or below, CFI values of approximately .95 or above, and NNFI of approximately .90 or higher suggest adequate model-data fit (Hu & Bentler, 1998; Vandenberg & Lance 2000). Because the $\chi2$ is sensitive to sample size, Hoelter (1983) recommended reporting the $\chi2/df$ ratio and suggested that ratios below 2.0 indicate a reasonable fit.

Measurement invariance

A decision was made a prior to follow a protocol for conducting a measurement invariance analysis of the PPASC according to how the results came out at each step in the analysis. If the same CFA model fits the data from male and female groups well, additional invariance tests will be conducted to determine if the measure is functioning differentially across the two groups. If, however, a different structure is supported across the two groups, additional invariance testing will not be completed and the focus will turn instead to examine these differences in structure. The invariance testing process involves several steps in which increasingly restrictive levels of measurement invariance are explored. Four levels of measurement invariance will be tested in the following order: (a) configural invariance, (b) metric invariance, (c) scalar invariance, and (d) mean

invariance (Byrne, Shavelson, & Muthen, 1989; Cheung & Rensvold, 2002; Meredith, 1993; Steenkamp & Baumgartner, 1998).

Configural invariance

Configural invariance tests whether the same basic factor structure holds for the two groups. This level of invariance tests only the overall structure: whether the same number of factors is relevant and whether the same items are salient to each factor across groups (i.e., same pattern of fixed and freed loadings). This model then served as a baseline model for comparisons with more restricted models.

Metric invariance

Metric invariance can be tested - if configural invariance is established- by constraining the matrix of factor loadings to be invariant across groups. Metric invariance is met if the strength of the relations between each item and its associated factor are equivalent across groups. If metric invariance is supported, one can conclude that the two groups of interest are interpreting the items in the same way. A lack of metric invariance may imply that some items are more important to the construct for one group than for the other. When full metric invariance is not established, the researcher can determine the source of the noninvariance by freeing, progressively, the loadings in the baseline model for items across the groups, until a final partial metric invariance model is obtained. This final partial metric invariance model will have only those items with equal loadings constrained equally across the groups. When a few items are found to be noninvariant, meaningful cross-group comparisons can still be performed because few items will not heavily influence such comparisons.

Scalar invariance

Scalar invariance tests the equality of intercept term by imposing an equality constraint on the intercepts of the items found to have invariant factor loadings (i.e., metric invariance). At this step, one is essentially testing to see whether the two groups use the response scale in a similar way. For instance, a male student and a female student with the same latent level of perceived physical ability should choose the same response option for a perceived physical ability item if the perceived physical ability item does in fact represent perceived physical ability and functions simi-

larly across the two groups. Scalar invariance is important because it must be satisfied before interpreting mean differences between the groups. If this initial model is not supported, the source of the non-invariance can be explored, using a similar strategy to that described for testing partial metric invariance. The final partial scalar invariance model can be revised appropriately to test for invariance in latent means. In general, for testing invariance for the structural model it is necessary for at least one item to show scalar invariance in addition to the loadings that are equal due to their being fixed at unity (e.g., 1) for identification purposes (Steenkamp & Baumgartner, 1998).

Because each of the increasingly constrained invariance models is nested within the previous models, the change in fit will be assessed by comparing fit indices from one step to the next. Typically, model comparisons are made by examining the change in $\chi 2$ relative to the change in the degrees of freedom of the two compared models (Byrne et al., 1989). However, additional indices have been recommended for comparing nested models. Therefore, in addition to the difference in $\chi 2$ ($\Delta \chi 2$), the change in model fit will be assessed by examining the change in the CFI index (ΔCFI). Decreases of .01 or less in the CFI index will be interpreted to suggest invariance at that step (Cheung & Rensvold, 2002; Wu, Li, & Zumbo, 2007).

