Measurement in Physical Education and Exercise Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmpe20

Psychometric Evaluation of the Norwegian Versions of the Modified Group Environment Questionnaire and the Youth Sport Environment Questionnaire

Tommy Haugen, Derek M. Peters, Yngvar Ommundsen, Luc J. Martin, Andreas Stenling & Rune Høigaard

MEASUREMENT IN PHYSICAL EDUCATION AND EXERCISE SCIENCE
https://doi.org/10.1080/1091367X.2021.1917414

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Psychometric Evaluation of the Norwegian Versions of the Modified Group Environment Questionnaire and the Youth Sport Environment Questionnaire

Tommy Haugen[a], Derek M. Peters[b], Yngvar Ommundsen[a,c], Luc J. Martin[d], Andreas Stenling[a,e], and Rune Høigaard[a]

[a]Department of Sport Science and Physical Education, University of Agder, Kristiansand, Norway; [b]School of Allied Health & Community, University of Worcester, Worchester, Massachusetts, UK; [c]Department of Sport and Social Sciences, Norwegian School of Sport Science, Oslo, Norway; [d]School of Kinesiology and Health Studies, Queen's University, Kingston, Canada; [e]Department of Psychology, Umeå University, Umeå, Sweden

ABSTRACT

This study aimed to translate the modified Group Environment Questionnaire (GEQ) and the Youth Sport Environment Questionnaire (YSEQ) into Norwegian, examine the factor structure and reliability of the scales through independent clusters model confirmatory factor analysis (ICM-CFA) and exploratory structural equation modeling (ESEM), and examine differential item functioning (DIF) as a function of sex. Three-hundred-and-thirty-three athletes ($M(SD)_{age}$ = 18.7(2.60) years; 33% females) completed the GEQ. Three-hundred-and-three athletes ($M(SD)_{age}$ = 15.0(1.48) years; 26% females) completed the YSEQ. Results indicated acceptable fit indices for a four-factor, a second-order two-factor (task and social), and a second order one-factor ESEM model for the GEQ. Cross-loadings and high latent factor correlations are issues in need of attention. The study supported the structural validity and reliability of the Norwegian YSEQ, with no major differences between the ICM-CFA and ESEM. No evidence of DIF as a function of sex was identified in either of the scales.

KEYWORDS
Team cohesion; group dynamics; psychometrics; team sports; validation

A prominent feature of being involved in sport is working within groups, so it is not surprising that a rich body of literature pertaining to group dynamics exists (Eys et al., 2019, 2020). Within this area of research, few would dispute that group cohesion has been the concept most heavily examined (Eys & Brawley, 2018). Indeed, across numerous studies, researchers have demonstrated its associations with situational (e.g., *level of competition*; Granito & Rainey, 1988, *group size;* Widmeyer et al., 1990), personal (e.g., *anxiety*; Eys et al., 2003, *social loafing;* Høigaard, Tofteland et al., 2006), and team factors (e.g., *performance*; Carron, Colman et al., 2002). It is important to highlight here, the critical prerequisites that enabled the continued investigation of cohesion in sport – the development of its conceptual and operational definitions.

Within the field of sport and exercise psychology, there is relative consensus on the definition and measurement of group cohesion (Carless & De Paola, 2000; Carron & Brawley, 2000; Cota et al., 1995). Carron et al. (1998) described it as "a dynamic process that is reflected in the tendency for a group to stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of members affective needs" (p. 213). Further, these authors proposed that a group's level of cohesiveness could be measured through individual member perceptions, supported by the following assumptions: (a) groups have observable properties; (b) individuals are socialized and integrated into groups and develop beliefs about those groups; (c) individuals' beliefs are based on the information gathered about their group; (d) individuals' beliefs are reflective of common within-group values; and (e) individuals' perceptions of the cohesion can be assessed through self-report questionnaires (Carron et al., 1998).

Following the abovementioned conceptualization, Carron et al. (1985) and Carron, Brawley et al. (2002) undertook a systematic research program to develop the Group Environment Questionnaire (GEQ) to assess individual perceptions of group cohesion. Within the GEQ, items were specifically developed to examine a group's integration (i.e., a member's perception of the group as a whole) or their individual attraction to the group (i.e., a member's personal attraction to the group). In addition, cohesion could be assessed through task and social perspectives (Carron et al., 1998; Carron, Brawley et al., 2002; Carron et al., 1985; Dion, 2000). The task perspective focuses on how well members are

CONTACT Tommy Haugen tommy.haugen@uia.no Department of Sport Science and Physical Education, University of Agder, Post Box 422, Kristiansand S, 4604, Norway.
Supplemental data for this article can be accessed on the publisher's website.

integrated and motivated toward the group's task objectives. Conversely, the social perspective focuses on how well members are integrated socially, and the degree to which they are attracted to the social aspects of the group. Both the group-individual and the task-social distinction have been equally suggested in the cohesion literature (see Dion, 2000 for a discussion). Taken together, the combination of these orientations comprises the original four-factor model of the GEQ (Carron et al., 1985): Group Integration-Task (GIT), Group Integration-Social (GIS), Attraction to the Group-Task (ATGT), and Attraction to the Group-Social (ATGS).

In addition to the four-factor model described previously, several alternative first- and second-order models have been proposed (Leeson & Fletcher, 2005; Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004; Schutz et al., 1994; Whitton & Fletcher, 2014). Second-order models may be considered when a conceptual and/or theoretical rationale is provided, and when the first-order factors are highly correlated. In their original Carron et al. (1985), (1998) advocated for the higher order two-factor model involving Attraction to the Group (ATG) and Group Integration (GI). Others, however, have argued for a different conceptualization of the hierarchical structure of the phenomenon, with task and social cohesion as the main higher order factors (Casey-Campbell & Martens, 2009; Dion, 2000). Indeed, researchers have provided varying levels of empirical support for different GEQ models (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004). For instance, one study (Schutz et al., 1994) failed to support any of the proposed GEQ models and instead found support for an alternative bifactor model with one global cohesion factor and four specific factors. Schutz et al. (1994) argued that their findings could indicate the presence of a general cohesion factor that explains a considerable amount of the covariance among the items. They also called for additional research on the bifactor representation of the GEQ, however, that call remains unanswered. Others have demonstrated acceptable CFA fit indices for most of the investigated models (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004), with the latter study showing slightly better fit for the GI-ATG model. In contrast, another study found support for the two-factor model of task and social cohesion (Leeson & Fletcher, 2005).

Despite conflicting results regarding the most appropriate structural model, most research indicates moderate-to-high correlations between the different cohesion-factors, indicating less than optimal discriminant validity (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004). Moreover, previous research has also tested single-factor structure models (i.e., a general group cohesion construct), both at first-order and second-order levels (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004). Although early research proposed a unidimensional perspective of cohesion (see Casey-Campbell and Martens (2009) for a discussion), the idea of a single-factor structure may be more driven by parsimony than conceptual logic. Nevertheless, empirical findings yield inconclusive results regarding single-factor models, from poor fit (Li & Harmer, 1996) to acceptable fit (Ntoumanis & Aggelonidis, 2004). In sum, these inconsistencies in previous research may indicate that further examinations of the dimensional structure of the GEQ are warranted.

Although the GEQ is the most widely accepted and used cohesion questionnaire in sport (Eys & Brawley, 2018; Eys et al., 2019), numerous studies have provided reliability and validity evidence (Carron & Brawley, 2000; Carron et al., 1998; Eys et al., 2007), studies have reported variable reliability estimates of one or more of the four subscales (Eys et al., 2007; Westre & Weiss, 1991). For example, researchers have reported reliability estimates (Cronbach's alpha) below .70 for all four subscales (Schutz et al., 1994; Westre & Weiss, 1991) or for the task subdimensions (Høigaard, Säfvenbom et al., 2006). It is important to highlight that the original scale construction for the GEQ involved both positively and negatively worded items, which has been deemed a cause of the variability in reliability estimates. Due to these shortcomings, Eys et al. (2007) developed a modified version whereby the 12 negative items were rephrased positively (e.g., "I do not enjoy being part of the social activities of this team" became "I enjoy being part of the social activities of this team"), resulting in significantly higher reliability estimates (Cronbach's alpha) for three of the four subscales.

