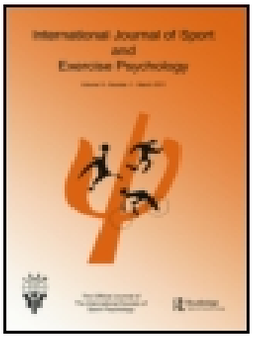


International Journal of Sport and Exercise Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rijs20

Validation of modified youth experience survey for sport (MYES-S) using multidimensional Rasch model

Kyungun R. Kim, Matthew T. Bowers, Woo-young Lee & Robert Slana

To cite this article: Kyungun R. Kim, Matthew T. Bowers, Woo-young Lee & Robert Slana (2021): Validation of modified youth experience survey for sport (MYES-S) using multidimensional Rasch model, International Journal of Sport and Exercise Psychology, DOI: 10.1080/1612197X.2021.1938633

To link to this article: https://doi.org/10.1080/1612197X.2021.1938633

Published online: 09 Jun 2021.

INTERNATIONAL JOURNAL OF SPORT AND EXERCISE PSYCHOLOGY
https://doi.org/10.1080/1612197X.2021.1938633

Validation of modified youth experience survey for sport (MYES-S) using multidimensional Rasch model

Kyungun R. Kim[a], Matthew T. Bowers[b], Woo-young Lee[a] and Robert Slana[a]

[a]School of Business Administration, University of Central Missouri, Warrensburg; [b]Department of Kinesiology and Health Education, The University of Texas at Austin, Austin

ABSTRACT

Youth sports provides the foundation for successful dynamics and organization within a sports ecosystem. Thus, it is necessary to explore the experiences of youth who either drop out of or persist in a sport. The study's main goal is to provide empirical evidence of the validity and reliability of the modified Youth Experiences Survey for Sport (MYES-S), which assesses the personal experiences of youths in sports activities using the multidimensional Rasch model. The present study is a first-of-its-kind attempt to introduce a multidimensional Rasch measurement approach into the sports psychology field while still addressing the limitations of the classical test theory approach. A total of 1,199 athletes from ninth through 12th grade participated in the survey. The findings support a six-dimensional model of the MYES-S. The person separation reliability estimates for each dimension were close to the acceptable range. Most student-athletes' logit scores were above the range of the item endorsibilities except in cognitive skill (CS); thus, more items need to be added to better capture the youth experience in sports. Future research directions include addressing differential functioning items across gender and race groups and having more appropriate response categories. The significance of this study lies in its potential to contribute to the investigation and the calibration of the MYES-S to provide practical and meaningful guidance to sports psychology researchers and practitioners vis-à-vis the understanding of the youth sports experience.

ARTICLE HISTORY
Received 21 July 2020
Accepted 21 May 2021

KEYWORDS
Youth sports; sport experience; positive youth development; YES-S; multidimensional Rasch model

Youth sports provides the foundation for successful organizational dynamics in the sports ecosystem. Providing children and adolescents with access to high-quality sports programs and strengthening a sports pipeline that helps train youth to be athletics-ready are priorities for a thriving sports ecosystem. According to the 2017 National Survey of Children's Health (The Child and Adolescent Health Measurement Initiative, 2017), 58% of U.S. youth between the ages of 6 and 17 participated in organized sports programs or took sports lessons after school during the previous year. However, the number of youths participating in sports is declining. The Sports and Fitness Industry Association

CONTACT Kyungun R. Kim kykim@ucmo.edu School of Business Administration, University of Central Missouri, Warrensburg, MO

and Aspen Institute (2019) data indicated organized-sports participation for children aged 6–12 decreased by almost 8% over the last decade. Moreover, other studies have indicated that more than 30% of children quit sports every year (Balish et al., 2014). This dropout rate accelerates after age 11 (Gardner et al., 2016).

Mitigating the abandonment of organized sports requires a clear understanding of how the benefits and advantages associated with engaging in sports can be accentuated for youth and how the negative experiences in sports can be eliminated. Researchers have typically highlighted the outcomes of participation in youth and adolescent sports instead of understanding the nature of the experience. Therefore, it is necessary to further study which experiences cause youth to either drop out of or continue with sports. This information will provide better guidance on which areas sports practitioners and policymakers should focus their efforts on (Hansen et al., 2003; Holt, 2016).

The asset-building paradigm of the positive youth development (PYD) for fostering positive outcomes has been used within sports participation literature to generate questions as to where and how we can optimally influence the personal and social development of youths and how sports can provide a worthwhile setting to promote a wide range of developmental assets (Fraser-Thomas et al., 2005; Gould & Carson, 2008; Lerner, 2002). The PYD framework has gained significant attention from sports researchers and practitioners for use in monitoring the youth sports experience through studies measuring the effectiveness of sports programs (Bean et al., 2015, 2016; Denault & Poulin, 2016), leadership studies (Newland et al., 2019; Vella et al., 2013), and youth development studies (Bruner et al., 2014, 2017; Denault & Poulin, 2016). Beyond the potential positive benefits of sports, the PYD framework indicated that a sports experience could be negatively affected by the physical and psychosocial factors of sports participation in a competitive environment and the context of elite focus. Fraser-Thomas and Côté (2009) presented a variety of antisocial and stressful experiences, choices, and influences that led youths to drop out of sports, such as decreased self-perception, pressure from parents and coaches, stress from an overemphasis on winning, and lack of playing time (Difiori et al., 2014; Fraser-Thomas et al., 2008b). Moreover, playing sports, especially contact sports, can increase a youth's aggressive orientation and actions, negatively affecting behavior (Kreager, 2007; Pappas et al., 2004).