Latent mean analysis

If configural, metric, and scalar invariance were supported, one is able to estimate latent mean differences and effect sizes (Hancock, 1997). Byrne et al. (1989) proposed that latent means can still be compared under partial intercept invariance assuming that the latent means will not be affected by noninvariant intercepts to a great extent. A structured means modeling approach was chosen over other mean differences techniques such as ANOVA and t test because in these statistical procedures, group differences are assessed on the linear composites of measured variables while including the measurement error of variables in the composite. On the other hand, structural equation modeling approaches allow for the use of theoretically error-free constructs in tests of group differences. Therefore, they are more powerful than mean differences

techniques as they parse out the measurement error (Thompson & Green, 2006). The AMOS 7.0 program (Arbuckle, 2006) was used to run all the CFA analyses.

Correlation analysis

We used Pearson's correlation to examine the relationships among the PPASC, teacher nominations of students' physical ability, in-class physical activity, and social desirability.

RESULTS

Descriptive statistics

Table 1 summarizes means, standard deviations, the item-total correlations with the item deleted, and the inter-item Pearson correlations of the PPASC. The means of items for males were generally greater 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-balled comparison showed that the inter-item and the item-total correlations were generally greater in the male group than those in the female group. Cronbach's alpha for the PPASC was .89 for the male group, and .85 for the female group. The full information maximum likelihood estimation was used to analyze the variance-covariance matrices, estimate model parameters, and obtain fit indices (Byrne, 2010).

Factorial structure

The CFA results showed that the unidimensional model of the PPASC had a close fit to the data ($\chi 2$ = 15.42, df = 9; p = .08; $\chi 2$/df = 1.71, RMSEA = .04, SRMR = .04, CFI = .98, NNFI = .97). Table 2 shows that the item loadings of the PPASC ranged from .62 to .75 and that the Z statistics associated with these factor loadings were well above 1.96, suggesting that these factor loadings were statistically significant. The Z statistic is distributed as t statistic and is calculated by dividing the unstandardized beta for an item by its associated standard error. The Z statistic is statically significant at .05 when its value is > 1.96 (Byrne, 2010).

Table 1
Means, standard deviations, inter-item correlations with item deleted, and the item total correlations for the 6 items of the PPASC in the male and female groups

	Items	Correlations						M	SD	Item total
		1	2	3	4	5	6			
Males	1	1.00						3.25	.32	.69
(n= 130, α = .89)	2	.66	1.00					3.36	.44	.63
	3	.65	.62	1.00				3.18	.46	.72
	4	.74	.70	.60	1.00			3.57	.52	.70
	5	.70	.69	.70	.71	1.00		3.46	.30	.64
	6	.62	.62	.63	.66	.74	1.00	3.31	.41	.77
Females	1	1.00						2.57	.84	.54
(n = 120, α = .85)	2	.50	1.00					2.46	.70	.58
	3	.52	.54	1.00				2.84	.81	.60
	4	.49	.50	.45	1.00			2.70	.77	.51
	5	.41	.44	.50	.55	1.00		2.44	.65	.57
	6	.53	.47	.56	.51	.46	1.00	2.50	.60	.53

Note: *N* = 250, *p* < .001 for all instances

Table 2
Unstandardized beta, standard error, standardized beta, and z statistics for the 6 items of the PPASC

Item	Unstandardized beta	Standard error	Standardized beta (item loadings)	Z statistic
Item 1	.89	.14	.66	6.35
Item 2	.76	.10	.75	7.68
Item 3	.32	.06	.62	5.30
Item 4	.24	.05	.70	4.92
Item 5	1.10	.15	.65	7.33
Item 6	.74	.12	.72	6.22

Note. $N = 250$, $p < .01$ for all instances of the z statistic

Gender invariance

Configurable invariance

Table 3 shows that the unidimensional model fitted the data reasonably well for male and female groups. We concluded that males and females have the same basic conceptualization of perceived physical ability. More specifically, the satisfaction of configural invariance implies that both males and females distinguished one distinct component of perceived physical ability and viewed the same items as salient to this single factor.

