



Development and Initial Validation of the Sport Education Scale

Rafael Burgueño ^{a,b}, Antonio Calderón ^c, Oleg Sinelnikov ^d, and Jesús Medina-Casaubón^b

^aFaculty of Health Sciences, Universidad Isabel I, Burgos, Spain; ^bDepartment of Physical Education and Sport, University of Granada, Granada, Spain; ^cDepartment of Physical Education and Sport Sciences, University of Limerick, Limerick Ireland; ^dDepartment of Kinesiology, University of Alabama, Tuscaloosa, United States

ABSTRACT

This research developed and psychometrically tested the Sport Education Scale, a measure of students' perceptions of the structural features of a Sport Education season. In the first study ($N = 277$ students), a pool of 28 items was developed, and an exploratory factor analysis found a 7-factor solution. In the second study ($N = 656$ students), a confirmatory factor analysis verified the 28-item 7-factor correlated model, which was invariant across gender and age. Convergent validity was met by suitable average variance extracted values. Discriminant validity was underpinned by appropriate values for heterotrait-monotrait ratio of correlations and correlations among latent factors. Reliability was adequate for each factor of the Sport Education Scale. A structural equation modeling showed that Sport Education positively predicted perceived competence, sportspersonship, and intrinsic motivation among students. The Sport Education Scale is a valid and reliable instrument to assess the core elements of this pedagogical model.

KEYWORDS

physical education; model-based practice; teaching and learning model; psychometric analysis

Introduction

Current sport pedagogy literature describes a number of pedagogical models that allow for achieving the breadth and depth of learning in physical education in different contexts (Casey & Kirk, 2020). Presently referred to as models-based practice or pedagogical models, these approaches to teaching physical education are becoming a significant component of any quality physical education program (Casey et al., 2021; Sinelnikov & Hastie, 2016). Although there is an emerging, developing or established body of research on most pedagogical models, it is Sport Education, initially conceived by Daryl Siedentop (1994), that has been described as the most well researched and philosophically justified (Sinelnikov & Hastie, 2016).

Over the last three decades, empirical research on Sport Education has mainly focused on the model's effect on its three main goals, namely, the development of student's competence, literacy and enthusiasm (Hastie & Wallhead, 2016). Siedentop et al. (2020) conceptualize competence as the acquisition of the sufficient level of skill to successfully participate in games and activities, as well as adequate understanding and execution of appropriate strategies for the complexity of the task. They define literacy as the understanding, respect and valuation of the rules, rituals and traditions of sport and learning to distinguish between good and bad

practices. The goal of enthusiasm within Sport Education refers to student behavior that preserves, protects and enhances sporting culture and the desire for sport participation based on the meaningfully authentic experiences and enjoyment. The overarching results of the established body of research on the model relative to its outcomes have been predominantly positive with findings described in three major reviews of research (Hastie et al., 2011; Hastie & Wallhead, 2016; Wallhead & O'Sullivan, 2005).

With respect to Sport Education's goal of the development of competence, there is robust recent evidence that supports students' improvement in skills and techniques (Bessa et al., 2021), decision-making and game tactical performance (Mahedero et al., 2015), as well as perceived motor competence (Chu & Zhang, 2018; Evangelio et al., 2018; Manninen & Campbell, 2021). In terms of the Sport Education's goal of development of literate sportspeople, a substantial body of research suggests that students improve their knowledge of the rules and understanding of the sport (Hastie & Wallhead, 2015). After participation in Sport Education, students' respect for social conventions, respect for rules and referees, full commitment, and respect for opponents (four out of five dimensions comprising the construct of sportspersonship) significantly improved (Burgueño & Medina-Casaubón, 2020; Wahl-Alexander et al., 2017). Furthermore, participating

in Sport Education results in marked increases in students' ethical behavior measured with indicators such as cooperation (Fernandez-Rio & Casey, 2020), social justice, peer support, inclusion (Harvey et al., 2014; Hastie & Wallhead, 2015), and social and personal responsibility (Guijarro et al., 2021). On the other hand, high levels of students' enthusiasm resulting from participation in Sport Education across different grade levels and in different countries have contributed to the model's goal of developing enthusiastic sportspeople (Hastie & Sinelnikov, 2006; Wahl-Alexander et al., 2017). Furthermore, increased levels of enthusiasm provide for higher student involvement and investment in the learning process (Calderón et al., 2013), an increased intrinsic motivation (Chu & Zhang, 2018; Manninen & Campbell, 2021) and result in greater levels of fun, enjoyment, and engagement (Bessa et al., 2019; Hastie & Wallhead, 2015).

The empirical research evidence highlighting the use of Sport Education in teaching and learning contexts rests on assumptions that Sport Education, as a pedagogical model, is carried out with its six primary structural features that characterize institutional sport and that these features should be a part of any Sport Education season (Siedentop et al., 2020). These non-negotiable features of Sport Education include seasons (i.e., teaching units with a higher number of sessions than the normal), affiliation (i.e., students are arranged into teams throughout the entire unit), formal competition (i.e., a game schedule in which training and competitions are interspersed), record keeping (i.e., formal records are kept on game play), festivity (i.e., continual festivity during the unit to celebrate improvements, fair play, and genuine efforts), and a culminating event (i.e., a final game to determine the overall winner of the unit and a celebration to recognize all accomplishments). Additionally, during any Sport Education season, participants learn diverse roles such as the player, official, coach, manager, physical trainer, game analyst or journalist (Siedentop et al., 2020). It is evident that each Sport Education season is unique since it is predicated on teachers' interpretation and delivery of it, as some teachers may use the model fully, while others may only implement certain aspects of the model and omit others which would result in a "watered down" version of the model (Casey et al., 2021). By consequence of this variability, it is crucial that fidelity of the model and its features are well established when conducting research on any pedagogical model (Hastie & Casey, 2014).

While the research area of students' perceptions of Sport Education has received considerable attention within the literature (Hastie et al., 2011; Hastie & Wallhead, 2016; Wallhead & O'Sullivan, 2005), there is a noted paucity of valid and reliable instruments that

examine students' perceptions of the model. In particular, there has been only one instrument described in published literature that attempts to assess the structural features of Sport Education. Specifically, Mohr et al. (2003) created the Physical Education Season Survey as an instrument that aimed to examine students' perception of Sport Education features, although these items were never published. Expert judgment was used as the foundation of providing validity evidence based on content of survey questions (Kinchin et al., 2004); however, this particular survey has been sparingly used (Burgueño, Lirola et al., 2018; Calderón et al., 2010; Hastie & Sinelnikov, 2006). One of the reasons for seldom use of the instrument lies in its design that severely limits its usefulness for research. In particular, the survey instrument uses only one item to measure each structural feature of Sport Education (Mohr et al., 2003). An accepted psychometric methodology recommends the use of a minimum of three items to measure the same construct in order to adequately explore the construct and perform psychometric tests (Hair et al., 2018; Kline, 2016). Consequently, it is impossible for researchers who use the Physical Education Season Survey to conduct any psychometric analysis and, more importantly, to provide evidence supporting validity and reliability of the measurement.

Based on the limitations of the available survey and the lack of other valid and reliable instruments, there is a need for researchers and teachers to develop psychometrically robust instruments to further investigate student perceptions of Sport Education and its core features in the physical education setting. Indeed, the use of these instruments could complement the existing knowledge base in the field, which previously heavily relied on qualitative data sources (e.g., interviews, diaries, etc.) and therefore, broaden the scope of methodological approaches for future research that could potentially strengthen the body of knowledge related to Sport Education as a pedagogical model in physical education.

Therefore, the objective of the study was to develop and provide validity and reliability evidence for the Sport Education Scale (SES), an instrument that examines students' perceptions of the structural features of a Sport Education season. In line with previous research (Andrade et al., 2020; Rocchi et al., 2017; Sabourin et al., 2020; Tilga et al., 2017), we present two sequential studies that allow us to meet the goal of the study.

Study 1: development of the SES

The objectives of the first study were: (a) to develop a pool of items that refer to the structural features of Sport Education, and (b) to explore the factor

175 composition and structure of the items by means of an
180 exploratory factor analysis.

Method

Participants

180 Purposive sampling technique was used in which the use
of Sport Education during regular Physical Education
185 courses served as a criterion for inclusion into a sample.
The participants were 277 students (137 boys and 140
girls) aged between 12 and 18 years ($M_{age} = 15.14$, $SD =$
1.69) from the five public middle and high schools from
a city in the Spanish southeast. A total of 55 students
were enrolled in seventh grade, 60 in eighth grade, 30 in
ninth grade, 84 in tenth grade and 48 students in ele-
venth grade.