The Youth Experiences Scale for Sport (YES-S), based on the PYD framework, is a popular instrument for acquiring information on the youth sports experience. The YES-S measures youths' positive or negative developmental outcomes across five dimensions: personal and social skills, cognitive skills, goal setting, initiative, and negative experiences (MacDonald et al., 2012; Sullivan et al., 2015). It offers a compelling combination of accessibility and efficiency and a set of tangible, experience-focused items; thus, numerous studies have adopted this instrument across disciplines.

The YES-S originated from the Youth Experience Survey (YES 1.0) developed by Hansen and Larson (2002). This instrument contained seven domains of personal youth experiences in extracurricular activities or community-based programs. Six domains represented positive experiences of identity, initiative, teamwork and social skills, interpersonal relationships, basic skills (e.g., emotional regulation), and adult networks. The seventh domain was related to negative experiences of stress, negative

influences, social exclusion, negative group dynamics, and inappropriate adult behaviors. Hansen and Larson (2002) further developed the Youth Experience Survey 2.0 (YES 2.0). YES 2.0 included a total of 70 items. Twenty-two items were removed from the original survey, and the factors were relabeled. MacDonald et al. (2012) tested the factor structure of YES 2.0 with young athletes to determine its validity for adapting this instrument to a sports-specific environment. The results of the confirmatory factor analysis (CFA) on the original measure did not support Hansen and Larson's seven-factor model. These results allowed MacDonald et al. (2012) to reexamine the factor structure of youth sports experiences using EFA. This process led to the development of YES-S built on the five-factor model of four positive domains and one negative domain: personal and social skills (PSS), cognitive skills (CS), goal setting (GS), initiative (INI), and negative experiences (NE). However, other scholars (e.g., Sullivan et al., 2015) criticized the authors because they supplied minimal information on the procedures used, and they omitted or neglected many of the conventional psychometric criteria. Additionally, some previous studies employing this measure reported several issues with the factor structure and internal consistency of YES 2.0 (Gould & Carson, 2011; Strachan et al., 2009). Sullivan et al. (2015) evaluated the YES-S using CFA to confirm the factor structure of the scale and determine how well the model explained the data for 350 youth athletes. All indices and criteria, such as χ^2, comparative fit index, standardized root-mean-square residual, root-mean-square error of approximation, and factor loadings were intended to assess the goodness of the fit of the questionnaire and provided evidence of the YES-S construct validity. More important, the Sullivan et al. (2015) five-factor model suggested that a short-form, 22-item version of the YES-S was also valid and could be more appealing to practitioners than the original scale.

Despite the sustained efforts and calls to use the YES-S in the aforementioned studies, the validated scales did not show desirable results because of unacceptable internal consistency for some items and inconsistent factor models for each study (Hansen & Larson, 2002). Moreover, the scales were analyzed based on classical test theory (e.g., CFA, PCA), which raises significant psychometric issues (e.g., nonadditive features of ordinal data, item difficulty, category functioning, and treatment of missing data; Bond & Fox, 2007).

Our study extends the Sullivan et al. (2015) work by adding a physical and psychological health factor following the hypothesis that the physical and psychological health construct can be essential in analyzing the youth sports experience. Numerous studies have proven that physical health conditions for youth are significantly affected by their sports experiences (e.g., Cairney et al., 2018; Coakley, 2011; Marques et al., 2016). The decision to add the health factor was vetted and approved by an advisory of experts across sports industries and academia in multiple meetings.

This study contributes to overcoming the limitations of the traditional psychometric approach and provides a psychometrically sound measure for youth sports research using a multidimensional Rasch measurement approach (Briggs & Wilson, 2003). We applied the multidimensional Rasch model to identify multiple dimensions in the MYES-S scale and provided the process for the evidence-based judgment that supports the appropriateness of the interpretations, uses, and decisions based on the survey results.

Methods

Data and participants

This study was a part of a national project led by the Aspen Institute Sports and Society Program, which used the Healthy Sport Index (HSI). The HSI is a multisided index aimed at better understanding how youth participation in U.S. high school sports affects their overall health (Aspen Institute, 2019). The total number of participants was 1,199. Of the cases, 159 (12.32%) showed incomplete responses or very similar response patterns across all questions (i.e., simply filling in the same response for every question). Therefore, we employed a listwise deletion method for these cases, reducing the final sample size used in the analyses to 1,131. The gender distribution included 583 females (51.5%), 541 males (49.5%), and seven others (0.6%). The participants displayed the following diversity in race and ethnicity: 864 White (76.4%), 142 African American (12.6%), 49 Hispanic (4.4%), 23 Asian (2.0%), eight Native American (0.7%), and 42 others (3.7%).

Scale development and content validity

In this study, we modified the Sullivan et al. (2015) YES-S for additional psychometric analyses, adding an additional factor—physical and psychological health (PPH)—producing a six-dimensional model. The three items for the PPH factor were developed based on the sports experience and health literature (e.g., Cairney et al., 2018; Swann et al., 2018) and the ideas of sports management experts.