Metric invariance

We tested a model in which the item loadings on their designated factors were constrained to be equal for male and female groups. Table 3 shows that the $\Delta\chi^2$ difference between the constrained model and the unconstrained model (i.e., configural model) was not statistically significant. In addition, the change in the CFI (ΔCFI) was less than the cutoff score of .01. We concluded that metric invariance was established and that the strength of the relations between each item and the underlying latent factor of perceived physical ability are equivalent across males and females.

Scalar invariance

Table 3 shows that the difference between the constrained model and the unconstrained model (i.e., metric invariance model) was not significant. In addition, the change in the CFI was less than the cutoff score of .01. We concluded that scalar invariance was established and that the males and females use the response scale in a similar way. Given that invariance was supported at each of the three levels tested: configural, metric, and scalar (see Table 3), we can feel more confident when comparing mean levels of perceived physical ability in male and female groups.

Table 3
Results of measurement invariance tests for male and female groups [1]

Model	χ²	df	χ²/ df	Δχ²	Δdf	CFI	ΔCFI	RMSEA	SRMR	NNFI
1. Configural: Factor structure constrained to be equal	28.11	18	1.56	-	-	.982	-	.03 (CI: .01 -.07)	.06	.97
2. Metric: Factor loadings constrained to be equal	33.67	23	1.46	5.56†	5	.986	.004	.04 (CI: .01 - .06)	.05	.97
3. Scalar: Intercepts constrained to be equal	37.91	28	1.35	9.80†	10	.981	.005	.04 (CI: .01 - .06)	.07	.96

Note. N = 250 (n = 130 and 120 for males and females). CI represents 90% confidence interval. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR= Standardized Root Mean Square Residual; NNFI = Non-normed Fit Index. † *p* > .05 for all instances of χ²

Latent means invariance

In latent mean analysis, the mean of a latent variable cannot be estimated directly. However, it is possible to estimate the difference between the means of a latent variable across groups by fixing one of these means to zero in one of the groups (i.e., the reference group). Accordingly, the estimated value of the latent mean in the other group indicates the latent mean difference between the two groups. The significance test for the latent mean estimate is the test for significance of the latent mean difference between the two groups. The effect sizes of latent mean difference between the two groups were calculated from: Effect size = (M2 -M1)/SD, SD = $\sqrt{(N_1V_1 + N_2V_2)/(N_1 + N_2 - 2)}$,

where M is latent variable estimated mean, SD is the standard deviation, N is the sample size, V is a latent variable estimated variance, and the subscripts denote groups (Fukunaka, 2007). Values of 0.2, 0.5, and 0.8 were considered small, medium, and large effects, respectively (Cohen, 1988). The analysis showed that males had higher levels of perceived physical ability than females (t = 6.4, M females = .84, effect size = .56)

Correlation analysis

Table 4 shows that children's perceived physical ability correlated positively with teachers' nominations (r = .59, p < .01) and in-class physical activity (r = .46, p < .01), but not with social desirability (r = .10, ns).

Table 4
Correlations among perceived physical ability, teachers' nominations, in-class physical activity, and social desirability

Variables	1	2	3	4
1. Perceived physical ability	-			
2. Teachers' nominations of students' physical ability	.59**	-		
3. In-class physical activity	.46**	.40**	-	
4. Social desirability	.10	.09	.11	-

Note. N = 250. ** p < .01

DISCUSSION

One important finding of the present study is that the PPASC single-factor structure fit the data from the Omani children well. Although no other alternative models have been tested, the present finding demonstrates that the Arabic version of the PPASC preserves the conceptual content of Colella et al.'s original perceived physical ability model as articulated by the PPASC. This finding 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). This finding sustains the usage of participants' self-rating total score on the PPASC items when investigating perceived physical ability.

Our data further demonstrated that the PPASC was entirely equivalent across gender because all items did not display significant DIF between males and female groups. Specifically, our data supports the three levels of invariance tested: configural, metric, and scalar. This finding indicates that the PPASC items are not influenced by external irrelevant variables such as gender. 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.