Instruments

190 The development of the SES followed classical test the-
ory. The subsequent stages proposed by Muñiz and
Fonseca-Pedero (2019) were followed to develop the
instrument. Namely, these stages included: (a) construc-
195 tion of a pool items, (b) expert appraisal, (c) pilot study,
and (d) second expert appraisal (see Figure 1).

Construction of a pool of items

During the first stage of instrument development a table
of item specifications was created to measure the seven
structural features that define Sport Education. This
200 table included the following elements: item format,
Likert-type scale with five answer choices, definition of
the construct, an example of shafted item, and a ratio of
items per factor. Since the process of questionnaire
construction should begin with the creation of
205 a number of items twice or three times higher than
those that the instrument will have in its final version
(Muñiz & Fonseca-Pedero, 2019), the rate of six items
per factor was chosen. Second, four University profes-
sors with expertise in Sport Education conducted
210 a discussion panel during which they created a 42-item
pool of questions in accordance with the pre-established
criteria listed in the table (first version of the scale). All
experts had the following qualifications: (a) PhDs in
physical education, (b) have published peer reviewed
215 research on the model-based practice, (c) have taught
numerous Sport Education seasons themselves, and (d)
have formally taught others how to teach using the
model.

Expert appraisal

220 After construction of a pool of items, a second panel of
four Sport Education experts independently assessed
and judged the content of each item. This panel

consisted of two University professors and PhDs in
physical education and two secondary physical educa-
225 tion teachers. These experts have also taught numerous
Sport Education seasons and have taught or supervised
others when teaching or using the model. During this
phase, experts estimated the grade of belonging (i.e., the
extent to which each item on a test is relevant to the
230 targeted domain, Sireci & Faulkner-Bond, 2014) of each
item to the construct for which it was created and, in
turn, indicated the importance attributed to the defini-
tion of the same construct, as well as the clarity in its
drafting. Following the Sireci and Faulkner-Bond (2014)
235 recommendations, three quantitative estimates were
performed using a 5-point Likert-type scale to respec-
tively judge appropriateness, conceptual clarity, and
relevance of each item with respect to the construct
which it purported to represent. Additionally, the
240 Aiken's V coefficient (Aiken, 1980) was calculated as
a content validity measure, showing acceptable values
when the lower bond of its CI at 95% (95%CI) is equal to
.70 or higher (Merino & Livia, 2009). Regarding domain
representation, Aiken's V values suggested that five
245 items had to be removed due to their redundancy of
the assessment of the target construct. Concerning to
relevance of each item, Aiken's V values suggested five
items had to be removed since their minor importance
relative to the assessment of construct. With respect to
250 conceptual clarity, Aiken's V values indicated that three
items were found to have confusing wording.
Consequently, these three items were redrafted in
order to clarify their meaning. After this process, the
scale was reduced to 32 items grouped into the seven
255 initially hypothesized factors (second version of the
scale).

Pilot study

The 32-item scale was applied to an independent sample
of 20 students in one Physical Education class for the
purpose of checking its interpretation and understand-
260 ability. Pilot study participants were 9th graders (12 boys
and 8 girls; $M_{age} = 14.50$, $SD = 0.89$) who have previously
participated in a Sport Education season during their
physical education classes. The researchers asked parti-
cipants to identify items they considered to be similar or
265 repetitive, as well as those questions that they considered
difficult to understand. As result of this process, four
additional items were removed, and the scale was
reduced to 28 items (third version of the scale).

Second expert appraisal

270 After the pilot study, three experts (two PhDs in physical
education and University professors and a secondary
physical education teacher) reviewed the 28-item scale.

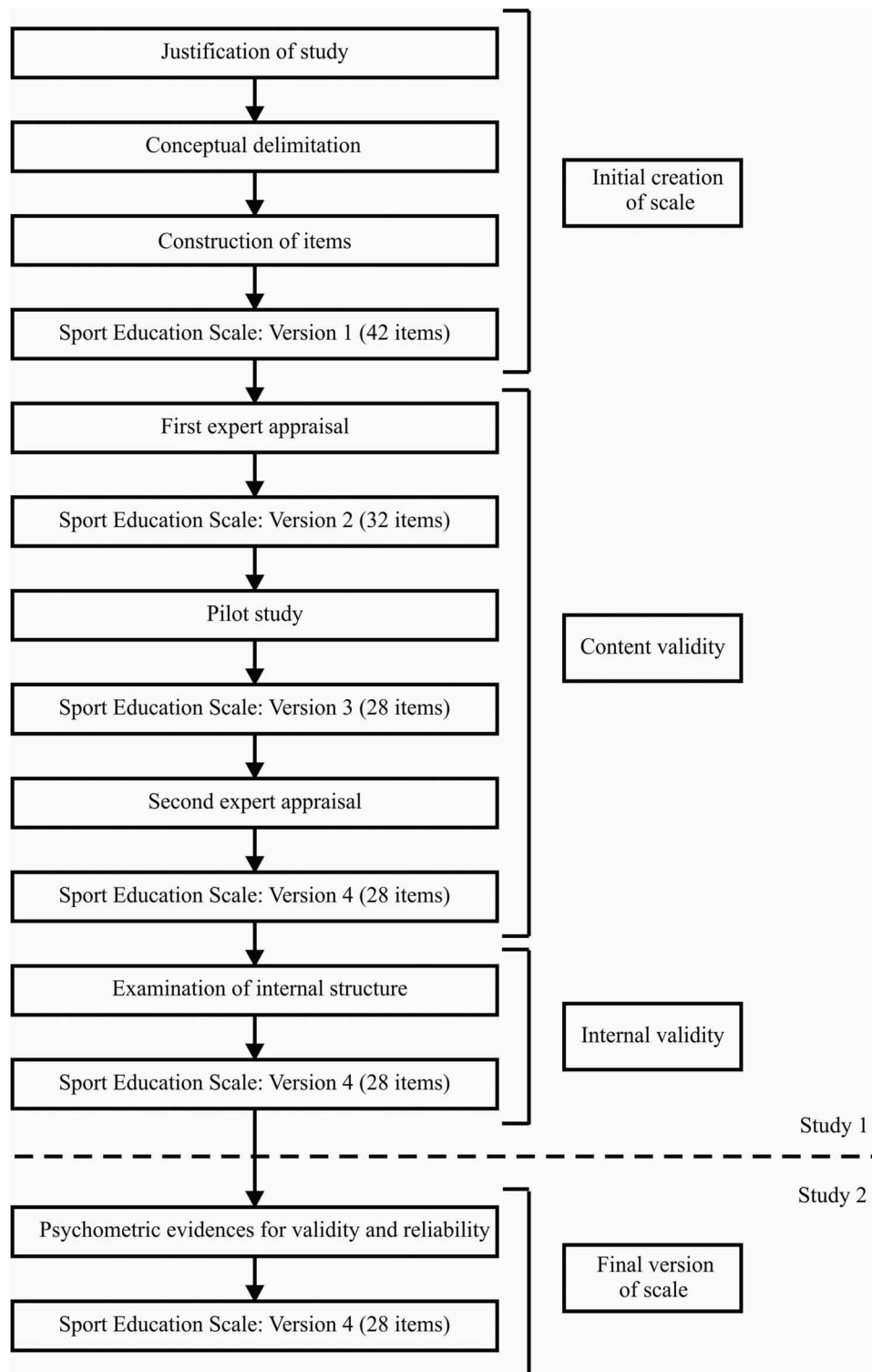


Figure 1. Sport Education Scale Development Process.

275 As a result of this new assessment, Aiken's V values above .70 for the lower bond of its 95%CI were obtained for every item in terms of domain representation, relevance and clarity. Finally, the SES included 28 items grouped into 4 items per factor (fourth version of the scale). Altogether, the development process of the SES

280 offered validity evidence based on the instrument content.

Procedure

This research was approved by the Granada University's Ethical Committee and authorized by each participating

285 school. Additionally, participants and their guardians
 provided assent and informed consent, respectively. The
 SES was administered after the conclusion of a Sport
 Education season. The researchers informed participants
 that their participation was voluntary and anonymous
 290 and that there were no correct and incorrect responses.
 Prior to the administration of the scale, participants had
 an opportunity to answer any questions. The average
 time for instrument administration was 20 min.

Intervention program

295 The authors along with each of the physical education
 teachers designed a Sport Education intervention based
 on the guidelines outlined by Siedentop et al. (2020), for
 Sport Education. This intervention was conducted in
 regular physical education classes over a period of six
 300 weeks. The season in each class was conducted across
 twelve 60-min sessions, twice a week, for a total of
 720 min of instructional time (Fernandez-Rio et al.,
 2013). Following Siedentop et al. (2020) recommenda-
 tions, the implemented Sport Education season had
 305 three main phases with students gradually assuming
 more responsibilities as the season progressed. These
 were: (a) a skill development phase, including an intro-
 duction session and teacher-directed lessons, (b) non-
 consequence games or practice competition, and (c)
 310 a formal competition phase, including regular competi-
 tion and a culminating event.