Another important consideration pertaining to the original GEQ is that it was developed for athletes between the ages of 18 and 30 years. Accordingly, researchers have developed instruments appropriate for different age groups, emphasizing youth athletes as a population of interest (Carron et al., 2007; Eys et al., 2009a). The Youth Sport Environment Questionnaire (YSEQ; Eys et al., 2009b) was developed for adolescents aged 13–17 years. Through a comprehensive development process, the authors determined how youth perceived and understood the concept of cohesion and then created items and established content and factorial validity. The resulting YSEQ contained 16 positively worded items in addition to two spurious negatively worded items (Eys et al., 2009b). The final structure for the YSEQ involved a two-factor model distinguishing task and social cohesion, a conceptualization that had been proposed previously by several group dynamics researchers (Carron et al., 1985; Dion, 2000) and subsequently supported with other age groups (Martin et al., 2012, 2013).

In addition to recognizing the need for age-appropriate questionnaires, the importance of cohesion in sport has also spurred the need for language-specific versions. Notably, the GEQ has been translated and validated in several different language versions, including French (Heuzé & Fontayne, 2002), Spanish (Iturbide et al., 2010; Leo et al., 2015), German (Ohlert, 2012), Greek (Ntoumanis & Aggelonidis, 2004), and Italian (Steca et al., 2013). Similarly, the YSEQ has been translated into Portuguese, Czech, Slovak, and Farsi (Benson et al., 2016; Eshghi et al., 2015; Junior et al., 2018). Clearly, the need to explore cohesion in sport is not restricted to English-speaking athletes, and advances in understanding the construct across various nations and cultures necessitates the development of valid and reliable questionnaires. In this regard, although researchers have used cohesion questionnaires that have been translated to Norwegian (De Backer et al., 2015; Erikstad et al., 2018), these have only involved basic translation methods. By extension, there are no validated questionnaires available for use with Norwegian speaking athletes. The translation process involved when developing a language-specific version of a questionnaire has been described as important (Brislin, 1970; Clark & Watson, 2019), because linguistic or cultural aspects may affect the interpretation of questionnaires. A comprehensive process ensures that the results obtained through research within one culture are not due to linguistics or contextual errors in translation, but rather to real differences or similarities in the phenomena being measured.

In a meta-analysis investigating the relationship between performance and cohesion, the findings revealed higher effect sizes for females compared to males (Carron, Colman et al., 2002). However, to fully establish a conceptual difference between the sexes, the instruments used across different populations (i.e., males and females) need to exhibit identical psychometric properties. Otherwise, the reported difference could have, at least partially, been driven by the questionnaires. Schutz et al. (1994) argued that potential sex differences in the psychometric properties of the GEQ have mostly been ignored in the literature. Similar arguments may be used for age-specific instruments (i.e., YSEQ). Regarding the GEQ, a previous study has shown different factorial structures for males and females (Schutz et al., 1994), whereas others have reported measurement invariance across sex (Ntoumanis & Aggelonidis, 2004; Leo et al., 2015). Regarding the YSEQ, Junior et al. (2018) claimed the instrument to be equivalent between boys and girls in the Portuguese version of the scale. Based on these conflicting or limited results from previous research, one may argue that potential sex-differences should be considered and needs to be investigated, also when adapting the questionnaire for a new language/culture.

Based on the importance of cohesion in sports teams, the inconsistent factor-structure of previous cohesion-measures in sports, and suggestions for expanding research involving youth, the purpose of this study was threefold: (1) To translate the modified GEQ and YSEQ items into Norwegian using recommended scale adaptation and translation procedures (Brislin, 1970; Clark & Watson, 2019); (2) to examine the factor structure and reliability of the scales in Norwegian adult and youth interdependent sport team athletes; and (3) to examine differential item functioning (DIF) as a function of sex.

Materials and methods

The study was approved by The Norwegian Social Science Data and the Faculty ethical board at the first author's University.

Translation process

The modified GEQ and YSEQ were translated into Norwegian using a process of back-translation (Brislin, 1970; Clark & Watson, 2019). The English versions were first translated into Norwegian by one professional translator and one experienced bilingual sport researcher. An expert group of native Norwegians consisting of one football coach (pro soccer license) and two sport researchers (PhD in Sport Science) independent of the study completed the translated versions while using a think-aloud procedure with the purpose of improving the clarity of the items (Jääskeläinen, 2010). The Norwegian versions of the modified GEQ and YSEQ were then back-translated to English by an independent bilingual researcher who was unfamiliar with the research topic, and a native English-speaking researcher evaluated the back translated versions against the original English versions. The content and meaning of each item were evaluated as entirely similar and therefore, no additional adjustments were required (please see appendix I for the translated scales in full). The translated questionnaires were completed by six appropriately aged Norwegian interdependent sport team athletes (aged 14–28 years) who confirmed the clarity of both the instructions and the individual items. Further, readability of the individual items was assessed using a Norwegian version (available at www.skriftlig.no/

liksres) of Björnsson's (1968) LIX readability index with the following interpretation: <30: very easy to read; 30–40: easy to read; 40–50: moderately difficult to read; 50–60: difficult to read; >60: very difficult to read. The translated YSEQ-items scored a median value of 22.5 (range 6 to 46), indicating that they were mostly "easy to read" (all but one item <40), and the translated GEQ-items scored a median value of 42 (range 20 to 48), indicating that they were at worst "moderately difficult to read".

Participants

Sample 1

This sample consisted of 333 adult athletes (age range 17–35; M_{age} = 18.7 years, *SD* = 2.6; 33% female) from different competitive teams playing football (n = 208), handball (n = 97), ice-hockey (n = 19), basketball (n = 2), and volleyball (n = 7). The athletes had participated in their sport for an average of 11.6 years (*SD* = 3.4) and on their current teams for 6.1 years (*SD* = 4.7).

Sample 2

This sample consisted of 303 youth athletes (age range 13–17; M_{age} = 15.0 years, *SD* = 1.48; 26% females) from age-restricted teams playing football (*n* = 120), handball (*n* = 59), ice-hockey (*n* = 122), and volleyball (*n* = 2). The athletes had participated in their sports for 8.7 years (*SD* = 2.73) and on their current teams for 5.1 years (*SD* = 3.56).

Procedure and measures

Sample 1. The participants were recruited through four universities in Norway with initial contact made via their coach/teacher. Participation was voluntary and required informed consent. Although the participants were recruited from educational institutions, they completed the questionnaire package in relation to their participation in competitive sport teams (outside of the universities). The pen-and-paper questionnaires were completed in person during class with a representative from the research group present and took approximately 15 minutes to complete. All participants completed the Norwegian version of the 18-item modified positively worded GEQ (Eys et al., 2007) that assessed four first-order dimensions of cohesion: ATG-S (five items; e.g., "I enjoy being part of the social activities of this team" to "Jeg liker å være med på sosiale aktiviteter sammen med laget"); ATG-T (four items; e.g., "I am happy with the amount of playing time I get" to "Jeg er fornøyd med den spilletiden jeg får i kamper"); GI-T (five items; e.g., "Our team have similar aspirations for the team's performance" to "Spillerne på laget vårt har like ambisjoner når det gjelder lagets resultater"); and GI-S (four items; e.g., "Members of our team communicate freely about each athlete's responsibilities during competition and practice" to "Spillerne på laget vårt snakker åpent om den enkelte spillers ansvar under kamp eller trening"). Participants responded to each item on a 9-point Likert scale anchored by 1 (strongly disagree) and 9 (strongly agree). Thus, higher scores reflected stronger perceptions of cohesion.