To ascertain the quality of the items, it was necessary to establish content validity. Content validity refers to content relevance and content representations (McPhail, 2007). Although there are numerous methods leading to agreement among experts regarding content validity, we used CVI, which quantifies the degree of agreement among experts (Lynn, 1986). This study adopted the existing scale but added a new factor that allowed us to evaluate content validity for the MYES-S. Table 1 represents the six factors and their descriptions. Our version consisted of 26 items in a four-point

Table 1. Youth Sport Experience Factors

Dimension	Description	Supporting Citations
Personal and Social Skills	The impact of their sport participation on sharing responsibility, compromise, patience, and self-control.	Hellison (2003); Prichard & Deutsch (2015)
Cognitive Skills	The improvement of a variety of general learning skills, including searching for information, computer skills, artistic/creative skills, and discipline-based achievement related to reading and math.	Wretman (2017); Van Boekel et al. (2016)
Goal Setting	Experiences related to achieving their goals, consideration of possible barriers, and overcoming challenges.	Burton & Weiss (2008); Widmeyer & Ducharme (1997)
Initiative	An opportunity to push themselves, focus attention, direct energy toward the task, and improve their skills.	Hansen and Larson (2005); Larson (2000)
Physical & Psychological Health	Interests in being physically active, making more friends and feeling happiness.	Eime et al. (2013)
Negative Experiences	Negative aspects of sports experiences such as injury, intimidation and disparaging comments from adult leaders, controlling and manipulative adult leaders, cheating behaviors, various types of discrimination, and the possibility of exposure to alcohol or drugs.	Cairney et al. (2018); Dunn & Dunn (1999); Law, Côté, & Ericsson (2007)

Likert-type scale, where 4 = *Yes, definitely*; 3 = *Quite a bit*; 2 = *A little*; and 1 = *Not at all*. Table 1 illustrates the list of factors along with their descriptions.

Survey administration

The North Carolina High School Athletic Association and Michigan High School Athletic Association helped distribute the survey. Athletes from ninth through 12th grade across North Carolina ($N = 544$), Michigan ($N = 520$), and additional samples scattered nationally ($N = 135$) participated in this study. Participation in the web-based survey was voluntary after signing an informed consent form.

Multidimensional Rasch analysis for construct validity

The specific steps made to fulfill the study's objective: (a) evaluation of dimensionality by determining how many dimensions can be represented in the survey, (b) assessment of item fit and person fit, (c) examination of the item and person separation and reliability, (d) comparison of the "endorsibility" of the items and individual level of sports experience using an item–person map, (e) investigation of the appropriateness of a specific four-point scale within the MYES-S, and (f) identification of whether items differ across genders and races.

1) Dimensionality:

This study proposed three hypothesized models to find the most plausible model: (a) a unidimensional model, (b) a two-dimensional model (i.e., a model with two dimensions of positive and negative components), and (c) a six-dimensional model. We examined the appropriateness of the proposed models of the MYES-S as to whether the items of the instrument involve one trait (youth sports experience) or have different multidimensions proposed by previous studies (Hansen & Larson, 2002; MacDonald et al., 2012; Sullivan et al., 2015). This allowed us to compare the dimensional structures and determine which model provides the best representation of the MYES-S structure. To compare the fit of models, we conducted deviance difference testing based on the chi-square for the nested models. A nonsignificant and smaller change in deviance would indicate that the simpler model was preferable. In addition to the likelihood ratio test outcomes, we used the information-based fit index Akaike information criterion (AIC) to identify a better-fitting model. A smaller AIC index would indicate a better model fit with the data.

1) Item Fit

The item analysis involves an examination of fit statistics for each item on a scale. Two fit statistics, referred to as the weighted mean square fit and the unweighted mean square fit, are mean-squared residual statistics with an expected value of 1.0 (in a range from 0 to infinity). The weighted fit allows us to know how much the actual residuals (observed score: expected score for a specific person and item) vary, compared with how they are expected to vary randomly if the data fit the theoretical model (Wilson, 2005). The weighted mean square statistic captures the extent to which response patterns are

consistent with the rank-ordered items. When residuals vary as much as expected, the value of the weighted fit should be close to 1.0. An acceptable range of mean square values of .5 (little variation) to 1.5 (large variation) is used to determine the model fit (Linacre, 2002).

1) Person Separation Reliability

The reliability of person separation is conceptually equivalent to Cronbach's alpha—obtaining a calculation of the adjusted person standard variation divided by the mean square error. This index provides an estimate of the proportion of variance attributed to the variation in the late trait (Wright & Masters, 1982). Values of .7 or higher indicate acceptable internal consistency (Nunally, 1978).

1) Distribution of Item Endorsibility and Individual Sports Experience

A Wright map is a graphical representation showing "item endorsibility" and the respondents' estimates on the same logit scale (see Figure 1). Thus, the map shows the distribution of item endorsibility and the youth experience levels in organized sports, which allows for comparisons between the item difficulties and person measures.

1) Rating Sale Functioning:

This study identifies the optimal rating scale categories by estimating the distance between thresholds (i.e., boundaries between rating categories). When the responses to items correspond to the construct (e.g., levels of the sports experience) being measured, the thresholds should be ordered in increasing value from low to high along with the rating scale, reflecting the levels of severity across each response category. If the thresholds do not increase in a linear fashion, or the distances between two response category thresholds are ordered in an irregular pattern, the categorical function may indicate unnecessary category usage and can be improved by collapsing response categories (Wright & Linacre, 1992).

1) Differential Item Functioning:

The Rasch model requires that subgroups with equal levels of an underlying construct respond in a similar manner to the items (Tennant & Pallant, 2007). Differential item functioning (DIF) analyses provide information on whether items in the MYES-S functioned differently across genders and races with the same level of latent trait. The difference between the means for groups and the difference in the item difficulty estimates for groups are evaluated.

We used ConQuest 4.13 (Wu et al., 2015) to conduct multidimensional Rasch analyses. Additionally, before conducting the analyses, we cleaned up the data using the listwise deletion method, reverse-scoring seven items in NE originally designed in a negative direction using SPSS 26 (IBM Corp., released, 2019) to give all items a score in a more positive direction, with higher scores indicating a less negative status or a more positive experience.

Results

Content validity

The seven experts reviewed the three added items for the PPH factor. According to the CVI criteria suggested by Lynn (1986), all experts rated the three items as either 3 or 4 (i.e., 1 = not relevant; 2 = unable to assess relevance without item revision or item is in need of so much revision that it would no longer be relevant; 3 = relevant, but needs minor alteration; 4 = very relevant), which met the minimum criteria (CVI = 0.8).