To ensure fidelity of the Sport Education interven-
 tion, the teachers received a theoretical-practical course
 of 14 hr in relation to the core features defining this
 pedagogical model. The structure and content of this
 315 course was designed based on the studies developed by
 Sinelnikov (2009) and Calderón and Martínez de Ojeda
 (2014) was delivered by the research team.
 Additionally, continuous on-site and remote support-
 320 ing process was implemented to facilitate delivery of
 Sport Education (Calderón & Martínez de Ojeda, 2014;
 Sinelnikov, 2009). This process consisted of: (a) inspec-
 tion sessions per season throughout the intervention
 program; (b) telephone conversations and e-mails with
 325 teacher to provide support and resolve problems, con-
 cerns and doubts; and (c) external observations with
 the observational record sheet created by Sinelnikov
 (2009) and adapted by Calderón et al. (2010) to the
 Spanish context. These external observations aimed at
 330 verifying the presence or absence of mismatches
 between the planned content and the implemented
 one, as well as confirming model fidelity to make sure
 that the pedagogical model was being enacted with all
 of its features

Data analysis

335 The robustness of applying an exploratory factor analy-
 sis was assessed by determining if the Kaiser-Meyer-
 Olkin (KMO) statistic was over .80 (Hair et al., 2018),
 and the Bartlett's sphericity test. Subsequently, to pro-
 340 vide validity evidence based on internal structure, an
 exploratory factor analysis was performed using the
 principal components method with the *varimax* ortho-
 gonal rotation. This technique tends to minimize the
 small coefficients and to maximize the large coefficients
 345 in a way that allows one to obtain profiles that are more
 easily associated with a concrete subset of variables and
 to facilitate interpretation of the factors obtained (Hair
 et al., 2018). To perform this analysis, a pre-determined
 number of factors was not set. Furthermore, we followed
 the criterion proposed by Hair et al. (2018) for inter-
 350 pretation of the items extracted which shows that an
 item with factor loading higher than .50 might be con-
 sidered as acceptable when its secondary factor loading
 is not above .32. The IBM Statistical Package for the
 Social Science version 22.00 (SPSS v.22) was used to
 355 perform statistical analysis.

Results

The Kaiser-Meyer-Olkin statistic ($KMO = .96$), as well as
 the Bartlett's sphericity test ($\chi^2 = 5270.60$, $df = 378$, $p <$
 360 $.001$), showed satisfactory results, indicating the robustness
 of applying an exploratory factor analysis. The exploratory
 factor analysis revealed a solution comprising of seven
 factors, which accounted for 74.77% of the total variance
 in the items. Table 1 presents the factor loading of each
 365 of the items for each factor, where they obtained a primary
 factor loading higher than .50 on the factor aiming to
 measure. Therefore, the final instrument consisted of
 seven factors and 28 items (final version of the scale).

Study 2: psychometric assessment of the SES

370 The objectives of a second study were to verify the
 instrument's 28-item and seven-factor internal structure
 established in the Study 1 via a confirmatory factor
 analysis and to examine invariance across gender and
 age, and reliability. This study also gathered criterion
 375 validity evidence by analyzing the predictive relationship
 of Sport Education on perceived competence, sports-
 personship, and intrinsic motivation.

Method

Participants

380 Purposive sampling method was used, in which the use
 of Sport Education during regular Physical Education

Table 1. Contents and Activities of the Sport Education Season.

Session	Session Content	Teacher's Role	Student's Role
Skill development phase: Teacher directed			
1	Presentation of Sport Education, team building, role assignment and choice of color of clothes and shield.	Leader	Participant
2	Familiarization with Sport Education. Assumption of roles. Training via small-sided games.	Leader	Participant
3	Familiarization with Sport Education. Assumption of roles. Training via small-sided games.	Leader	Subordinate
Pre-season: Students to work in their teams within peer-teaching. Scrimmages			
4	Training via small-sided games. Pre-season tournament	Head coach/ Referee advisor	Duty team roles
5	Training via small-sided games. Pre-season tournament	Head coach/ Referee advisor	Duty team roles
6	Training via small-sided games. Pre-season tournament	Head coach/ Referee advisor	Duty team roles
7	Training via small-sided games. Pre-season tournament	Head coach/ Referee advisor	Duty team roles
8	Training via small-sided games. Pre-season tournament	Head coach/ Referee advisor	Duty team roles
Formal competition: Round robin tournament			
9	Competitions (regular phase)	Guide/ Mediator	Duty team roles
10	Competitions (regular phase)	Guide/ Mediator	Duty team roles
11	Competitions (regular phase)	Guide/ Mediator	Duty team roles
12	Culminating event: 3 rd /4 th place match and award ceremony ceremonies	Master of Participant	

Note. Teachers were offered then different strands (i.e., basketball, handball, volleyball, floorball, badminton, futsal) and the small side games practiced varied from 2 × 2 to 3 × 3.

courses served as a criterion for inclusion into this study. A new sample of 656 students (351 boys and 305 girls) aged between 12 and 19 years ($M_{age} = 15.01$, $SD = 1.60$) from 13 public middle and high schools which were located in a city from the southeast of Spain participated in Study 2. Ninety-six seventh grade, 102 eighth grade, 122 ninth grade, 250 tenth grade, and 86 eleventh grade students took part in Study 2.

Instruments

Core features of the Sport Education. The SES, with its 28-item and seven-factor correlated structure obtained from EFA in Study 1, was used to examine the structural features of the model.

Perceived competence

The Spanish version adapted to Physical Education (Moreno et al., 2008) of the Basic Psychological Needs in Exercise Scale (Vlachopoulos & Michailidou, 2006)

was utilized to measure perceived competence through 4 items ($\alpha = .83$; e.g., “I feel I perform successfully the activities”). Participants responded to each item using a 7-point Likert-type scale, ranging from 1 (I don’t agree at all) to 7 (I completely agree).

Sportspersonship

The Spanish physical education version (Burgueño, Sánchez-Gallardo et al., 2018) of the Multidimensional Sportspersonship Orientations Scale (Vallerand et al., 1997) was used to measure the students’ sportspersonship. This instrument consisted of 17 items to assess social conventions (through 5 items, e.g., “Win or lose, shake hand with opponents”; $\alpha = .86$), commitment (through four items; e.g., “Give maximum effort”; $\alpha = .77$), respect of rules and referees (through five items; e.g., “Respect the rules”; $\alpha = .78$), respect of opponent (through three items; e.g., “Help opponent after a fall”; $\alpha = .72$). The instrument utilized a 5-point Likert-type scale, ranging from 1 (no), mid-point 3 (I don’t know), to 5 (yes).

Intrinsic motivation

The Spanish version (Martín-Albo et al., 2009) of the Situational Motivation Scale (Guay et al., 2000) was utilized to measure intrinsic motivation through 4 items ($\alpha = .85$; e.g., “Because I think that this activity is pleasant”). Each item used a 7-point Likert-type scale, ranging from 1 (not at all in agreement) to 7 (completely in agreement).

Procedure

The procedure was similar to Study 1 with participants and their guardians providing assent and informed consent and researchers informing participants about voluntary and anonymous participation. The scale was administered following the conclusion of participants’ Sport Education seasons and took approximately 20 min to complete.

Intervention program

The design of the program and the fidelity of the implementation were based on the same guidelines described in Study 1. The research team in charge of monitoring the Sport Education intervention confirmed that the teachers met the main features of Sport Education.

Data analysis

The IBM SPSS version 22.00 as well as the SPSS AMOS version 22.00 were used to perform the statistical analysis of data. To provide validity evidence based on internal structure, a confirmatory factor analysis approach was used to verify the underlying structure obtained for

Q5 **Table 2.** Exploratory Factor Analysis for the Sport Education Scale.