Sample 2. The participants were recruited from high schools across Norway with initial contact being made via their coach/teacher. Participation was voluntary and involved individual consent for those aged 15 years and older (*n* = 196), in addition to parental/guardian consent for those younger than 15 years of age (*n* = 107). The participants were asked to respond to questions based on their participation in age-restricted sport teams (outside of their schools). Similar to sample 1, the pen-and-paper questionnaires were completed in person during class with a representative from the research group present and took approximately 15 minutes to complete. All participants completed the Norwegian version of the YSEQ (Eys et al., 2009b) that assessed task (eight items) and social cohesion (eight items). Two translated item examples are: task cohesion, "We like the way we work together as a team" to "Vi liker måten vi jobber sammen på, som et lag" and social cohesion, "Some of my best friends are on this team" to "Noen av mine beste venner er på dette laget." Participants responded to each of the 16 items on a 9-point Likert scale anchored at 1 (strongly disagree) and 9 (strongly agree), with a higher score reflecting stronger perceptions of cohesion.

Statistical analyses

Mplus (Muthén & Muthén, 1998–2017) version 8.4 was used to estimate the measurement models with the robust full information maximum likelihood estimator (MLR), which provide standard errors and a chi-square test statistic that are robust to non-normality (Satorra & Bentler, 1994). Item-level missing data were accounted for by the full information MLR (Enders, 2010). Following recommendations in the literature (Marsh et al., 2013) and current practice in psychometric research (e.g., Garn & Webster, 2017; Myers et al., 2016) both Independent Clusters Model Confirmatory Factor Analysis (ICM-CFA) and Exploratory Structural Equation Modeling (ESEM) were used to examine the factor structure of the GEQ and YSEQ. A common observation in previous studies is very strong correlations between the first-order latent factors of the GEQ (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004).

Given that most previous studies have relied on ICM-CFA to examine the factor structure of the GEQ, the factor correlations are likely inflated due to the highly restrictive nature of the model specification (Marsh et al., 2014). ESEM have been found to mitigate some of these problems with CFA because it allows for the inclusion of cross-loadings between items and non-target factors. Instruments, such as the GEQ and the YSEQ, typically include cross-loadings that can be justified by substantive theory, item content, or simply represent another source of measurement error. As a consequence, the items are fallible indicators of the constructs and tend to have small residual associations with other constructs (Asparouhov & Muthén, 2009). As most items have multiple determinants, it is reasonable to assume non-zero cross-loadings in psychological measurement (Marsh et al., 2014). Further, simulation and empirical studies show that ignoring small cross-loadings and forcing them to be zero usually result in inflated factor correlations that undermine discriminant validity and lead to biased estimates in SEM with other variables (Marsh et al., 2013). We used oblique target rotation (Asparouhov & Muthén, 2009) in the ESEM models and all cross-loadings were specified to be close to zero but not exactly zero, and the target loadings were freely estimated. There are no established criteria regarding the magnitude of target factor loadings and cross-loadings in ESEM models; however, some guidelines have recently been provided in the literature (Morin et al., 2020). Target factor loadings greater than 0.500 were considered fully satisfactory and those below 0.300 indicated an inadequate indicator. Target loadings falling in between can be retained in a specific application but must be scrutinized in future studies. Cross-loadings below 0.300 were considered negligible, whereas cross-loadings larger than 0.300 were inspected further. Note that these are guidelines, not "golden rules," and should be interpreted in combination with other criteria to determine the adequacy of the model.

For the GEQ, based on previous research and the logic of the conceptual framework (Carron et al., 1985; Casey-Campbell & Martens, 2009; Dion, 2000), the *primary models* of interest were the first order four-factor model and two different second-order models (i.e., GI-ATG and task-social). However, due to the inconsistent findings in the literature regarding the factor structure of the GEQ (Leo et al., 2015; Li & Harmer, 1996; Schutz et al., 1994), we also examined multiple *alternative models* that have been examined in previous research. In total, we examined eight-factor structures of the GEQ using ICM-CFA and ESEM: a first order one-factor cohesion model (Model 1); a first order two-factor ATG and GI model (Model 2); a first order two-factor task and social cohesion model (Model 3); a first order four-factor model (Model 4); a second order one-factor cohesion model (Model 5); a second order two-factor ATG and GI model (Model 6); a second order two-factor task and social cohesion model (Model 7); and a bifactor model (Model 8). The hierarchical ESEM models were estimated using the ESEM-within-CFA approach developed by Morin and Asparouhov (2018). For the YSEQ, we compared the hypothesized first order two-factor task and social cohesion model using ICM-CFA and ESEM. Visual illustrations of the models are presented in appendix II.

The DIF as a function of sex was examined using a multiple indicator multiple causes (MIMIC) approach (Morin et al., 2016). In the MIMIC models, latent variables are regressed onto observed predictors and can be extended to examine DIF. Morin et al. (2016) defined this approach to DIF as "a form of measurement non-invariance that is characterized by direct relations between predictors and item responses over and above the effects of the predictors on the latent factors" (p. 184). The MIMIC approach provides a more parsimonious approach than multi-group measurement invariance testing, which suits the relatively small sample and uneven distribution of boys and girls in the current study. Following Morin et al. (2013), Morin et al. (2016), three models were estimated: (i) a null effect model, in which all paths from the predictor to the latent variables and item responses were constrained to zero; (ii) a factors-only model, where the paths from the predictor to the latent variables, but not the item responses, were freely estimated; (iii) a saturated model, where the paths from the predictor to the item responses, but not the latent factors, were freely estimated. A better model fit of the saturated model compared to the factors-only model indicates a presence of DIF, whereas improved model fit in the factors-only and saturated models compared to the null effects model indicates relations between the predictor and the ratings.

Given the oversensitivity of the chi-square test of exact fit to sample size and minor model misspecifications (Marsh et al., 2005), we relied on goodness-of-fit indices to assess model fit. More specifically, we relied on the Tucker Lewis index (TLI), comparative fit index (CFI), root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR) based on the following criteria: TLI and CFI > .90; RMSEA < .08, and SRMR < .08 (Marsh, 2007). For nested model comparisons (e.g., ICM-CFA vs. ESEM, MIMIC), a difference of less than 0.010 on the CFI and 0.015 on the RMSEA between two models can be considered evidence that they provide an equivalent fit to the data (Chen, 2007). The CFI was used as the main criterion because it is less sensitive to sample size and model complexity.

We used the fixed factor method to scale the latent factors in the ICM-CFA models. Composite reliability was computed according to McDonald's (1970) $\omega = (\Sigma|\lambda i|)^2/([\Sigma|\lambda i|^2]+ \Sigma\delta ii)$ using standardized parameter estimates from the ICM-CFA or ESEM models where λi are the factor loadings and δii are the error variances. McDonald's omega coefficient has the same practical interpretation as coefficient alpha but is a more flexible alternative for reliability estimation and does not rely on the tau-equivalence assumption (McNeish, 2018). Researchers have proposed various criteria for determining discriminant validity based on the magnitude of the latent variable correlations (Cheung & Wang, 2017; Kline, 2015). We used 0.700 (Cheung & Wang, 2017) as a lower bound and considered correlations larger than 0.700 as an indication of problems with discriminant validity.

Results

Sample 1. modified GEQ

Primary models

Inspection of skewness- and kurtosis-values revealed that all items fell within ± 1 (see Table 1). First, we compared model fit of the first-order four-factor model based on ICM-CFA (M4a) and ESEM (M4b). As seen in Table 2, the model fit of the ESEM model (χ^2 = 155.324, df = 87, p < .001; TLI = 0.943, CFI = .967; RMSEA = .049 [.036 – .061], SRMR = 0.028) was superior to the ICM-CFA model (χ^2 = 291.702, df = 129, p < .001; TLI = 0.908, CFI = .922; RMSEA = .062 [.052 – .071], SRMR = .049). The latent factor correlations of the ICM-CFA ranged from 0.656 to 0.911, whereas the latent factor correlations in the ESEM model ranged from 0.341 to 0.547. These results indicate problems with discriminant validity in the ICM-CFA model.