Dimensionality

On the basis of the original framework for the YES-S, we expected a six-dimensional model to represent the best fit when compared with other models. Table 2 shows the final deviance, the number of parameters, and the AIC fit indices for each model. When comparing the one-dimensional and two-dimensional models, the difference was statistically significant (χ^2 difference = 4184.679, $df = 1$, $p < .001$). The difference between the two-dimensional model and six-dimensional model was also significant (χ^2 difference = 836.667, $df = 4$, $p < .001$). This suggested that the six-dimensional model fit the data significantly better than the other two models. The smaller AIC of the six-dimensional model also indicated that it was the most desirable model. The following results demonstrate the psychometric properties of the six-dimensional model.

Item Fit

In the Rasch measurement, the fit analysis usually identifies items that contribute poorly to the measurement of a trait (Bond & Fox, 2007). Therefore, it is expected that all items in this study must fit the rating scale model and measure some part of a perceived sports experience classified in the six factors. Otherwise, each poorly fitting item will have to be revised to perform accordingly or be discarded. As shown in Table 3, the analysis revealed weighted statistics ranging from .76 to 2.31, and all items except for two (NE1: "*I have had an injury in the past year in my primary sport that caused me to miss at least one day of school,*" and NE6: "*Youth in this activity got me into drinking alcohol or using drugs*") were within the acceptable range of .5–1.5 (Linacre, 2002). Only item NE1 and the unweighted statistics of all items ranged from .70–2.14 showed an overfit. A large positive value suggested that the responses to this item were significantly more variable than the model would have predicted. This indicates that the item was noisy, meaning that other factors could have been driving responses. After deleting NE1 and NE6, we reanalyzed the data to examine the model fit of the remaining 24 items, which showed a good fit (i.e., items were in the acceptable range of criteria: .67–1.46 logits).

Table 2 Fit Statistics for Proposed Three Models

Model	Final Deviance	Number of Estimated Parameters	AIC
One-factor	48239.332	25	48297.332
Two-factor	44055.625	24	44117.625
Six-factor	43218.985	20	43316.985

Table 3. Calibration Summary of 26 MYES-S Items

Items	Estimate	SE	Unweighted MNSQ	Weighted MNSQ
Personal and Social Skills (PSS):				
PSS1: I became better at sharing responsibility	.217	.044	.93	.85
PSS2: I learned that working together requires some compromising	-.18	.048	.70	.80
PSS3: I learned to be patient with other group members	.168	.045	.96	.92
PSS4: I learned how my emotions and attitude affect others in the group	-.206	.048	.93	1.01
Cognitive Skills (CS):				
CS1: I have improved: Skills for finding information	-.797	.037	.87	.89
CS2: I have improved: Academic skills (reading, writing, math, etc.)	-.137	.033	.83	.86
CS3: I have improved: Computer/internet skills	.559	.034	.94	.97
CS4: I have improved: Artistic/creative skills	.375	.033	.98	.99
Goal Setting (GS)				
GS1: I learned to find ways to achieve my goals	-.072	.049	.74	.76
GS2: I set goals for myself in this activity	-.484	.054	.71	.89
GS3: I learned to consider possible obstacles when making plans	.181	.047	.80	.80
GS4: Observed how others solved problems and learned from them	.374	.045	.82	.87
Initiative (INI):				
INI1: I learned to push myself	-.398	.068	.72	1.02
INI2: I learned to focus my attention	.292	.058	.81	.85
INI3: I put all my energy into this activity	.500	.056	1.35	.94
INI4: I have improved athletic/physical skills	-.394	.068	.73	1.18
Physical and Psychological Health** (PPH):				
PPH1: I felt like a happier person by playing this sport*	.017	.048	1.11	1.18
PPH2: I made more friends by playing this sport*	-.165	.049	1.15	1.25
PPH3: I was more interested in being physically active*	.148	.047	1.08	1.17
Negative Experiences (NE):				
NE1: I have had an injury in the past year in my primary sport that caused me to miss at least 1 d of school* [R]	1.427	.043	2.14	2.31
NE2: Adult leaders in this activity are controlling and manipulative [R]	.759	.044	1.19	1.04
NE3: Adult leaders intimidate me [R]	.258	.048	1.34	1.11
NE4: Adult leaders make personal comments that I find upsetting [R]	.455	.046	1.07	1.01
NE5: Adult leaders encouraged me to do something I believed morally wrong [R]	-.825	.062	.99	1.40
NE6: Youth in this activity got me into drinking alcohol or using drugs [R]	−1.153	.068	1.18	1.60
NE7: I was treated differently because of my gender, race, ethnicity, disability, or sexual orientation [R]	-.923	.064	1.13	1.47

Note. [R] Values reversed from standard YES-S, * Added dimension/item

Additionally, this suggests that item NE1 is statistically uninformative in regard to placing the respondents along a continuum. The misfit item might represent a different concept from the remaining items in the factor. Items in the GS and INI dimensions have relatively smaller fit values than the other items. The smaller value indicates responses were less varied than the model would have predicted.