	F1	F2	F3	F4	F5	F6	F7
F1. Season							
1. Esta unidad didáctica tuvo una mayor duración [The duration of this unit was longer]	.76						
8. La unidad tuvo más sesiones de lo normal [The unit had more sessions than usual]	.80						
15. Esta unidad fue más larga de lo habitual [This unit was longer than common]	.71						
22. Hubo más clases de lo que es típico [There were more classes than typical]	.60						
F2. Affiliation							
2. El equipo fue el mismo durante toda la unidad [The team was the same during the unit]		.56					
9. Fui parte siempre de un mismo equipo [I was part on a same team]		.69					
16. Mi equipo fue el mismo en todas las clases [We stayed constantly in the same group]		.62					
23. Permanecimos siempre al mismo equipo [We belonged to a same team]		.70					
F3. Regular Competition							
3. Se combinaron los entrenamientos con las competiciones [We combined training and competitions]			.72				
10. Las actividades, los juegos y los partidos estuvieron intercalados [Activities and game with matches were intercalated]			.68				
17. Hubo entrenamiento y al mismo tiempo competición [There were trainings and competing at the same time]			.60				
24. Hubo un calendario de entrenamientos y partidos [There was a schedule for trainings and competitions]			.54				
F4. Culminating Event							
4. Hubo un evento final [There was a closing event]				.78			
11. Tuvimos un evento final [We participated in a final event]				.85			
18. Me ha parecido buena idea que la temporada termine en un día de fiesta [It is a good idea to be over the unit with an award ceremony]				.58			
25. Celebramos un acontecimiento para terminar las clases [An event was celebrated to finish the classes]				.60			
F5. Record Keeping							
5. Hubo anotación de puntuaciones, faltas, sanciones durante las sesiones. [There was a record of scores and fouls during the sessions]					.60		
12. Apuntamos los resultados de todas las competiciones [We had to collect data in the classes]					.50		
19. Tomamos datos en las clases [We kept the results from every competition]					.60		
26. Desarrollamos un registro de información durante las clases [We had to keep records during the classes]					.80		
F6. Festivity							
6. Hubo un clima de fiesta mediante cartelería, colorido y premios [We made a festive climate with signage, colorful, and award ceremony]						.70	
13. Hubo un ambiente festivo con colores, carteles y celebración de los esfuerzos genuinos [There was a festive atmosphere with colors, posters and celebration of genuine efforts]						.97	
20. Se creó un ámbito lúdico mediante la publicidad, la ceremonia de premios. [A ludic climate through signage and award ceremony was created]						.66	
27. Existió un ambiente festivo con el reconocimiento de los logros, el uso de publicidad y color. [A festive atmosphere existed with the recognition of achievements, the use of advertisement and color]						.96	
F7. Roles							
7. Cada uno de nosotros desempeñamos un rol [Each one of us performed a role]							.65
14. Todos tuvimos que realizar un papel dentro del equipo [We all carried out a role for the team]							.62
21. Tuvimos que realizar un rol [We had to complete a role]							.80
28. Nos encargamos de cumplir con unos roles [We were in charge of complying with some roles]							.61

Note. Items from the English version are presented in squared-brackets. Only factor loadings over .32 are displayed.

this instrument in Study 1, and two multi-group analyses were, respectively, applied to test invariance across gender and age. To conduct the confirmatory factor analysis, the maximum likelihood method was utilized

together with the bootstrapping procedure due to the absence of multivariate normality (Mardia's coefficient = 212.86, $p < .01$). This statistical procedure allows one to calculate the typical error and the confidence interval at

455 95% (95%CI) of each parameter from the random construction of 5000 resamples of the original sample. The goodness of fit was evaluated by a variety of fit indexes: ratio between χ^2 and degrees of freedom (χ^2/df), the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), the Standardized Root Mean Square Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA) with its 90% confidence interval. The χ^2/df statistic shows an excellent fit with values below 3, and a good fit with values less than 5 (Kline, 2016). CFI and TLI are indicative of an excellent fit with values over .95, while a good fit is obtained with values higher than .90 (Kline, 2016). SRMR and RMSEA are representative of an excellent fit with values below .08 and .06, while values lower than .08 for RMSEA would be representative of a reasonable fit with data (Kline, 2016). The standardized residual covariances are suitable with values lower than 2.58 in absolute terms (Hair et al., 2018). The standardized regression weights are acceptable with values higher than .50 (Hair et al., 2018).

475 In order to run multi-group analyses testing invariance across gender and age, the methodological approach by Putnick and Bornstein (2016) was followed. This approach judges the tenability of four increasingly constrained models by analyzing configural invariance (i.e., no equality constraints), metric invariance (i.e., equal item loadings), strong invariance (i.e., equal item loadings and item intercepts, simultaneously) and strict invariance (i.e., equal item loadings, item intercepts and item error variance, simultaneously). Assessment of differences among two multi-group models was made considering CFI and RMSEA. Values below .010 in CFI paired with values lower than .015 in RMSEA between two progressively constrained modes reflect no differences among them, the tenability of equality constrains and, thus, the instrument's invariance assumption (Putnick & Bornstein, 2016). For the age invariance analysis, two groups were created by the median. The first group included 392 (212 male and 180 female) students aged between 12 and 15 years ($M = 13.99$, $SD = 0.98$), while the second group consisted of 264 (139 male and 125 female) students aged between 16 and 19 years ($M = 16.53$, $SD = 0.79$).

500 To provide convergent validity evidence for the SES, the average variance extracted (AVE) was used as a criterion (Rocchi et al., 2017; Tilga et al., 2017). This is acceptable when values are over .50 (Hair et al., 2018). To gather discriminant validity evidence, heterotrait-monotrait (HTMT) ratio of correlations (Henseler et al., 2015) was estimated, indicating that values as high as .85 are indicative of a good level of discriminant validity (Henseler et al., 2015). Complementary to this, discriminant validity is also met when correlations

among latent variables are below .85 (Kline, 2016) and when the upper limit of the 95%CI of the correlation in question does not exceed 1.00 (Anderson & Gerbing, 1988). 510

To provide the instrument's reliability evidence, a series of coefficients were computed. For the primary-order factors, Cronbach's alpha (α), Raykov's composite reliability coefficient (ρ) and McDonald's omega (ω) were estimated. For the hierarchical Sport Education factor, coefficient H was used to inspect construct reliability. These reliability coefficients are suitable with values equal to .70 or greater (Hair et al., 2018). 515

To gather criterion validity evidence, a structural equation modeling was performed following a two-step approach (Kline, 2016). The first step tested the robustness of the measurement model by analyzing the bidirectional relationships among all target study variables. The hypothesized structural model, the second step, examined the predictive association between Sport Education, and perceived competence, sportspersonship and intrinsic motivation, respectively. Finally, descriptive statistics were computed for every variable. Prior to analyzing differences by gender and age, the univariate normality assumption was inspected by standardized scores for skewness and kurtosis coefficients. Standardized scores as high as 1.96 are indicative of a normal data distribution (Hair et al., 2018). To examine differences by gender and age in the target variables under normal conditions, independent- t -tests and Cohen's d effect sizes were estimated. 520 525 530 535

Results

Confirmatory factor analysis

The confirmatory factor analysis revealed acceptable fit indexes for the seven-factor correlated structure: χ^2 (329, $N = 656$) = 963.42, $p < .001$; $\chi^2/df = 2.93$; CFI = .96; TLI = .95; SRMR = .029; RMSEA = .054 (90%CI = .050 – .058). The standardized residual covariances ranged from –2.16 to 2.18. Figure 2 demonstrates that the standardized regression weights were between .56 and .85, with each being statistically significant ($p < .001$). 540 545

Given the moderate correlations among factors and consistent with Siedentop et al.'s (2020) characterization of Sport Education by its seven core features, we tested the robustness of a hierarchical factor model (i.e., Sport Education) formed by seven primary-order factors (i.e., seven structural features). This model obtained a good fit to data: χ^2 (343, $N = 656$) = 1117.53, $p < .001$; $\chi^2/df = 3.26$, CFI = .94; TLI = .93; SRMR = .033; RMSEA = .059 (90%CI = .055–.063). In this model, all the standardized regression weights were statistically significant ($p < .001$): .93 for season, .96 for affiliation, .96 for regular 550 555