An inspection of the factor loadings of the first order four-factor ESEM model (see Table 3) indicated that two of the items (ATGT2: "I am pleased with my playing time" and GIS11: "members in our team would rather do something together as a team than be together with others") were weakly related to their target factors (ATGT2: standardized factor loading = −0.134; and GIS11: standardized factor loading = −0.020). Further, item ATGT2 had a stronger factor loading onto the ATGS factor (standardized factor loading = 0.329). Overall, the two items (i.e., AGT2, GIS11) performed relatively poorly in all inspected models.

There were also some relatively strong cross-loadings (>0.300) in the ESEM model (i.e., ATGS7, ATGS9 on ATGT; ATGT2 on ATGS; ATGT4 on GIT; GIT10 on ATGT; GIT18 on GIS) indicating a substantial overlap between individual items on one factor with a non-target factor. Composite reliability (ω) based on the standardized parameter estimates from the first order four-factor ICM-CFA/ESEM was 0.752/0.710 (ATGS), 0.678/0.567 (ATGT), 0.701/0.625 (GIS), and 0.716/0.674 (GIT).

Taken together, these results suggest that the first order four-factor ESEM model provides a better fit to the data than the ICM-CFA model. However, it also reveals some items with weak target factor loadings (i.e., items ATGT2 and GIS11) and, as previously noted, items with substantial cross-loadings on non-target factors.

As seen in Table 2, the model fit of the second-order task and social cohesion ESEM model (M7b; χ^2 = 153.571, df = 88, p < .001; TLI = 0.945, CFI = .969; RMSEA = .047 [.035 – .060],

Table 1. Descriptive statistics of the items of the geq (top part) and yseq (bottom part).

geq (sample 1)	m	sd	skewness	kurtosis	range	n
atgs1	7.01	1.87	−0.80	−0.15	1–9	332
atgt2	6.84	2.24	−0.95	−0.03	1–9	327
atgs3	6.17	2.11	−0.41	−0.55	1–9	327
atgt4	6.74	1.84	−0.59	−0.37	2–9	324
atgs5	6.36	2.52	−0.60	−0.93	1–9	325
atgt6	6.22	2.10	−0.57	−0.57	1–9	326
atgs7	5.32	2.24	−0.16	−0.83	1–9	321
atgt8	6.14	1.85	−0.53	−0.21	1–9	320
atgs9	6.07	2.22	−0.55	−0.60	1–9	323
git10	6.53	1.86	−0.65	−0.20	1–9	319
gis11	5.17	1.92	−0.08	−0.57	1–9	317
git12	6.12	1.98	−0.37	−0.63	1–9	320
gis13	5.27	2.07	−0.19	−0.84	1–9	325
git14	6.01	1.96	−0.30	−0.59	1–9	326
gis15	5.78	2.05	−0.34	−0.68	1–9	324
git16	5.80	2.02	−0.31	−0.59	1–9	327
gis17	5.92	1.88	−0.38	−0.35	1–9	321
git18	6.17	1.77	−0.38	−0.29	1–9	326
yseq (sample 2)	m	sd	skewness	kurtosis	range	n
task1	6.92	1.77	−0.63	−0.07	1–9	294
task3	5.68	1.92	−0.23	−0.59	1–9	298
task5	7.06	1.75	−1.04	0.89	1–9	298
task8	7.28	1.74	−0.93	0.21	1–9	296
task10	7.44	1.78	−1.41	1.68	1–9	300
task14	7.10	1.83	−0.90	0.10	1–9	296
task16	6.46	1.84	−0.52	−0.44	1–9	294
task18	7.10	1.64	−0.80	0.31	1–9	302
social2	6.31	2.19	−0.60	−0.46	1–9	299
social4	7.47	2.29	−1.57	1.43	1–9	298
social7	5.44	2.28	−0.06	−0.84	1–9	298
social9	6.95	2.12	−1.04	0.26	1–9	295
social11	6.66	2.34	−0.89	−0.15	1–9	291
social13	7.04	2.26	−1.24	0.84	1–9	300
social15	6.46	2.10	−0.75	−0.12	1–9	298
social17	6.73	2.24	−0.89	−0.11	1–9	301

atgs = attraction to group social, atgt = attraction to group task, gis = group integration social, git = group integration task.

Table 2. Goodness-of-fit statistics and information criteria for the models estimated on the full geq and yseq.

	χ^2	*df*	cfi	tli	rmsea	rmsea 90% ci	srmr
geq							
icm-cfa							
m1: one-factor icm-cfa	488.153*	135	0.831	0.809	0.089	[0.080, 0.097]	0.063
m2a: two-factor icm-cfa (atg/gi)	444.611*	134	0.851	0.830	0.083	[0.075, 0.092]	0.061
m3a: two-factor icm-cfa (task/social)	376.051*	134	0.884	0.868	0.074	[0.065, 0.083]	0.055
m4a: four-factor icm-cfa	291.702*	129	0.922	0.908	0.062	[0.052, 0.071]	0.049
m5a: h-cfa (cohesion)	323.403*	131	0.908	0.893	0.066	[0.057, 0.076]	0.053
m6a: h-cfa (atg/gi)			*inadmissible solution. factor correlation larger than 1.0.*				
m7a: h-cfa (task/social)	302.934*	130	0.917	0.903	0.063	[0.054, 0.073]	0.050
m8a: b-cfa[a]	295.590*	118	0.915	0.890	0.067	[0.058, 0.077]	0.053
esem							
m2b: two-factor esem (atg/gi)	355.551*	118	0.886	0.853	0.078	[0.069, 0.087]	0.046
m3b: two-factor esem (task/social)	355.551*	118	0.886	0.853	0.078	[0.069, 0.087]	0.046
m4b: four-factor esem	155.324*	87	0.967	0.943	0.049	[0.036, 0.061]	0.028
m5b: h-esem (cohesion)	153.458*	89	0.969	0.947	0.047	[0.034, 0.059]	0.028
m6b: h-esem (atg/gi)			*inadmissible solution. factor correlation larger than 1.0.*				
m7b: h-esem (task/social)	153.571*	88	0.969	0.945	0.047	[0.035, 0.060]	0.028
m8b: b-esem	173.867*	73	0.952	0.899	0.064	[0.052, 0.077]	0.024
yseq							
icm-cfa	246.402*	103	0.936	0.926	0.068	[0.057, 0.079]	0.053
esem	196.069*	89	0.952	0.936	0.063	[0.051, 0.075]	0.030

df = degrees of freedom; cfi = comparative fit index; tli = tucker-lewis index; rmsea = root-mean-square error of approximation; ci = confidence interval; gi = group integration; atg = attraction to group; icm-cfa = independent clusters model confirmatory factor analysis; h-cfa = hierarchical confirmatory factor analysis; b-cfa = bifactor confirmatory factor analysis; esem = exploratory structural equation modeling; h-esem = hierarchical exploratory structural equation; modeling, b-esem = bifactor exploratory structural equation modeling. esem models were conducted with target oblique rotation. *all χ^2 values are significant (p < .001). [a]This model resulted in a negative residual variance estimate for item atgt8, which was fixed to a value of .1 to achieve an interpretable solution.

SRMR = 0.028) was superior to the second-order ICM-CFA (M7a; χ^2 = 302.934, df = 130, p < .001; TLI = 0.903, CFI = .917; RMSEA = .063 [.054 – .073], SRMR = .050). The factor correlation between the second order latent factors was .875 in both the ESEM and ICM-CFA, which indicates problems with discriminant validity. Again, the ESEM model provided a better fit to the data; however, the same issues (e.g., cross-loadings and weak factor loadings) with the factor structure observed for the first-order ESEM applies to the second-order ESEM.