Person separation reliability

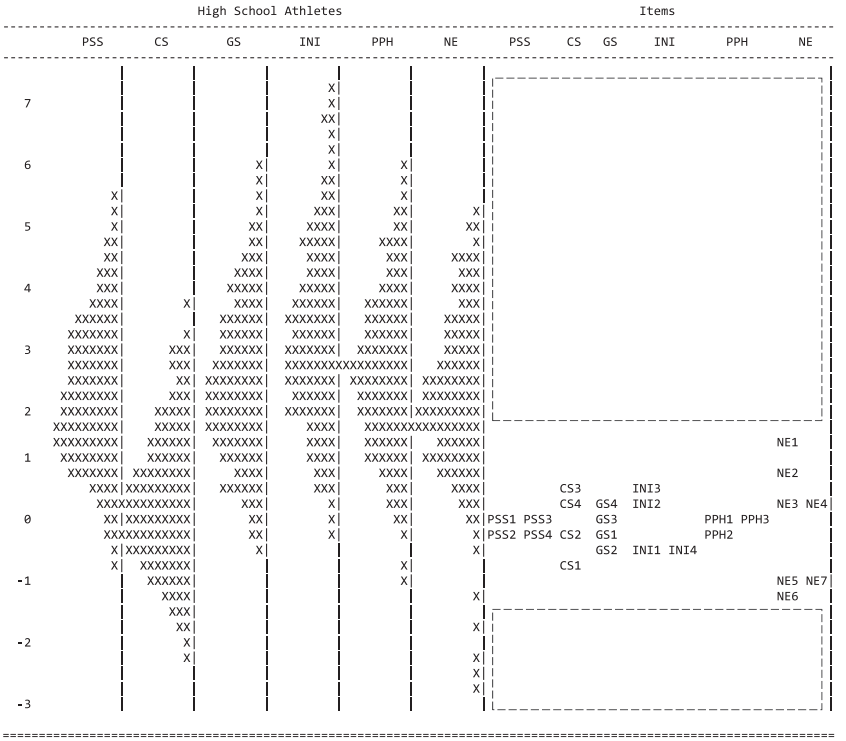
Higher values of the person separation reliability are desired because they provide more consistent results when repeated. A commonly acceptable value of the person separation reliability is considered acceptable for $\alpha \geq .7$ (Nunally, 1978). The person separation reliability estimates for each of the dimensions were close to the acceptable range:

.753 (PSS), .793 (CS), .755 (GS), .693 (physical and psychological health), .663 (INI), and .785 (NE). Overall, the six dimensions had acceptable reliabilities.

Distribution of item endorsibility and individual sports experience

Figure 1 shows the Wright map of person and item parameters of the MYES-S. The left side of each vertical line on the map displays the level of youth experience for the six dimensions. The right side of the map illustrates the distribution of item endorsibility from those that are easy to endorse to those that are difficult to endorse. The Wright map helps to compare the relative positions of item endorsibility and the youth sports experience using the same metric.

The item parameter (i.e., endorsibility—item side) indicates the likelihood of agreement on the survey item, whereas the person parameter (i.e., the youth experience level—the high school athlete side) refers to the student-athletes' experience outcome in regard to their sports participation. The items with higher logit scores (on the top of the map in the item section) indicate they are less likely to be endorsed by the student-athletes, whereas the items with lower logit scores (bottom of the map) are more easily endorsed by the student-athletes. However, because the NE items were reverse coded, they should be interpreted in the opposite way. Hence, NE6 (*"Youths in this activity got me into drinking alcohol or using drugs"*) is the least-endorsed item, which implies that youth are less likely to be exposed to alcohol or drugs while



Note. Each 'x' represents 9.9 cases. NE items were reverse-coded.

Figure 1. Wright map of person and item distributions for the six-factor model of the MYES-S.

participating in sports. A total of 603 participants (almost half the total) did not endorse this item, choosing "1" (not at all). In contrast, NE1 (*"I have had an injury in the past year in my primary sport that caused me to miss at least one day of school"*) is the most-endorsed item, which shows that injury experiences are common. The NE factors showed a wider range of item endorsibility compared to items of other factors.

The person parameters with higher logit scores meant greater developmental outcomes from sports participation. Using the Wright map, researchers can quickly understand how item parameters relate to person parameters and how the item parameters can explain the degree to which a survey "fits" the local population. The shape of the distribution of youth for each factor approached a normal distribution. As shown in Figure 1, the item parameter ranged from −1.153–1.427 logits. The student-athletes' experience level based on the logit scores showed that the INI through sports experience gained the most among the six different experiences. Most student-athletes' logit scores were located above the range of the item's endorsibilities except for CS; thus, the items' distribution did not fully cover individuals with higher experience levels. More items (i.e., capturing higher levels of outcomes) need to be added to better capture the level of youth sports experience. Two-dashed empty boxes represent locations in which more items should be added (see the boxes in Figure 1).

Rating scale functioning

We used the original four-point rating scale (4 = *Yes, definitely*; 3 = *Quite a bit*; 2 = *A little*; and 1 = *Not at all*) for all items in the MYES-S and examined the response category structure to evaluate how well the four-point options work (Linacre, 2002). Table 4 shows the step estimates in the MYES-S. The threshold estimates increased monotonically; however, of the four rating scale categories, only Category 1 had unacceptable infit or outfit mean square (MNSQ) scores outside the −2 to +2 range for both unweighted and weighted fit (Bond & Fox, 2007). Category 3 had the maximum acceptable MNSQ value for an unweighted fit (1.99), but it had an unacceptable MNSQ value for a weighted fit (2.13). These two categories were problematic. Additionally, Linacre (2002) suggested that step thresholds must advance by at least 1.4 logits for items with a three-category scale, and for a four-category scale, a shorter distance between consecutive steps is acceptable. These relatively narrow gaps were pushing items with a narrow range of endorsibility by respondents; thus, if there were many categories with up to five or six scales, the range of the item difficulty could be expanded, and the items might better match person locations.