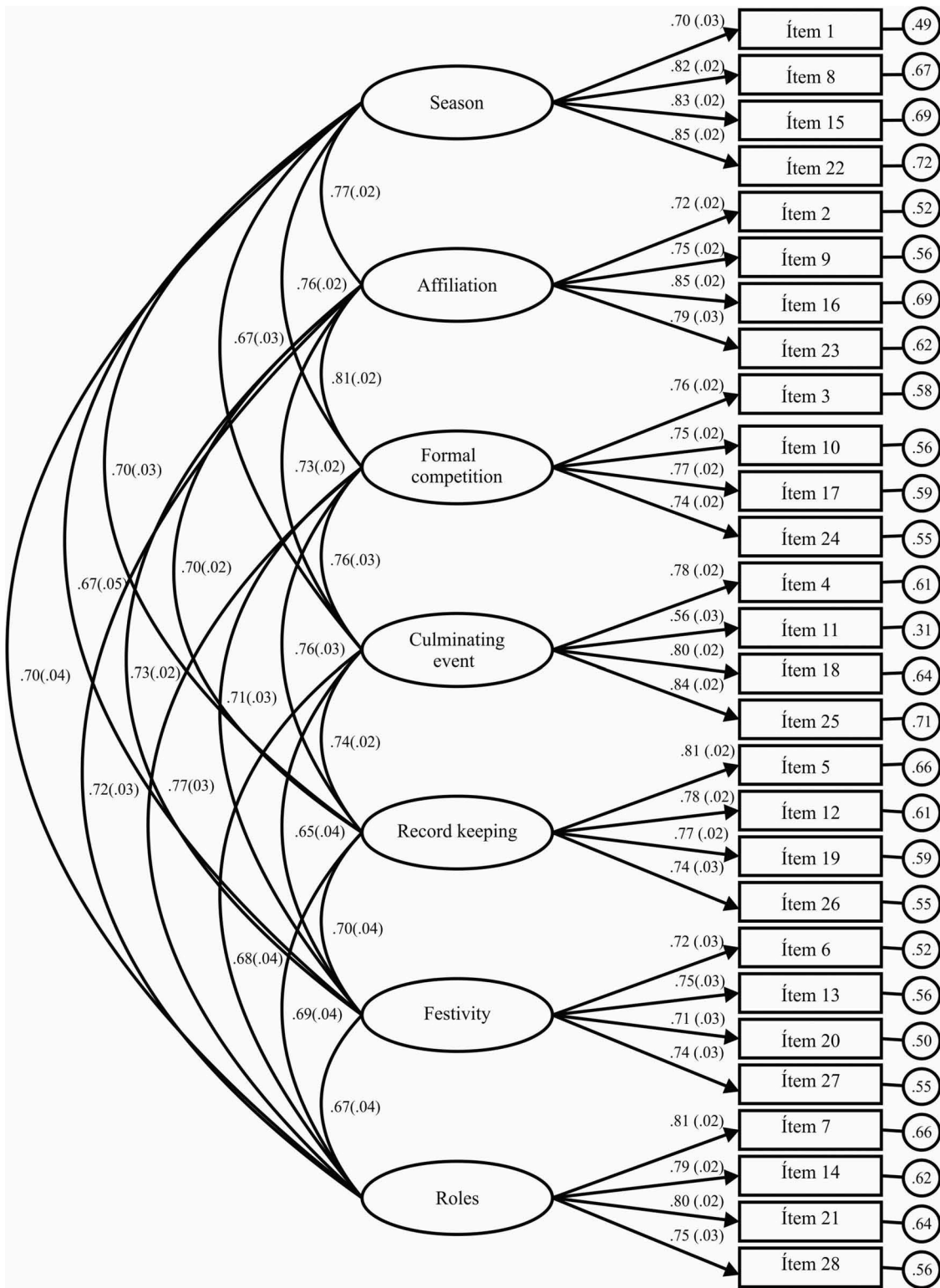


Figure 2. Confirmatory Factor Analysis of the Sport Education Scale. Note. The number in parentheses constitute the typical error estimated by bootstrapping. The ellipses represent the different latent factors, the rectangles symbolize the distinct items, and the cycles show the multiple squared correlations.

Table 3. Gender and Age Invariance Analysis.

Gender invariance analysis												
	χ^2	df	χ^2/df	CFI	TLI	SRMR	RMSEA (90%CI)	Models Comparison	$\Delta\chi^2$	Δdf	ΔCFI	$\Delta RMSEA$
1. Configural invariance	1555.30	658	2.36	.931	.921	.033	.046 (.043 – .049)	-	-	-	-	-
3. Metric invariance	1569.41	679	2.31	.932	.924	.033	.045 (.042 – .048)	2 versus 1	14.11	21	.001	-.001
4. Strong invariance	1613.08	707	2.28	.930	.926	.033	.044 (.041 – .047)	3 versus 2	43.76*	28	-.002	-.001
5. Strict invariance	1662.65	735	2.26	.929	.927	.035	.044 (.041 – .047)	4 versus 3	49.57**	28	-.001	.000
Age invariance analysis												
	χ^2	df	χ^2/df	CFI	TLI	SRMR	RMSEA (90%CI)	Models Comparison	$\Delta\chi^2$	Δdf	ΔCFI	$\Delta RMSEA$
1. Configural invariance	1606.07	658	2.44	.928	.917	.037	.047 (.044 – .050)	-	-	-	-	-
2. Metric invariance	1644.64	679	2.42	.926	.918	.039	.047 (.044 – .050)	2 versus 1	38.57*	21	-.002	.000
3. Strong invariance	1679.63	707	2.36	.926	.921	.039	.046 (.043 – .048)	3 versus 2	25.99	28	.000	-.001
4. Strict invariance	1745.19	735	2.37	.923	.921	.041	.046 (.043 – .049)	4 versus 3	74.55***	28	-.003	.000

Note. *** $p < .001$ ** $p < .01$, * $p < .05$.

560 competition, .95 for culminating event, .92 for record keeping, .94 for festivity and .91 for roles.

Invariance analysis

565 Table 3 shows changes in CFI and RMSEA values lower than .010 and .015, respectively, between the successive constrained models for the seven-factor correlated structure. Therefore, the null hypothesis of invariance across gender and age cannot be rejected.

Reliability analysis

570 The analysis of reliability reflected adequate values for season ($\alpha = .87$, $\rho = .88$, $\omega = .90$), affiliation ($\alpha = .85$, $\rho = .86$, $\omega = .91$), regular competition ($\alpha = .84$, $\rho = .84$, $\omega = .89$), culminating event ($\alpha = .87$, $\rho = .88$, $\omega = .90$), record keeping ($\alpha = .86$, $\rho = .86$, $\omega = .90$), festivity ($\alpha = .80$, $\rho = .80$, $\omega = .83$), and roles ($\alpha = .87$, $\rho = .87$, $\omega = .91$). Coefficient H was .98 for the hierarchical Sport Education factor.

Convergent and discriminant validity analysis

580 Table 4 shows AVE scores over .50, indicating a good level of convergent validity for each of the seven variables comprising the SES. Moreover, the HTMT ratio of correlations obtained values between .69 and .76. Additionally, correlations among latent variables ranged from .67 to .81 with upper limits of its 95%CI that did not exceed 1.00 in any case. Altogether, the results

585 provided evidence supporting the instrument's discriminant validity.

Criterion validity analysis

590 The measurement model, the first of the two steps, revealed acceptable fit indexes: χ^2 (146, $N = 656$) = 422.00, $p < .001$; $\chi^2/df = 2.89$; CFI = .97; TLI = .96; SRMR = .044; RMSEA = .056 (90%CI = .050–.062). The standardized regression weights were between .70 and .90, which reached the statistical significance ($p < .001$). The inter-factor correlations ranged between .60 and .80. Thus, the results obtained were representative of the appropriateness of measurement model.

595 The structural equation model, the second of the two steps, reflected a good fit to data: χ^2 (164, $N = 656$) = 462.11, $p < .001$; $\chi^2/df = 2.82$; CFI = .96; TLI = .96; SRMR = .037; RMSEA = .055 (90%CI = .049–.061). As shown Figure 3, Sport Education positively and significantly predicted perceived competence, sportspersonship and intrinsic motivation.

Descriptive statistics and differences by gender and age

600 Table 5 shows mean scores for each target variable higher than the midpoint of its respective measurement scale. While there were no significant differences by age, there were significant differences by gender, with boys scores being higher on the construct of competence than the corresponding scores of girls.

Table 4. Convergent and Discriminant Validity Analysis for the Sport Education Scale.

	AVE	1	2	3	4	5	6	7
1. Season	.64	-	.77(.73–.81)	.76(.73–.80)	.67(.61–.72)	.70(.62–.76)	.67(.61–.72)	.70(.62–.76)
2. Affiliation	.61	.73	-	.81(.77–.86)	.73(.68–.78)	.70(.64–.76)	.73(.67–.77)	.72(.68–.77)
3. Formal competition	.58	.72	.76	-	.78(.74–.82)	.76(.72–.80)	.71(.65–.75)	.68(.64–.72)
4. Culminating event	.57	.68	.74	.70	-	.70(.66–.75)	.65(.59–.70)	.68(.62–.73)
5. Record keeping	.60	.65	.73	.72	.75	-	.72(.67–.77)	.69(.63–.75)
6. Festivity	.54	.69	.72	.73	.73	.76	-	.67(.62–.72)
7. Roles	.62	.69	.75	.73	.71	.70	.73	-

Note. Numbers over diagonal show correlations together with its 95% confidence interval in parenthesis, both from confirmatory factor analysis. Numbers below diagonal display the heterotrait-monotrait ratio of correlations.