Our data also provide support for the convergent and the discriminant validity of the PPASC. Perceived physical ability correlated strongly and positively with teachers' nominations. One possible explanation for this finding is that students' confidence to participate in physical activity, overcome physical activity related barriers, and organize time and responsibilities around physical activity are associated with striking and therefore salient and observable physical activity behaviors. Furthermore, perceived physical ability correlated strongly with in-class physical activity as a criterion behavioral measure. 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' achievement outcomes (e.g., effort, persistence, performance) are closely linked to their beliefs about their own competence, with self-efficacy being conceptualized as determining factor for achievement behaviors. However, our data showed that perceived physical ability did not correlate significantly with social desirability. This finding supports the discriminant validity of the PPASC and highlights the notion that students' responses to the PPASC are not biased by the need to respond in a culturally appropriate and acceptable manner.

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 contradicts the findings of Draun and Stevens (2009) and also Rabi and Swanson (2010) who reported gender differences in perceived physical ability favoring females. Yet other studies found nonsignificant differences in perceived physical ability scores (Carmen & Shineder, 2011). This finding can be interpreted within the cultural values of Oman as a masculine conservative society. It is possible that perceived physical ability is intended to maintain dominance and express adherence to masculine gender norms. As such, physical activity is represented as a training ground for manhood. By expressing high levels of perceived physical ability, male students are able to demonstrate several key components of masculinity including vigorousness, hardship, competition, aggression, toughness, dominance, and physicality.

Our research conclusions are constrained by three important limitations. First, although we used teacher nominations, in-class physical activity, and a measure of social desirability to collect criterion and discriminant information on the PPASC, it would be advantageous to include other physiological criterion measures of perceived physical ability. This would increase confidence that the PPASC actually measures perceived physical ability. Second, we used a sample of primary school student only in Oman, and therefore our results might not generalize to older students. Third, it should be noted that participant teachers were not trained observers and some of the students surely were well known to all teachers, but others were not so well known, which probably led to some unreliability in teacher nominations.

In summary, the development of the Arabic version of the PPASC is one of the strengths of this study. The findings demonstrated that the PPASC had acceptable psychometric properties as a means of measuring children's perceived physical ability. This measure 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 more confident in their interpretation of the PPASC scores when used with gender diverse samples.

REFERENCES

Al-Moseny, A., & Al-Rabie, R. (2008). Psychometric properties of the Children's Social Desirability Scale among Omani and Kuwaiti children. *Educational and Psychological Research, 3*, 37-54.

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

Arbuckle, J. L. (2006). AMOS (Version 7.0) [Computer Program]. Chicago: SPSS.

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.* New York: Freeman.

Benn, T. C., Dagkas, S., & Jawad, H. (2011). Embodied faith: Islam, religious freedom and educational practices in physical education. *Sport, Education and Society, 16*, 17-34.

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.

Byrne, B. M., Shavelson, R. J. & Muthen, B. (1989). Testing for the equivalence factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2 ed.). New York: Routledge.

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.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

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.

Crandall, V. C., Crandall, V. J., & Katkovsky, W. (1991). Children's Social Desirability Scale (CSD). In J. Robinson, P. Shaver & L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 43-46). San Diego, California: Academic Press.

Dagkas, S., & Benn, T. C. (2006). Young Muslim women's experiences of Islam and physical education in Greece and Britain: A comparative study. *Sport, Education and Society 11,* 21-38.

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.

Fukunaka, K. (2007). Latent structured mean analysis for multi-group. In H. Toyoda (Ed.), *Covariance structure analysis: AMOS* (pp. 105-134). Tokyo, Japan: TokyoTosho.

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.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development, 30,* 91-105.

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.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research, 11,* 325-344.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

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.

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

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525-543.

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.

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.

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

Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent

variable means. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 119-169). Greenwich, CT: Information Age.

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.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurements invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

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.

Wu, A. D., Li, Z., Zumbo, B. D (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment Research and Evaluation*, 12, 1-26.

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.

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.