Differential item functioning

DIF analyses were performed to examine whether items in the MYES-S functioned differently across genders and races with the same level of latent trait. The category *Other* in

Table 4. Summary of Rating Scale Structure

Category	Estimate	Error	Unweighted Fit MNSQ	Weighted Fit MNSQ
1			3.00	3.59
2	-.602	.030	1.37	1.72
3	-.099	.027	1.99	2.13
4	.701		1.72	1.71

the gender variable and several subcategories of *nonwhites* (Asian, Native American, and others) in the race variable had sample sizes too small to be compared; thus, the DIF analysis included two groups for gender (male and female) and two groups for race (White and nonwhite). For gender, the male student-athletes ($M = -.072$) were less likely to endorse the overall items in the MYES-S scale than the female student-athletes ($M = .072$), meaning the level of sports experience was lower for female students. The male students' actual experience level was almost six times larger than the standard error estimate, which indicates that the difference between males and females is clearly significant. Additionally, the significant chi-square ($\chi^2 = 125.13$, $df = 25$, $p < .001$) showed the existence of DIF in the specific items. The interaction results between the item and gender show that eight items (see CS1–CS3, PPH1, NE1–NE4) were more likely to be endorsed by males than by females; five items (see PSS2, PSS3, INI2, NE6, NE7) were more likely to be endorsed by females than by males; and the remaining 13 items had the same difficulty.

For scores based on self-identified race, White student-athletes ($M = .027$) scored higher than nonwhite student-athletes ($M = -.027$), indicating that White student-athletes had a higher score in the MYES-S than nonwhite student-athletes. The parameter equality estimates were significantly different in the DIF analysis ($\chi^2 = 173.76$, $df = 25$, $p < .001$), indicating several items varied by race. Eight items scored high in the White group (see PPH2, NE1–7), and nine items scored high in the nonwhite group (see PSS1–4, CS1–4, INI3).

Discussion

The main goal of this study was to provide empirical evidence of the validity and reliability of MYES-S. As such, the study examined the psychometric properties of the MYES-S in multiple ways using the multidimensional Rasch model. In terms of the findings for dimensionality, the MYES-S supported that all subdimensions exist independently instead of on a unidimensional scale. Additionally, the successful fit of the data to the Rasch model after eliminating the misfit items (i.e., NE1 and NE6) suggested that the empirical evidence for the scale could be improved through item revisions. The narrow distribution of the 26 items on the map compared with the respondent locations toward the top and bottom of the map mean the items did not match with respondents on the continuum; thus, an ideal survey would have more items covering the entire range of person estimates. The construct reliability of the scale showed inconsistency between theoretical suppositions about the construct and the hierarchy of the items, which indicates a need to refine the theory about the trait. Alternatively, it might indicate that the respondents understood the items in ways that were different from the theoretical assumptions of the item developers.

Further examination of the rating scale structure revealed some problematic functioning in the categories used. The original four-point Likert-type scale did not function well enough for the respondents to express the relative strength or intensity of their experience accurately. These aspects of the data indicate that further investigation, rewording, or revising of the categories will be necessary as well as with other steps.

The significant difference of item difficulty across gender and race groups can distort trait level estimates, which is often considered threatening to an instrument's validity

(Myers et al., 2006). The presence of DIF in the MYES-S could originate from diverse construct-irrelevant factors, such as other cognitive, social, and cultural factors in sports settings. For example, differential experience levels from sports participation between male and female student-athletes may be because of the presence of construct-irrelevant characteristics, such as physical aggression (Benenson, 2013) and social cohesion (Eys et al., 2015). The current study has limited value in explaining where the biases have occurred; thus, additional effort should be made to identify potential reasons for these biases in future revisions of the MYES-S scale.

Previous YES-S studies are based on the summary or averaged information of the entire scale, where the sum of participants' response scores or mean scores are used to analyze the measure. This approach potentially sacrifices details about information at the item and individual participant level. Furthermore, previous YES-S validation studies provided only the information for the factor structure of the scale and the internal consistency. The Rasch approach complements the disadvantages and provides the psychometric results at the individual item and person level.

The Rasch analysis offers psychometric properties to guide future revisions of the MYES-S to adequately apply it to young athletes. If the MYES-S is further improved as stated above, researchers, coaches, and administrators will be able to assess young athletes' experiences in their sports to reveal and further address the challenges and issues that youths face. Additionally, the MYES-S can be used to discover the relationship between a set of retention- and dropout-related variables, allowing us to explore the youth experience in sports using quantitative methods. To provide comprehensive insight into their experiences, the MYES-S can be used with other qualitative research methodologies, such as ethnography and multiple case studies.

Program leaders could use the MYES-S to evaluate their programs' effectiveness to improve their quality. However, evidence-based judgment about the evaluation must rely on the instruments; thus, having validated instruments is essential to supporting the appropriateness of the interpretations that are made from the survey results.

Conclusion

Young athletes can have various positive and negative experiences in sports. The experiences could be the consequences of a number of cultural, economic, and systemic issues that affect youths. Thus, it is vital to be informed about the youth sports experience. The steady decline of youth participation in sports is a serious problem. The MYES-S instrument may be particularly beneficial for capturing and monitoring high school athletes' experiences over time, which enables one to provide the needed feedback for the continuous promotion of PYD. The use of a vetted tool could provide practical and meaningful guidance to sports psychology researchers and practitioners vis-à-vis the understanding of the youth sports experience. Improved understanding of the developmental benefits of participating in sports as well as the negative experiences some athletes face could lead to theoretically grounded intentional sports programming or participant support, potentially increasing participants' movement through the "funnel."