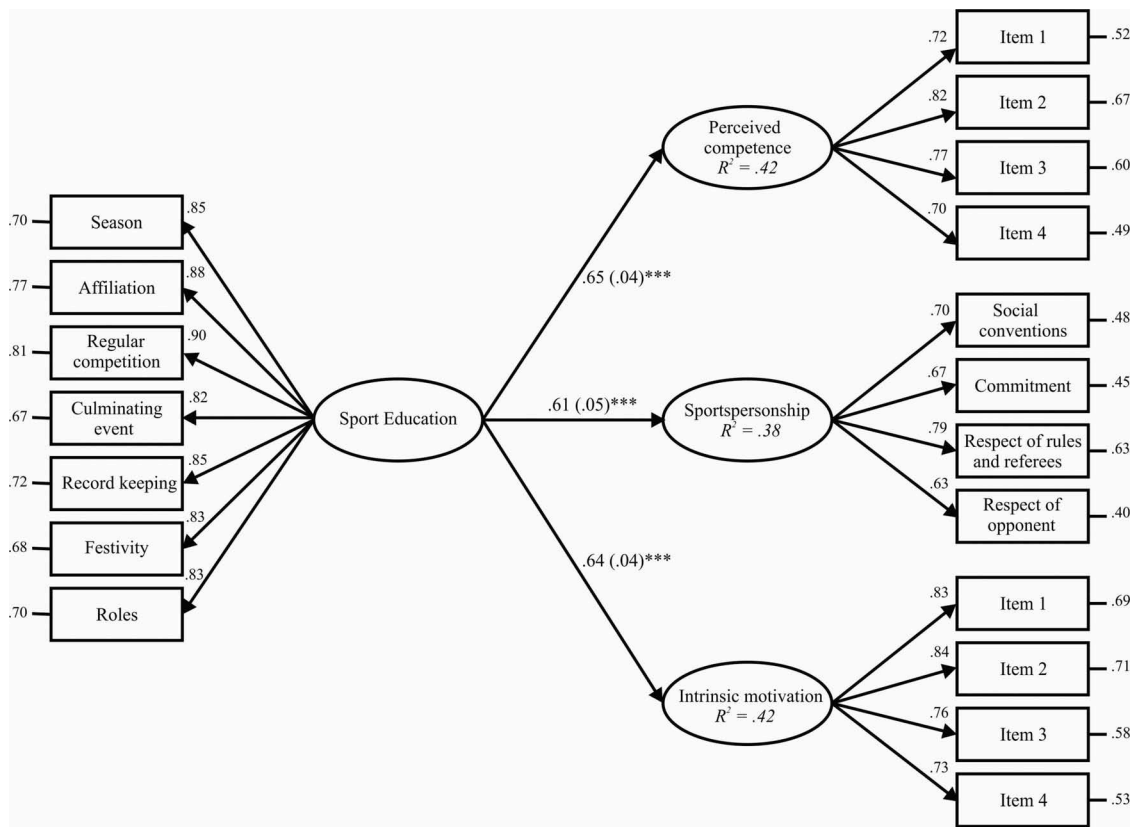


Figure 3. Predictive Relationships between Sport Education, and Perceived Competence, Sportsmanship and Intrinsic Motivation. Note. The numbers in the brackets represent the standard error calculated by bootstrapping. *** $p < .001$

Table 5. Descriptive Statistics and Differences by Gender and Age.

	Total sample				Boys M (SD)	Girls M (SD)	t	d	Younger students M (SD)	Older students M (SD)	t	d
	Range	M (SD)	γ_1	γ_2								
Season	1-5	4.04 (0.88)	-0.81	0.59	4.05 (0.91)	4.03 (0.84)	0.29	.02	4.05 (0.89)	4.03 (0.86)	0.36	.02
Affiliation	1-5	4.07 (0.83)	-0.75	0.60	4.04 (0.84)	4.10 (0.82)	-0.85	.07	4.07 (0.84)	4.07 (0.83)	-0.05	.00
Formal competition	1-5	4.15 (0.82)	-0.83	0.65	4.16 (0.85)	4.14 (0.78)	0.36	.03	4.15 (0.83)	4.15 (0.82)	0.24	.00
Culminating event	1-5	4.13 (0.80)	-0.89	0.97	4.12 (0.82)	4.14 (0.75)	-0.34	.03	4.16 (0.81)	4.09 (0.76)	1.02	.09
Record keeping	1-5	4.10 (0.83)	-0.85	0.85	4.11 (0.85)	4.09 (0.79)	0.47	.02	4.11 (0.81)	4.09 (0.84)	0.26	.02
Festivity	1-5	4.06 (0.81)	-0.77	0.38	4.02 (0.84)	4.12 (0.77)	-1.60	.12	4.10 (0.78)	4.02 (0.85)	1.37	.10
Roles	1-5	4.16 (0.79)	-0.85	0.52	4.15 (0.83)	4.17 (0.76)	-0.32	.03	4.18 (0.79)	4.14 (0.80)	0.55	.05
Competence	1-7	5.45 (1.13)	-0.63	0.31	5.59 (1.12)	5.28 (1.11)	3.53***	.28	5.40 (1.17)	5.50 (1.10)	1.44	.09
Sportsmanship	1-5	3.99 (0.61)	-0.80	1.24	4.00 (0.62)	3.98 (0.59)	0.28	.03	4.02 (0.61)	3.96 (0.59)	1.66	.10
Intrinsic motivation	1-7	5.55 (1.18)	-0.56	-0.24	5.56 (1.22)	5.54 (1.13)	0.07	.02	5.60 (1.18)	5.50 (1.17)	1.73	.09

Note. *** $p < .001$

Discussion

615 The purpose of the study was to create and validate the
 620 SES. The results from the confirmatory factor analyses
 provided validity evidence based on the instrument's
 internal structure. These results showed adequate fit
 indexes for a seven-factor correlated model: season,
 affiliation, regular competition, culminating event,
 record keeping, festivity, and roles, in other words all
 the features that correspond with the structural elements
 of Sport Education described by Siedentop et al. (2020).
 Indeed, all standardized regression weights reached the

statistical significance levels of greater than .50, indicat-
 ing that each suitably represented the factor for which it
 was theoretically intended. Furthermore, standardized
 residual covariances did not exceed 2.58, which
 endorsed the absence of misspecifications in the SES's
 internal structure. This indicated the congruence
 between the seven-factor correlated model and the
 observed data.

Given moderate correlations obtained among the
 seven factors, the robustness of a hierarchical factor
 model was hypothesized, formed by the seven structural
 features that define this pedagogical model, which was

625

630

635 supported by acceptable fit indexes obtained from the factor confirmatory analysis. This structure acquires special relevance in the empirical research given that it underpins the use of a general value formed by mean scores of the seven factors. In other words, one of the advantages of the hierarchical factor structure (i.e., Sport Education) is to allow one to calculate its arithmetic mean (i.e., general value) via the sum of the arithmetic average of each one of the primary-order factors that composed it (i.e., seven features of this pedagogical model).

645 From a methodological perspective, it might be advantageous and practical to consider Sport Education as a single construct within a more complex model that implies the association with other variables. The structural equation model used in this study to provide validity evidence based on relations to other variables is an example of such sophisticated view of Sport Education. In this complex model, the hierarchical factor structure (in this case – Sport Education) should be related to other variables to allow for analysis of its predictive power on such variables. From a theoretical perspective, these findings provide empirical support to the Siedentop et al.'s (2020) tenets outlining the critical nature of the seven Sport Education structural features. One of the important contributions of this study is that the researchers now can use either a seven-factor correlated model or a hierarchical factor model depending on their research purpose. For instance, to examine the effects of each Sport Education core feature independently researchers can use a seven-factor correlated model while alternatively, a hierarchical model can be used for investigations relating to the effects of the model itself. This sophisticated modeling could lead to a more tailored research on Sport Education and could allow investigators to further delineate and more accurately define their research methodology.

660 The results from the two multi-group analyses provided evidence to support the instrument's invariance across gender and age, signifying that the SES performed equally well for both, male and female students, as well as students of different ages. Thus, these findings are of great practical utility since they support the use of the SES in analyzing the possible differences between boys and girls of different ages about their perceptions of the seven core features of Sport Education. Moreover, this is the first study that robustly examined gender and age differences relative to their perception of the seven structural features of Sport Education. The findings demonstrated no gender or age differences. Furthermore, HTMT ratio of correlations together with correlations among latent factors provided evidence confirming the instrument's discriminant validity. The SES's

convergent validity was established by suitable values for AVE in every structural feature of the model. On the other hand, the results of the reliability examination showed acceptable levels of reliability for each primary-order factor of the SES in terms of Cronbach's alpha, Raykov's coefficient and McDonald's omega. It is important to note that the hierarchical factor model (i.e., Sport Education) obtained a coefficient over .70, indicating a good construct reliability.