Alternative models

Several alternative factor structures for the GEQ were also examined (Table 2). The first-order one-factor and two-factor models (i.e., M1, M2, and M3) all had relatively poor model fit. The hierarchical (i.e., second order) and bifactor ESEM models provided a better fit to the data than the ICM-CFA models; thus, we focus on the results from the ESEM models. The second-order one-factor cohesion model (M5b) provided a good fit to the data (χ^2 = 153.458, df = 89, p < .001; TLI = 0.947, CFI = .969; RMSEA = .047 [.034 – .059], SRMR = 0.028), with Composite reliability (ω) for the global cohesion factor of 0.703. The standardized second-order factor loadings ranged from 0.464 to 0.833. The second order two-factor ATG and GI model (M6b) yielded an inadmissible solution; the second order latent factor correlation was larger than 1.0. The bifactor ESEM model (M8b) provided an adequate fit to the data (χ^2 = 173.867, df = 73, p < .001; TLI = 0.899, CFI = .952; RMSEA = .064 [.052 – .077], SRMR = 0.024). The factor loading pattern from the bifactor ESEM shows that all items, except ATGT2 and GIS11, loaded relatively strongly onto the global cohesion factor (standardized factor loadings > .50). However, the specific factors were relatively poorly defined (Table 4), with at best one significant factor loading for the specific first-order factors.

Sample 2. modified YSEQ

As seen in Table 2, the model fit of the two-factor ESEM of the translated YSEQ (χ^2 = 196.069, df = 89, p < .001; TLI = 0.936, CFI = .952; RMSEA = .063 [.051 – .075], and SRMR = .030) was superior to the two-factor ICM-CFA (χ^2 = 246.402, df = 103, p < .001; TLI = 0.926, CFI = .936; RMSEA = .068 [.057 – .079], and SRMR = .053). The standardized target factor loadings in the ESEM model ranged from 0.555 to 0.918 and all items loaded on their respective latent constructs (Table 5). The cross-loadings were relatively weak (< 0.20) in the ESEM model. Although the ESEM models had better model fit than the ICM-CFA model, the factor loading pattern and latent factor correlations (r_{ESEM} = 0.539 vs. $r_{\text{ICM-CFA}}$ = 0.558) were similar between the two models. Composite reliability (ω) based on the standardized

Table 3. Standardized factor loadings and item uniquenesses from the first-order four-factor icm-cfa and esem models of the geq.

	icm-cfa					esem				
	atgs	atgt	gis	git	δ	atgs	atgt	gis	git	δ
atgs1	**0.736***				0.458*	**0.701***	−0.072	0.072	0.104	0.390*
atgs3	**0.665***				0.557*	**0.702***	−0.187	−0.058	0.265*	0.865*
atgs5	**0.606***				0.633*	**0.703***	−0.020	0.141	−0.233*	0.367*
atgs7	**0.798***				0.362*	**0.553***	0.370*	0.256*	−0.176*	0.464*
atgs9	**0.799***				0.362*	**0.555***	0.328*	0.139	−0.036	0.536*
atgt2		**0.243***			0.941*	0.329*	**−0.134**	−0.105	0.160	0.475*
atgt4		**0.697***			0.514*	0.057	**0.284***	−0.020	0.559*	0.280*
atgt6		**0.727***			0.472*	0.280*	**0.472***	−0.044	0.201*	0.436*
atgt8		**0.777***			0.397*	0.284*	**0.388***	−0.002	0.297*	0.354*
gis11			**0.116***		0.987*	0.061	0.016	**−0.020**	0.089	0.306*
gis13			**0.697***		0.515*	−0.010	0.037	**0.678***	0.099	0.983*
gis15			**0.833***		0.306*	0.198*	−0.084	**0.618***	0.186*	0.578*
gis17			**0.831***		0.309*	0.208*	−0.095	**0.666***	0.138	0.463*
git10				**0.821***	0.326*	0.097	0.410*	0.111	**0.460***	0.476*
git12				**0.642***	0.588*	0.130	0.190*	−0.007	**0.476***	0.334*
git14				**0.659***	0.565*	−0.210*	0.238*	0.253*	**0.573***	0.587*
git16				**0.628***	0.605*	0.270*	0.004	0.128	**0.370***	0.300*
git18				**0.630***	0.603*	0.010	−0.048	0.379*	**0.442***	0.545*

atgs = attraction to group social, atgt = attraction to group task, gis = group integration social, git = group integration task, δ = item uniquenesses. *p < .05

Table 4. Standardized factor loadings and item uniquenesses from the bifactor esem model of geq.

	atgs	atgt	gis	git	cohesion	δ
atgs1	**0.399**	−0.022	0.065	0.007	0.654*	0.408*
atgs3	**0.570***	0.100	0.036	0.132	0.573*	0.318
atgs5	**0.341**	−0.142	0.081	−0.220	0.522*	0.536*
atgs7	**0.049**	−0.117	0.032	−0.201	0.807*	0.292*
atgs9	**0.086**	−0.091	−0.053	−0.099	0.794*	0.341*
atgt2	0.277*	**0.031**	−0.023	0.103	0.204*	0.87*
atgt4	−0.017	**0.418**	−0.025	0.208	0.597*	0.425*
atgt6	−0.050	**0.249**	−0.122	−0.053	0.677*	0.46*
atgt8	−0.015	**0.180**	−0.101	0.069	0.717*	0.439*
gis11	0.066	0.180	**0.067**	−0.053	0.094	0.947*
gis13	−0.080	−0.047	**0.445***	0.020	0.574*	0.464*
gis15	0.129	0.043	**0.525**	0.033	0.665*	0.262
gis17	0.083	−0.178	**0.411**	0.127	0.685*	0.307*
git10	−0.132	0.203	−0.047	**0.200**	0.771*	0.304*
git12	0.013	0.166	−0.073	**0.269***	0.562*	0.578*
git14	−0.176	0.305*	0.151	**0.280**	0.553*	0.469*
git16	0.143	−0.121	−0.037	**0.384**	0.590*	0.468
git18	0.006	−0.084	0.182	**0.419**	0.555*	0.476*

atgs = attraction to group social, atgt = attraction to group task, gis = group integration social, git = group integration task, cohesion = global cohesion factor, δ = item uniquenesses. *p < .05

Table 5. Standardized factor loadings and item uniquenesses from the first-order two-factor icm-cfa and esem models of the yseq.

	icm-cfa			esem		
	task	social	δ	task	social	δ
task1	0.679*		0.539*	0.769*	−0.142*	0.506*
task3	0.635*		0.596*	0.592*	0.070	0.600*
task5	0.808*		0.346*	0.813*	−0.010	0.347*
task8	0.793*		0.371*	0.727*	0.114*	0.370*
task10	0.664*		0.559*	0.623*	0.062	0.566*
task14	0.672*		0.548*	0.739*	−0.095*	0.521*
task16	0.717*		0.485*	0.619*	0.167*	0.477*
task18	0.852*		0.273*	0.883*	−0.047	0.263*
social2		0.790*	0.376*	0.028	0.769*	0.384*
social4		0.726*	0.472*	−0.048	0.750*	0.474*
social7		0.755*	0.430*	0.159*	0.661*	0.425*
social9		0.839*	0.296*	−0.113*	0.913*	0.266*
social11		0.808*	0.347*	−0.033	0.826*	0.346*
social13		0.593*	0.648*	0.067	0.555*	0.648*
social15		0.852*	0.275*	0.136*	0.771*	0.275*
social17		0.874*	0.236*	−0.063	0.918*	0.216*

δ = item uniquenesses. *p < .05

parameter estimates from the ICM-CFA and ESEM models was 0.758 and 0.760 for social cohesion and 0.802 and 0.802 for task cohesion, respectively.