Acknowledgments

The authors wish to express their gratitude to the Hospital for Special Surgery and the Aspen Institute's Project Play for funding and organizing the Healthy Sport Index, the overarching project from which this manuscript and the data presented herein were derived.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Aspen Institute. (2019). *State of play: 2019 trends and developments*. https://assets.aspeninstitute.org/content/uploads/2018/10/StateofPlay2018_v4WEB_2-FINAL.pdf

Balish, S. M., McLaren, C., Rainham, D., & Blanchard, C. (2014). Correlates of youth sport attrition: A review and future directions. *Psychology of Sport and Exercise*, *15*(4), 429–439. https://doi.org/ doi:10.1016/j.psychsport.2014.04.003

Bean, C., & Forneris, T. (2016). Re-examining the youth program quality survey as a tool to assess quality within youth programming. *Cogent Psychology*, *3*, 1–14. https://doi.org/10.1080/23311908.2016.1149265

Bean, C., Forneris, T., & Fortier, M. (2015). Girls just wanna have fun: Understanding perceptions of effective strategies and outcomes in a female youth-driven physical activity-based life skills programme. *Journal of Sport for Development*, *3*(4), 28–40.

Benenson, J. F. (2013). The development of human female competition: Allies and adversaries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1631), 20130079. https://doi.org/ doi:10.1098/rstb.2013.0079

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associates.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, *4*(1), 87–100.

Bruner, M. W., Balish, S. M., Forrest, C., Brown, S., Webber, K., Gray, E., McGuckin, M., Keats, M. R., Rehman, L., & Shields, C. A. (2017). Ties that bond: Youth sport as a vehicle for social identity and positive youth development. *Research Quarterly for Exercise and Sport*, *88*(2), 209–214. https://doi.org/ doi:10.1080/02701367.2017.1296100

Bruner, M. W., Eys, M. A., Wilson, K. S., & Côté, J. (2014). Group cohesion and positive youth development in team sport athletes. *Sport, Exercise, and Performance Psychology*, *3*(4), 219–227. https://doi.org/ doi:10.1037/spy0000017

Burton, D., & Weiss, C. (2008). The fundamental goal concept: The path to process and performance success. In T.S. Horn (Ed.), *Advances in sport psychology* (3rd ed., pp. 339–375). Human Kinetics.

Cairney, J., Clark, H. J., Kwan, M. Y., Bruner, M., & Tamminen, K. (2018). Measuring sport experiences in children and youth to better understand the impact of sport on health and positive youth development: Designing a brief measure for population health surveys. *BMC Public Health*, *18*(1), 446–456. https://doi.org/ doi:10.1186/s12889-018-5325-9

The Child and Adolescent Health Measurement Initiative. (2017). *National Survey of Children's Health (NSCH) data query. Data resource center for child & Adolescent Health [database online]*. Johns Hopkins Bloomberg School of Public Health. childhealthdata.org.

Coakley, J. (2011). Youth sports: What counts as "positive development?". *Journal of Sport and Social Issues*, *35*(3), 306–324. https://doi.org/ doi:10.1177/0193723511417311

Denault, A. S., & Poulin, F. (2016). What adolescents experience in organized activities: Profiles of individual and social experiences. *Journal of Applied Developmental Psychology*, *42*, 40–48. https://doi.org/ doi:10.1016/j.appdev.2015.11.004

Difiori, J., Benjamin, H., Brenner, J., Gregory, A., Jayanthi, N., Landry, G., & Luke, A. (2014). Overuse injuries and burnout in youth sports: A position statement from the American Medical Society

for Sports medicine. *British Journal of Sports Medicine*, *48*(4), 287–288. https://doi.org/ doi:10.1136/bjsports-2013-093299

Dunn, J. G., & Dunn, J. C. (1999). Goal orientations, perceptions of aggression, and sports personship in elite male youth ice hockey players. *The Sport Psychologist*, *13*(2), 183–200. https://doi.org/10.1123/tsp.13.2.183

Eime, R. M., Young, J. A., Harvey, J. T., Charity, M. J., & Payne, W. R. (2013). A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. *International Journal of Behavioral Nutrition and Physical Activity*, *10*(1), 1–21. https://doi.org/10.1186/1479-5868-10-98

Eys, M., Evans, M. B., Martin, L. J., Ohlert, J., Wolf, S. A., Van Bussel, M., & Steins, C. (2015). Cohesion and performance for female and male sport teams. *The Sport Psychologist*, *29*(2), 97–109. https://doi.org/ doi:10.1123/tsp.2014-0027

Fraser-Thomas, J., & Côté, J. (2009). Understanding adolescents' positive and negative developmental experiences in sport. *The Sport Psychologist*, *23*(1), 3–23. https://doi.org/ doi:10.1123/tsp.23.1.3

Fraser-Thomas, J., Côté, J., & Deakin, J. (2005). Youth sport programs: An avenue to foster positive youth development. *Physical Education & Sport Pedagogy*, *10*(1), 19–40. https://doi.org/ doi:10.1080/1740898042000334890

Fraser-Thomas, J., Côté, J., & Deakin, J. (2008a). Examining adolescent sport dropout and prolonged engagement from a developmental perspective. *Journal of Applied Sport Psychology*, *20*(3), 318–333. https://doi.org/ doi:10.1080/10413200802163549

Fraser-Thomas, J., Côté, J., & Deakin, J. (2008b). Understanding dropout and prolonged engagement in adolescent competitive sport. *Psychology of Sport and Exercise*, *9*(5), 645–662. https://doi.org/ doi:10.1016/j.psychsport.2007.08.003

Gardner, L. A., Vella, S. A., & Magee, C. A. (2016). Continued participation in youth sports: The role of achievement motivation. *Journal of Applied Sport Psychology*, *29*(1), 17–31. https://doi.org/ doi:10.1080/10413200.2016.1173744

Gould, D., & Carson, S. (2008). Life skills development through sport: Current status and future directions. *International Review of Sport and Exercise Psychology*, *1*(1), 58–78. https://doi.org/ doi:10.1080/17509840701834573

Gould, D., & Carson, S. (2011). Young athletes' perceptions of the relationship between coaching behaviors and developmental experiences. *International Journal of Coaching Science*, *5*(2), 3–29.