690 The results derived from the structural equation modeling provided criterion validity evidence for the SES. These findings are in line with previous research (Bessa et al., 2021; Hastie et al., 2011; Hastie & Wallhead, 2016; Wallhead & O'Sullivan, 2005), confirming that Sport Education positively predicted students' perceived competence, sportspersonship and intrinsic motivation. These results demonstrate that students who participate in Sport Education perceive to obtain a higher level of motor competence, consistent with previous research (Chu & Zhang, 2018; Evangelio et al., 2018; Manninen & Campbell, 2021). A possible explanation might be related to the structure of the model since Sport Education, as a pedagogical model, provide students with more opportunities to understand and execute appropriate strategies when facing the complexities of the game within a consistent group of peers (i.e., teams), the completion of the different roles adjusted to students' own interests and previous sport experience as well as the continuous combination of practice and meaningful competition (Siedentop et al., 2020). Furthermore, this pedagogical model contributes to creating both a well-structured learning environment (Washburn et al., 2019) and task-oriented motivational climate (Hastie et al., 2014) which will largely promote the students' competence via development of problem solving skills and learning that meets their educational needs (Chu & Zhang, 2018). Likewise, these results support the notion of that students who were taught Sport Education exhibit high level of sportspersonship, in congruence with findings from previous research (Burgueño & Medina-Casabón, 2020; Wahl-Alexander et al., 2017). This increase is possible because of Sport Education's intentionality of emphasis on student pro-social behaviors and fair play (Bessa et al., 2021; Fernandez-Rio & Casey, 2020; Harvey et al., 2014). Examples of such behaviors can include simple instances of students respecting the rules, not arguing with officials and shaking hands with opponents after the game (Wahl-Alexander et al., 2017). Similarly and consistent with previous findings (Chu & Zhang, 2018; Evangelio et al., 2018; Manninen & Campbell, 2021), the results of this study demonstrated that students had higher levels of

740 intrinsic motivation. Participation in Sport Education
has resulted in students experiencing increased levels of
enjoyment, effort, persistence or engagement (Bessa
et al., 2019; Hastie & Wallhead, 2015), factors asso-
745 ciated with intrinsic motivation (Guay et al., 2000).
These findings lend credence to model's robustness in
achieving its three major aims as judged from students'
perspectives, namely developing competent (construct
of perceived competence), literate (construct of sports-
750 personship), and enthusiastic (construct of intrinsic
motivation) sportspeople (Siedentop et al., 2020).

Limitations

While this research represents a first advance in measur-
ing the core features defining Sport Education through
a valid and reliable scale, the validation of any psycho-
755 metric instrument should be understood as an ongoing
process over time. Therefore, although items developed
for the SES to robustly measure core characteristics of
Sport Education outlined by Siedentop et al. (2020),
researchers could also take into account additional
760 aspects of the model to operationalize the measurement
of every structural feature. For instance, further instruc-
tional strategies such as team cheers, creation of a team
shield, same clothes and colors could be used to oper-
ationalize the core feature of affiliation. Further,
765 although this study exclusively aimed to develop
a measure of the students' perception of the core features
defining Sport Education, complementary observational
instruments to self-reported questionnaires are
required. Additional research, therefore, should investi-
770 gate the degree of agreement between students' percep-
tions of core features of Sport Education and the
observed instructional elements throughout the imple-
mentation of this model. It should also be noted that the
participants in this study comprised a purposeful sam-
775 ple, which may limit generalization of results.
Furthermore, new studies are needed to analyze the
psychometric properties of the SES in new populations
in order to support the results obtained in the present
research. In addition, future research should examine
780 the psychometric properties of the instrument in differ-
ent contexts. For example, a version of this instrument
should be tested for instance, in Anglo, Russian, or
Portuguese settings, contexts where contemporary lit-
erature describes the use of the Sport Education
785 predominantly.

Implications for practice

The availability of a robust psychometric measure of
students' perceptions of core features of Sport

Education will allow sport pedagogy scholars and prac-
790 titioners to explore students' experiences with Sport
Education. New findings might provide teachers with
useful information and valuable feedback for the model
to evolve and to adapt it to their local context. Similarly,
795 this scale opens up the possibility to examine boys' and
girls' perceptions of Sport Education, which may pro-
vide teachers with valuable information to provide
meaningful and inclusive experiences in physical educa-
tion lessons. It is also important to note that the use of
this instrument will be useful for teachers to ascertain
800 which features (or combination of features) are more
relevant for the full development of students' affective,
physical-motor, social and/or cognitive domains
throughout Sport Education seasons. Additionally, this
instrument might help new qualified teachers to
805 improve their knowledge of Sport Education in the
early seasons. Overall, the possible findings derived
from a valid and reliable measure of Sport Education
might have meaningful practical implications not only
for researcher and students' but also for physical educa-
810 tion teachers and teacher educators.

Conclusions

In conclusion, we would like to reiterate that until now,
a psychometrically rigorous instrument that examines the
core features proposed by Siedentop et al. (2020) for Sport
Education, has not been created. Thus, the objective of this
815 research was to develop and examine the psychometric
properties of the instrument with these characteristics.
After conducting two sequential studies, an instrument,
called the SES, was developed and demonstrated adequate
validity evidence based on content, internal structure and
820 relationships with other variables, as well as an acceptable
reliability. We further conclude that the SES seems to be
a psychometrically robust measure and, therefore, appro-
priate one to use when examining the structural features
that define Sport Education pedagogical model. 825

Funding

This work was supported by the Consejería de Educación,
Junta de Andalucía [Influencia del modelo de educación
830 deportiva sobr].

ORCID

Rafael Burgueño  <http://orcid.org/0000-0003-2354-0037>
Antonio Calderón  <http://orcid.org/0000-0003-1389-9898>
Oleg Sinelnikov  <http://orcid.org/0000-0003-3019-6705>