DIF as a function of sex in the GEQ and YSEQ

We examined DIF in the three best-fitting models of the GEQ (Model 4b, Model 5b, and Model 7b) and the ESEM of YSEQ (Table 6). Regarding the GEQ, none of the model fit comparisons indicated that the saturated models had superior model fit compared to the factors-only models (ΔCFI < 0.010 and ΔRMSEA < 0.015). Model fit comparisons of the factors-only and saturated models indicated a difference between males and females in Model 4b (ΔCFI = 0.011). Examination of the regression coefficients showed that males reported lower group integration-task (GIT) compared to females (β = −0.20, p = .005); however, the magnitude of the regression coefficient was relatively small.

The model fit comparisons of the YSEQ models did not indicate DIF as a function of sex or differences between boys and girls on the latent variables (ΔCFI < 0.010 and ΔRMSEA < 0.015). Taken together, these results do not suggest DIF in the item responses of the GEQ or YSEQ as a function of sex.

Table 6. Model fit statistics of the mimic models of the geq and yseq.

	χ^2	*df*	cfi	tli	rmsea	rmsea 90% ci	srmr
geq							
m4b: four-factor esem							
null effects model	199.030*	105	0.955	0.927	0.052	[0.041, 0.063]	0.034
factors-only model	172.128*	101	0.966	0.943	0.046	[0.034, 0.058]	0.028
saturated model	162.906*	87	0.964	0.929	0.052	[0.039, 0.064]	0.027
m5b: h-esem (cohesion)							
null effects model	192.559*	107	0.959	0.935	0.050	[0.038, 0.061]	0.035
factors-only model	194.026*	106	0.958	0.932	0.050	[0.039, 0.062]	0.032
saturated model	160.647*	89	0.966	0.934	0.050	[0.037, 0.062]	0.027
m7b: h-esem (task/social)							
null effects model	192.415*	106	0.959	0.934	0.050	[0.039, 0.061]	0.034
factors-only model	199.159*	108	0.957	0.931	0.051	[0.040, 0.062]	0.035
saturated model	160.745*	88	0.965	0.933	0.050	[0.038, 0.063]	0.027
yseq (two-factor esem)							
null effects model	223.472*	105	0.948	0.932	0.062	[0.050, 0.073]	0.033
factors-only model	223.001*	103	0.947	0.930	0.063	[0.051, 0.074]	0.033
saturated model	192.133*	89	0.955	0.931	0.062	[0.050, 0.075]	0.028

df = degrees of freedom; cfi = comparative fit index; tli = tucker-lewis index; rmsea = root-mean-square error of approximation; ci = confidence interval; srmr = standardized root mean squared residual; esem = exploratory structural equation modeling; h-esem = hierarchical exploratory structural equation; modeling, esem models were conducted with target oblique rotation. *all χ^2 values are significant (p < .001).

Discussion

Group cohesion is arguably one of the most important constructs for sport teams across a range of competition levels and age groups (Eys et al., 2019). Cohesion has also been extensively investigated internationally, yet there is a dearth of research in Scandinavia. One possible explanation could be the lack of suitable questionnaires. Therefore, this study translated, adapted, and carefully examined the factor structure of the GEQ and YSEQ questionnaire in a Norwegian context. First, the established English versions were translated using a protocol advocated for transcultural validation of psychometric instruments (Brislin, 1970; Clark & Watson, 2019). During forward and back translation and subsequent discussion of the constructs and items with experts, there were no major disagreements in terms of translation and content. Readability tests also indicated that items within the final instruments were generally easy (YSEQ) and at worst moderately difficult (GEQ) to read, indicating that the translated scales have both high face and content validity, and should be possible to understand for participants. Second, we implemented a robust–methodological approach involving ICM-CFA and ESEM to examine the factor structure of the two questionnaires and examined DIF as a function of sex.

The findings suggest that the original four-factor model (Carron et al., 1998, 1985) yielded acceptable fit indices when estimated as an ICM-CFA and ESEM model (model M4a and M4b, respectively), in line with previous research (Carron et al., 1985; Casey-Campbell & Martens, 2009; Dion, 2000). However, the model fit of the ESEM was superior to the ICM-CFA, like what has been found in other multidimensional measures in sport psychology (Perry et al., 2015). The relatively strong latent factor correlations in the ICM-CFA model and relatively strong cross-loadings in the ESEM model indicates a substantial overlap between the four factors. This finding mirrors previous research with the GEQ that had reported poor discriminant validity (Ntoumanis & Aggelonidis, 2004). In the traditional ICM-CFA model, all cross-loadings are constrained to zero. Both theoretically and empirically the dimensions – and thus, the items – pertaining to cohesion are interrelated. Forcing the cross-loadings to be zero could result in high factor correlations, poor discriminant validity, and biased estimates in multivariate SEM-models (Marsh et al., 2014). In the current study, the latent factor correlations in the ESEM model were weaker compared to the ICM-CFA model (ESEM: 0.341– 0.547; ICM-CFA: 0.656 to 0.911). It has been argued that if the fit and parameter estimates (e.g., latent factor correlations) of the ESEM model is acceptable and better than the ICM-CFA model, the ESEM model should be retained (Marsh et al., 2014). Due to the strict requirement of the ICM-CFA model, all potential cross-loadings will contribute to model misspecification (Marsh et al., 2014). Thus, our findings suggest that an ESEM model is more appropriate for a four-factor GEQ model than an ICM-CFA in the present Norwegian version.

Consistent with the propositions of Carron et al. (1998) and Carron, Brawley et al. (2002), and the recommendation that task and social dimensions should be given special emphasis (Carron et al., 1985; Casey-Campbell & Martens, 2009; Dion, 2000), we also investigated a second-order two-factor models of task and social cohesion. Like the four-factor models, the second order two-factor models of task and social cohesion yielded acceptable fit indices for both the ICM-CFA and ESEM solution (model M7a and M7b, respectively). Similar findings have been reported previously (Leeson & Fletcher, 2005; Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004). Again, the second-order ESEM model provided a superior fit compared to the ICM-CFA model. Nevertheless, the same issues (e.g., cross-loadings) with the factor structure observed for the first-order ESEM apply to the second-order ESEM. Acceptable psychometric properties of a second order two-factor task and social cohesion model may be considered important as Casey-Campbell and Martens (2009) stated that researchers "who ignore the potential impact of both social and task concepts on the group cohesiveness construct risk generating yet more confusion in the literature" (p. 231). In addition, using a two-rather than four-factor model could make it more feasible to explore more complex structural models using cohesion as an antecedent, mediator, or outcome in future research.

Despite acceptable model fit indices overall, some items had weak factor loadings for their intended latent factor. The weak factor loading on the corresponding factor for the ATGT2 item is not readily explainable, but could be explained by the Norwegians society's egalitarian and less individualistic orientation compared to the origin country of the scale (Canada) and other European countries (e.g., Denmark, Sweden, UK)[1] – and by extension, these athletes (i.e., "the team first"). Thus, satisfaction with their own playing time may not be perceived as important when evaluating their attraction to the team's task. However, the reason for the weak factor loading may not necessarily be cultural. Considering that the participants in the present study had played for their current team for a relatively long period of time (mean of 6.1 years), it is possible that the players had accepted their roles within the team. The weak factor loading for the GIS11 item on the intended factor may be considered slightly unexpected. One possible explanation is the current wording in the Norwegian version, because respondents were asked to evaluate their teammates' preferences, which can be difficult and superficial, and not necessarily reflective of the social bonding within the team. Both items should be investigated further and perhaps reworded so they may fit into a Norwegian sport context more accurately.