Hansen, D. M., & Larson, R. (2002). *The youth experience survey 1.0: Instrument development and testing*. University of Illinois.

Hansen, D. M., & Larson, R. (2005). *The youth experience survey 2.0: Instrument revisions and validity testing*. University of Illinois.

Hansen, D. M., Larson, R. W., & Dworkin, J. B. (2003). What adolescents learn in organized youth activities: A survey of self-reported developmental experiences. *Journal of Research on Adolescence*, *13*(1), 25–55. https://doi.org/ doi:10.1111/1532-7795.1301006

Hellison, D. (2003). *Teaching responsibility through physical activity* (2nd ed.). Human Kinetics.

Holt, N. L. (2016). *Positive youth development through sport*. Routledge.

IBM Corp. (Released 2019). *SPSS statistics for macintosh, version 26.0*.

Kreager, D. A. (2007). Unnecessary roughness? School sports, peer networks, and male adolescent violence. *American Sociological Review*, *72*(5), 705–724. https://doi.org/ doi:10.1177/000312240707200503

Larson, R. W. (2000). Toward a psychology of positive youth development. *American Psychologist*, *55*(1), 170–183. https://doi.org/10.1037/0003-066X.55.1.170

Law, M. P., Côté, J., & Ericsson, K. A. (2007). Characteristics of expert development in rhythmic gymnastics: A retrospective study. *International Journal of Sport and Exercise Psychology*, *5*(1), 82–103. https://doi.org/10.1080/1612197X.2008.9671814

Lerner, R. M. (2002). *Adolescence: Development, diversity, context, and application.* Pearson Education.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

MacDonald, D. J., Côté, J., Eys, M., & Deakin, J. (2012). Psychometric properties of the youth experience survey with young athletes. *Psychology of Sport and Exercise*, *13*(3), 332–340. https://doi.org/doi:10.1016/j.psychsport.2011.09.001

Marques, A., Ekelund, U., & Sardinha, L. B. (2016). Associations between organized sports participation and objectively measured physical activity, sedentary time and weight status in youth. *Journal of Science and Medicine in Sport*, *19*(2), 154–157. https://doi.org/doi:10.1016/j.jsams.2015.02.007

McPhail, S. M. (2007). *Alternative validation strategies: Developing new and leveraging existing validity evidence*. John Wiley & Sons.

Myers, N. D., Wolfe, E. W., Maier, K. S., Feltz, D. L., & Reckase, M. D. (2006). Extending validity evidence for multidimensional measures of coaching competency. *Research Quarterly for Exercise and Sport*, *77*(4), 451–463. https://doi.org/ doi:10.1080/02701367.2006.10599380

Newland, A., Newton, M., Moore, E. W. G., & Legg, W. E. (2019). Transformational leadership and positive youth development in basketball. *International Sport Coaching Journal*, *6*(1), 30–41. https://doi.org/ doi:10.1123/iscj.2018-0002

Nunally, J. C. (1978). *Psychometric theory (2nd ed.)*. McGraw-Hill.

Pappas, N. T., McKenry, P. C., & Catlett, B. S. (2004). Athlete aggression on the rink and off the ice: Athlete violence and aggression in hockey and interpersonal relationships. *Men and Masculinities*, *6*(3), 291–312. https://doi.org/ doi:10.1177/1097184X03257433

Prichard, A., & Deutsch, J. (2015). The effects of motivational climate on youth sport participants. *The Physical Educator*, *72*(5). https://doi.org/10.18666/TPE-2015-V72-I5-6999

Strachan, L., Côté, J., & Deakin, J. (2009). "Specializers" versus "samplers" in youth sport: Comparing experiences and outcomes. *The Sport Psychologist*, *23*(1), 77–92. https://doi.org/ doi:10.1123/tsp.23.1.77

Sullivan, P. J., LaForge-MacKenzie, K., & Marini, M. (2015). Confirmatory factor analysis of the youth experiences survey for sport (YES-S). *Open Journal of Statistics*, *05*((05|5)), 421–429. https://doi.org/

Swann, C., Telenta, J., Draper, G., Liddle, S., Fogarty, A., Hurley, D., & Vella, S. (2018). Youth sport as a context for supporting mental health: Adolescent male perspectives. *Psychology of Sport and Exercise*, *35*, 55–64. https://doi.org/ doi:10.1016/j.psychsport.2017.11.008

Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions*, *20*(4), 1082–1084.

Van Boekel, M., Bulut, O., Stanke, L., Zamora, J. R. P., Jang, Y., Kang, Y., & Nickodem, K. (2016). Effects of participation in school sports on academic and social functioning. *Journal of Applied Developmental Psychology*, *46*, 31–40. http://dx.doi.org/10.1016/j.appdev.2016.05.002

Vella, S. A., Oades, L. G., & Crowe, T. P. (2013). The relationship between coach leadership, the coach–athlete relationship, team success, and the positive developmental experiences of adolescent soccer players. *Physical Education and Sport Pedagogy*, *18*(5), 549–561. https://doi.org/ doi:10.1080/17408989.2012.726976

Widmeyer, W. N., & Ducharme, K. (1997). Team building through team goal setting. *Journal of Applied Sport Psychology*, *9*(1), 97–113. https://doi.org/10.1080/10413209708415386

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wretman, C. J. (2017). School sports participation and academic achievement in middle and high school. *Journal of the Society for Social Work and Research*, *8*(3), 399–420. https://doi.org/10.1086/693117

Wright, B. D., & Linacre, J. M. (1992). Combining and splitting categories. *Rasch Measurement Transactions*, *6*(3), 3233–3235.

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

Wu, M. L., Adams, R. J., & Wilson, M. R. (2015). ACER ConQuest: Generalised item response modeling software [Computer software] Version 4. Australian Council for Educational Research.