References

- Q6 835 Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- 840 Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- 845 Andrade, J. R., De Sousa, L., Morais-Freire, G. L., Vicentini de Oliveira, D., Fiorese, L., Donal-Cronin, L., et al (2020). Cross-cultural adaptation and psychometric properties of the Portuguese version of the life skills scale for sport. *Measurement in Physical Education and Exercise Science*, 24(1), 11–24. <https://doi.org/10.1080/1091367X.2019.1647208>
- Q7 850 Bessa, C., Hastie, P., Ramos, A., & Mesquita, I. (2021). What actually differs between traditional teaching and sport education in students' learning outcomes? A critical systematic review. *Journal of Sports Science & Medicine*, 20(1), 110–125. <https://doi.org/10.52082/jssm.2021.110>
- 855 Bessa, C., Hastie, P. A., Araújo, R., & Mesquita, I. (2019). What do we know about the development of personal and social skills within the sport education model: A systematic review. *Journal of Sports Science & Medicine*, 18(4), 812–829.
- Q8 860 Burgueño, R., Lirola, M.-J., Morales-Ortiz, E., Cueto-Martín, B., & Medina-Casaubón, J. (2018). Psychometric analysis of the physical education season survey in Spanish secondary education students. *Psychology, Society, & Education*, 10(2), 189–201. <https://doi.org/10.25115/psye.v10i1.1093>
- 865 Burgueño, R., & Medina-Casaubón, J. (2020). Sport education and sportsmanship orientations: an intervention in high school students. *International Journal of Environmental Research and Public Health*, 17(3), 837. <https://doi.org/10.3390/ijerph17030837>
- 870 Burgueño, R., Sánchez-Gallardo, I., & Medina-Casaubón, J. (2018). Adaptation of the multidimensional sportspersonship orientations scale to the Spanish physical education context. *SPORT TK*, 7(2 (Supl. 1)), 59–66. <https://doi.org/10.6018/spork.343251>
- 875 Calderón, A., Hastie, P. A., & Martínez de Ojeda, D. (2010). Learning to teach sport education: initial experience in elementary education. *Cultura, Ciencia y Deporte*, 6(5), 169–181. <https://doi.org/10.12800/ccd.v5i15.103>
- 880 Calderón, A., & Martínez de Ojeda, D. (2014). Training in-service teachers in physical education. A proposal to teach the sport education model. *Revista de Educacion*, 363(1), 128–153. <https://doi.org/10.4438/1988-592X-RE-2012-363-173>
- 885 Calderón, A., Martínez de Ojeda, D., & Hastie, P. A. (2013). Students and teachers' perception after practice with two pedagogical models in physical education. *RICYDE. Revista Internacional de Ciencias Del Deporte*, 32(9), 137–153. <https://doi.org/10.5232/ricyde2013.03204>
- 890 Casey, A., & Kirk, D. (2020). *Models-based practice in physical education* (1st ed.). Routledge.
- Casey, A., MacPhail, A., Larsson, H., & Quennerstedt, M. (2021). Between hope and happening: problematizing the M and the P in models-based practice. *Physical Education and Sport Pedagogy*, 26(2), 111–122. <https://doi.org/10.1080/17408989.2020.1789576> 895
- Chu, T. L., & Zhang, T. (2018). Motivational processes of sport education among high school students: A systematic review. *European Physical Education Review*, 24(3), 372–394. <https://doi.org/10.1177/1356336X17751231> 900
- Evangelio, C., Sierra-Díaz, J., González-Villora, S., & Fernández-Río, J. (2018). The sport education model in elementary and secondary education: A systematic review. *Movimento*, 24(2), 931–946. <https://doi.org/10.22456/1982-8918.81689> 905
- Fernandez-Rio, J., & Casey, A. (2020). Sport education as a cooperative learning endeavour. *Physical Education and Sport Pedagogy*. <https://doi.org/10.1080/17408989.2020.1810220> 905
- Fernandez-Rio, J., Méndez-Giménez, A., & Méndez-Alonso, D. (2013). Effects of three instructional approaches in adolescents' physical self-concept. *Cultura Y Educacion: Culture and Education*, 25(4), 509–521. <https://doi.org/10.1174/113564013808906870> 910
- Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (SIMS). *Motivation and Emotion*, 24(3), 175–213. <https://doi.org/10.1023/A:1005614228250> 915
- Guijarro, E., MacPhail, A., González-Villora, S., & Arias-Palencia, N. M. (2021). Relationship between personal and social responsibility and the roles undertaken in sport education. *Journal of Teaching in Physical Education*, 40(1), 76–85. <https://doi.org/10.1123/jtpe.2019-0097> 920
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning EMEA. 925
- Harvey, S., Kirk, D., & O'Donovan, T. M. (2014). Sport education as a pedagogical application for ethical development in physical education and youth sport. *Sport, Education and Society*, 19(1), 41–62. <https://doi.org/10.1080/13573322.2011.624594> 930
- Hastie, P. A., & Casey, A. (2014). Fidelity in models-based practice research in sport pedagogy: A guide for future investigations. *Journal of Teaching in Physical Education*, 33(3), 422–431. <https://doi.org/10.1123/jtpe.2013-0141> 935
- Hastie, P. A., Martínez de Ojeda, D., & Calderon, A. (2011). A review of research on sport education: 2004 to the present. *Physical Education and Sport Pedagogy*, 16(2), 103–132. <https://doi.org/10.1080/17408989.2010.535202>
- Hastie, P. A., Sinelnikov, O., Wallhead, T., & Layne, T. (2014). Perceived and actual motivational climate of a mastery-involving sport education season. *European Physical Education Review*, 20(2), 215–228. <https://doi.org/10.1177/1356336X14524858> 940
- Hastie, P. A., & Sinelnikov, O. A. (2006). Russian students' participation in and perceptions of a season of sport education. *European Physical Education Review*, 12(2), 131–150. <https://doi.org/10.1177/1356336X06065166> 945
- Hastie, P. A., & Wallhead, T. (2016). Models-based practice in physical education: the case for sport education. *Journal of Teaching in Physical Education*, 35(4), 390–399. <https://doi.org/10.1123/jtpe.2016-0092> 950
- Hastie, P. A., & Wallhead, T. L. (2015). Operationalizing physical literacy through sport education. *Journal of Sport*

- 955 *and Health Science*, 4(2), 132–138. <https://doi.org/10.1016/j.jshs.2015.04.001>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- 960 Kinchin, G. D., Wardle, C., Roderick, S., & Sprosen, A. (2004). A survey of year 9 boys' perceptions of sport education in one English secondary school. *Bulletin of Physical Education*, 40(1), 27–40.
- 965 Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Q10 Mahedero, P., Calderón, A., Arias-Estero, J. L., Hastie, P. A., & Guarino, A. J. (2015). Effects of student skill level on knowledge, decision making, skill execution and game performance in a mini-volleyball sport education season. *Journal of Teaching in Physical Education*, 34(4), 626–641. <https://doi.org/10.1123/jtpe.2014-0061>
- 970 Manninen, M., & Campbell, S. (2021). The effect of the sport education model on basic needs, intrinsic motivation and prosocial attitudes: A systematic review and multilevel meta-analysis. *European Physical Education Review*. <https://doi.org/10.1177/1356336X211017938>
- 975 Martin-Albo, J., Núñez, J. L., & Navarro, J. G. (2009). Validation of the Spanish version of the situational motivation scale (EMSI) in the educational context. *The Spanish Journal of Psychology*, 12(2), 799–807. <https://doi.org/10.1017/S113874160000216X>
- Q11 Merino, C., & Livia, J. (2009). Confidence intervals for the content validity: A visual basic computer program for the Aiken's V. *Annals of Psychology*, 25(1), 169–171.
- 980 Mohr, D. J., Townsend, S. J., Rairigh, R., & Mohr, C. (2003). Students' perceptions of sport education when taught using pedagogical approach to sport education (PASE) planning and instructional framework. *Research Quarterly for Exercise and Sport*, 74(supplement), A51–A52.
- 985 Moreno, J. A., González-Cutre, D., Chillón-Garzón, M., & Parra-Rojas, N. (2008). Adaptation of the basic psychological needs in exercise scale to physical education. *Revista Mexicana de Psicología*, 25(2), 295–303.
- Q12 Muñoz, J., & Fonseca-Pedero, E. (2019). Ten steps for test development. *Psicothema*, 31(1), 7–16. <https://doi.org/10.7334/psicothema2018.291>
- 990 Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Q13 Rocchi, M., Pelletier, L., & Desmarais, P. (2017). The validity of the interpersonal behaviors questionnaire (IBQ) in sport. *Measurement in Physical Education and Exercise Science*, 21(1), 15–25. <https://doi.org/10.1080/1091367X.2016.1242488>
- 995 Sabourin, C., Trottier, C., Frenette, E., Camiré, M., Drapeau, V., Goulet, C., Lemyre, P. N., et al (2020). French adaptation of the short form youth experiences 1010 survey for sport and life skills scale for sport. *Measurement in Physical Education and Exercise Science*, 24(1), 33–48. <https://doi.org/10.1080/1091367X.2019.1663351> Q16
- Siedentop, D. (1994). *Sport Education: Quality PE through 1015 positive sport experiences*. Human Kinetics.
- Siedentop, D., Hastie, P. A., & Van Der Mars, H. (2020). *Complete guide to sport education* (3rd ed.). Human Kinetics.
- 1020 Sinelnikov, O. A. (2009). Sport education for teachers: professional development when introducing a novel curriculum model. *European Physical Education Review*, 15(1), 91–114. <https://doi.org/10.1177/1356336X09105213>
- 1025 Sinelnikov, O. A., & Hastie, P. A. (2016). The learning of pedagogical models in physical education: The socialization perspective. In K. A. Richards & K. L. Gaudreault (Eds.), *Teacher socialization in physical education: New perspectives* (1st ed., pp.130–143). Routledge. <https://doi.org/10.4324/9781315679471>.
- 1030 Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- 1035 Tilga, H., Hein, V., & Koka, A. (2017). Measuring the perception of the teachers' autonomy-supportive behavior in physical education: development and initial validation of a multi-dimensional instrument. *Measurement in Physical Education and Exercise Science*, 21(4), 244–255. <https://doi.org/10.1080/1091367X.2017.1354296>
- 1040 Vallerand, R. J., Brière, N. M., Blanchard, C., & Provencher, P. (1997). Development and validation of the multidimensional sportspersonship orientations scale. *Journal of Sport & Exercise Psychology*, 19(2), 197–206. <https://doi.org/10.1123/jsep.19.2.197>
- 1045 Vlachopoulos, S. P., & Michailidou, S. (2006). Development and initial validation of a measure of autonomy, competence, and relatedness in exercise: the basic psychological needs in exercise scale. *Measurement in Physical Education and Exercise Science*, 10(3), 179–201. https://doi.org/10.1207/s15327841mpee1003_4
- 1050 Wahl-Alexander, Z., Sinelnikov, O. A., & Curtner-Smith, M. (2017). A longitudinal analysis of students' autobiographical memories of participation in multiple sport education seasons. *European Physical Education Review*, 23(1), 25–40. <https://doi.org/10.1177/1356336X15624246>
- 1055 Wallhead, T., & O'Sullivan, M. (2005). Sport education: physical education for the new millennium?. *Physical Education and Sport Pedagogy*, 10(2), 181–210. <https://doi.org/10.1080/17408980500105098>
- 1060 Washburn, N., Sinelnikov, O. A., & Richards, K. A. R. (2019). Enhancing student learning in sport education through the manipulation of need-supportive instruction. *The Physical Educator*, 76(3), 848–875. <https://doi.org/10.18666/tpe-2019-v76-i3-8873>