Contrary to other findings (Ntoumanis & Aggelonidis, 2004), this study did not support an acceptable model fit for a second order two-factor AGT – GI model in the ICM-CFA framework. Moreover, the same model tested in the ESEM framework yielded an inadmissible solution because of a factor correlation larger than 1.0. Thus, our findings do not support a second-order cohesion model with ATG and GI as second-order factors.

In addition to the abovementioned models, other alternative models of the GEQ were tested. The second-order one-factor ESEM model (M5b) yielded acceptable fit indices, which is in line with previous research (Li & Harmer, 1996; Ntoumanis & Aggelonidis, 2004). Similarly, the model fit of the bifactor ESEM (M8b) could also be considered adequate. Except for two items (i.e., ATGT2, GIS11), all items loaded strongly onto the global cohesion factor, which suggests that the global factor explains a substantial amount of the covariance among the items. These findings bear resemblance to Schutz et al. (1994) and suggest that the GEQ captures a strong common cohesion factor but there is multi-dimensionality caused by clusters of items representing subdomains (Reise, 2012). However, the results in the current study also suggest that the specific factors were poorly defined as indicated by relatively weak target factors. Indeed, the bifactor solution resulted in at best one significant factor loading for each specific construct. Although not completely consistent with the original theoretical propositions, our results combined with previous findings (e.g., Ntoumanis & Aggelonidis, 2004; Schutz et al., 1994) suggest that the second-order one-factor cohesion model (M5b) provides a parsimonious approach to account for the covariance structure with a single cohesion factor. More research is warranted to be able to determine the implications of these findings for the construct of cohesion and how it should be measured. The identification of different, but possibly equally plausible, factor structures within the complex cohesion framework makes it difficult to give universal recommendations, or to favor one-factor structure over another, purely based on fit-indices and amount of explainable variance. As recently highlighted in research on self-determination theory (Howard et al., 2020), selection of the ideal 'scoring method' will depend on theoretical alignment between the hypotheses and the method.

For the investigated ESEM models of the modified GEQ, substantial cross-loadings (>0.300) were evident for some of the items. Interestingly, all cross-loadings followed a theoretically and conceptually logical pattern; items loaded on factors that shared common features such as task or social and ATG or GI. Cross-loadings that align with theory and logic may be considered less problematic (Morin et al., 2020). For example,

[1]¥ for a nation-by-nation comparison, please see https://www.hofstede-insights.com/product/compare-countries/

respondents may not distinguish between ATG and GI when reporting on a task-related item. Thus, the cross-loadings seemed to reflect construct-relevant associations between items and non-target factors. However, the magnitude of two cross-loadings (ATGT4 on GIT = 0.559; GIT10 on ATGT = 0.410) were substantial, and item ATGT4 had a considerably higher cross-loading than target factor loading (0.284). The question remains whether these items are ascribed to the wrong factor and they should be examined further in future psychometric studies. One approach that could be adopted in future studies on the GEQ is Bayesian CFA (e.g., B. Muthén & Asparouhov, 2012; see also Stenling et al., 2015) that can incorporate cross-loadings as well as residual covariances among all items. Exploring the GEQ using various prior specifications on the residual variances and covariances might provide additional insight into the factor structure of the GEQ.

The composite reliability of the first order four-factor ESEM yielded somewhat low estimates for several of the subscales. Researchers have reported variable reliability estimates of the GEQ subscales. The original Cronbach's alpha estimates of the four GEQ subscales ranged from .64 to .76 (Carron et al., 1998). Other studies have reported low to moderate Cronbach's alpha estimates (.44 to .68; Westre & Weiss, 1991), some have reported low to high estimates across the four subscales (.49 to .91; Whitton & Fletcher, 2014), whereas others have reported estimates higher than .90 for all subscales (Ntoumanis & Aggelonidis, 2004). Eys et al. (2007) showed that the combination of positively and negatively phrased items may attenuate reliability estimates of the GEQ subscales, which may explain the low estimates observed in some studies. The variable reliability estimates may also be a consequence of the dynamic and multidimensional nature of the cohesion construct; not all dimensions may be salient for a group at a specific time point or across different types of groups (Carron et al., 1998; Whitton & Fletcher, 2014). Taken together, the findings from this study and previous literature highlight the importance of using latent variable models with the GEQ that corrects for unreliability (Marsh et al., 2013).

The YSEQ is based on a conceptual and proposed two-factor model (task and social cohesion) and was developed for participants between 13 and 17 years of age (Eys et al., 2009b). Both the ICM-CFA model and the ESEM-model of the translated YSEQ instrument yielded acceptable model fit and had similar factor loadings and latent factor correlations. Based on the CFI values, the ESEM model seemed to provide a slightly better fit compared to the more restricted ICM-CFA model; however, there were no major differences between the two models. The results from the psychometric analyses bear resemblance to Eys et al. (2009b) and other translated versions (Benson et al., 2016; Eshghi et al., 2015). Also, the composite reliability of both factors was considered acceptable. In sum, the 16-item Norwegian version of the YSEQ in the present sample is deemed to be appropriate both through ICM-CFA modeling and ESEM modeling.

Finally, examination of DIF as a function of sex did not indicate any differences between sexes. These results are in line with previous adaptations and translations of the GEQ to Greek (Ntoumanis & Aggelonidis, 2004) and Spanish (Leo et al., 2015) and suggest that the male and female athletes perceived the items of the GEQ and YSEQ in a similar way. Thus, these findings may indicate that sex differences such as the one in the cohesion–performance relationship reported in Carron, Colman et al. (2002) may be conceptual differences rather than caused by measurement. However, although the present findings provide initial support for measurement invariance, future studies with larger samples and a more even distribution of males and females are warranted to be able to conduct multi-group tests of measurement invariance.

Limitations

One major limitation of this study is that only one sample and one measurement point were used to examine each adapted scale. We were, therefore, not able to perform confirmatory analyses on the most appropriate models, to test criterion validity, or to investigate measurement invariance across multiple samples or across time. Thus, the proposed factor structures will need to be scrutinized in future psychometric studies. Further investigation of each scale is recommended and should include multiple samples and/or multiple measurement points to further examine the psychometric properties of the GEQ and YSEQ.

Moreover, the participants in the present study were recruited from schools and universities in Norway and represent an unknown number of sport teams. Thus, we were not able to account for potential group-level variance or examine cohesion at the group level (Whitton & Fletcher, 2014). Even though the current study investigated DIF as a function of sex, other subject or context characteristics may be similarly worthy of examination (e.g., individual vs team-sports, level of performance).

Conclusions

Findings from the present study provide initial support for the structural validity of the Norwegian version of the GEQ and suggest that ESEM provides a better representation of the data than the ICM-CFA. The results indicated acceptable fit indices for both a four-factor ESEM model and a second-order two-factor task and social cohesion ESEM model, in addition to a second-order one-factor cohesion model. However, the substantial cross-loadings and high latent factor correlations are issues in need of attention in future studies. Also, the items with low factor loadings on their intended latent factors should be targeted for reassessment in future psychometric studies.

The study also provides initial support for the structural validity and reliability of the Norwegian version of YSEQ. No major differences in model fit and parameter estimates were found between the ICM-CFA and ESEM, thus the ICM-CFA could be retained for subsequent analyses based on parsimony. However, we agree with the recommendations that researchers should routinely apply both ICM-CFA and ESEM to their measurement models (Marsh et al., 2014). In general – not only regarding the GEQ and YSEQ – researchers interested in interrelated constructs (and items) should consider testing their latent variable models with ESEM to examine, and if necessary account for, residual associations between items and constructs. Finally, examination of DIF as a function of sex did not indicate any differences between sexes in either of the two scales (GEQ and YSEQ).

Declaration of interests

The authors have no conflicts of interest to disclose.

ORCID

Luc J. Martin http://orcid.org/0000-0001-5336-6119
Andreas Stenling http://orcid.org/0000-0002-0834-1040

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Benson, A. J., Šiška, P., Eys, M., Priklerová, S., & Slepička, P. (2016). A prospective multilevel examination of the relationship between cohesion and team performance in elite youth sport. *Psychology of Sport & Exercise*, *27*(1), 39–46. https://doi.org/10.1016/j.psychsport.2016.07.009

Björnsson, C. H. (1968). *Läsbarhet. [Readability]*. Liber.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small Group Research*, *31*(1), 71–88. https://doi.org/10.1177/104649640003100104

Carron, A. V., & Brawley, L. R. (2000). Cohesion: Conceptual and measurement issues. *Small Group Research*, *31*(1), 89–106. https://doi.org/10.1177/104649640003100105

Carron, A. V., Brawley, L. R., & Widmeyer, W. N. (1998). The measurement of cohesiveness in sport groups. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 213–226). Fitness Information Technology.

Carron, A. V., Brawley, L. R., & Widmeyer, W. N. (2002). *The Group Environment Questionnaire test manual*. Fitness Information Technology.

Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance in sport: A meta analysis. *Journal of Sport & Exercise Psychology*, *24*(2), 168–188. https://doi.org/10.1123/jsep.24.2.168

Carron, A. V., Eys, M. A., & Burke, S. M. (2007). Team cohesion. In S. Jowett & D. Lavallee (Eds.), *Social psychology in sport* (pp. 91–102). Human Kinetics.

Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. *Journal of Sport & Exercise Psychology*, *7*(3), 244–266. https://doi.org/10.1123/jsp.7.3.244

Casey-Campbell, M., & Martens, M. L. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, *11*(2), 223–246. https://doi.org/10.1111/j.1468-2370.2008.00239.x

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Wang, C. (2017). Current approaches for assessing convergent and discriminant validity with SEM: Issues and solutions. In *Academy of Management Proceedings* (Vol. 2017, No. 1, p. 12706). Academy of Management, Briacliff Manor, NY 10510. https://doi.org/10.5465/AMBPP.2017.12706abstract

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/dx.doi.org/10.1037/pas0000626

Cota, A. A., Evans, C. R., Dion, K. L., Kilik, L., & Longman, R. S. (1995). The structure of group cohesion. *Personality & Social Psychology Bulletin*, *21*(6), 572–580. https://doi.org/10.1177/0146167295216003

De Backer, M., Boen, F., De Cuyper, B., Høigaard, R., & Vande Broek, G. (2015). A team fares well with a fair coach: Predictors of social loafing in interactive female sport teams. *Scandinavian Journal of Medicine & Science in Sports*, *25*(6), 897–908. https://doi.org/10.1111/sms.12303

Dion, K. L. (2000). Group cohesion: From" field of forces" to multidimensional construct. *Group Dynamics: Theory, Research, & Practice*, *4*(1), 7–26. https://doi.org/10.1037/1089-2699.4.1.7

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Erikstad, M. K., Martin, L. J., Haugen, T., & Høigaard, R. (2018). Group cohesion, needs satisfaction, and self-regulated learning: A one-year prospective study of elite youth soccer players' perceptions of their club team. *Psychology of Sport & Exercise*, *39*(1), 171–178. https://doi.org/10.1016/j.psychsport.2018.08.013

Eshghi, M. A., Kordi, R., Memari, A. H., Ghaziasgar, A., Mansournia, M. A., & Zamani Sani, S. H. (2015). Validity and reliability of Farsi version of youth sport environment questionnaire. *Journal of Sports Medicine*, 2015, 5. Article ID 985283. https://doi.org/10.1155/2015/985283 .

Eys, M., Bruner, M. W., & Martin, L. J. (2019). The dynamic group environment in sport and exercise. *Psychology of Sport & Exercise*, *42*(1), 40–47. https://doi.org/10.1016/j.psychsport.2018.11.001

Eys, M., Loughead, T., Bray, S. R., & Carron, A. V. (2009b). Development of a cohesion questionnaire for youth: The Youth Sport Environment Questionnaire. *Journal of Sport & Exercise Psychology*, *31*(3), 390–408. https://doi.org/10.1123/jsep.31.3.390

Eys, M. A., & Brawley, L. R. (2018). Reflections on cohesion research with sport and exercise groups. *Social & Personality Psychology Compass*, *12* (4), e12379. Article. https://doi.org/10.1111/spc3.12379 .

Eys, M. A., Carron, A. V., Bray, S. R., & Brawley, L. R. (2007). Item wording and internal consistency of a measure of cohesion: The Group Environment Questionnaire. *Journal of Sport & Exercise Psychology*, *29*(3), 395–402. https://doi.org/10.1123/jsep.29.3.395

Eys, M. A., Evans, B., & Benson, A. (2020). *Group dynamics in sport* (5th ed.). Fit Publishing.

Eys, M. A., Hardy, J., Carron, A. V., & Beauchamp, M. R. (2003). The relationship between task cohesion and competitive state anxiety. *Journal of Sport & Exercise Psychology*, *25*(1), 66–76. https://doi.org/10.1123/jsep.25.1.66

Eys, M. A., Loughead, T. M., Bray, S. R., & Carron, A. V. (2009a). Perceptions of cohesion by youth sport participants. *The Sport Psychologist*, *23*(3), 330–345. https://doi.org/10.1123/tsp.23.3.330

Garn, A. C., & Webster, E. K. (2017). Reexamining the factor structure of the test of gross motor development – Second edition: Application of exploratory structural equation modeling. *Measurement in Physical Education & Exercise Science*, *22*(3), 200–212. https://doi.org/10.1080/1091367X.2017.1413373

Granito, V. J., Jr, & Rainey, D. W. (1988). Differences in cohesion between high school and college football teams and starters and nonstarters. *Perceptual & Motor Skills*, *66* (2), 471–477. https://doi.org/10.2466/pms.1988.66.2.471

Heuzé, J. P., & Fontayne, P. (2002). Questionnaire sur l'Ambiance du Groupe: A French-language instrument for measuring group cohesion. *Journal of Sport & Exercise Psychology*, *24*(1), 42–67. https://doi.org/10.1123/jsep.24.1.42

Høigaard, R., Säfvenbom, R., & Tønnessen, F. E. (2006). The relationship between group cohesion, group norms, and perceived social loafing in soccer teams. *Small Group Research*, *37*(3), 217–232. https://doi.org/10.1177/1046496406287311

Høigaard, R., Tofteland, I., & Ommundsen, Y. (2006). The effect of team cohesion on social loafing in relay teams. *International Journal of Applied Sport Sciences*, *18*(1), 59–73. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1062.2113&rep=rep1&type=pdf

Howard, J. L., Gagné, M., Van Den Broeck, A., Guay, F., Chatzisarantis, N., Ntoumanis, N., & Pelletier, L. G. (2020). A review and empirical comparison of motivation scoring methods: An application to self-determination theory. *Motivation & Emotion*, *44*(4), 534–548. https://doi.org/10.1007/s11031-020-09831-9

Iturbide, L. M., Elosua, P., & Yanes, F. (2010). Medida de la cohesión en equipos deportivos. Adaptación al español del Group Environment Questionnaire (GEQ). *Psicothema*, *22* (3), 482–488. https://www.redalyc.org/pdf/727/72714400020.pdf

Jääskeläinen, R. (2010). Think-aloud protocol. In Y. Gambier & L. Van Doorslaer (Eds.), *Handbook of translation studies* (pp. 371–374). John Benjamins Publishing.

Junior, J. R. A. D. N., Granja, C. T. L., Fortes, L. D. S., Freire, G. L. M., De Oliveira, D. V., & Peixoto, E. M. (2018). Cross-cultural adaptation and psychometric properties of the Portuguese version of the Youth Sport Environment Questionnaire (P-YSEQ). *Journal of Physical Education & Sport*, *18*(3), 1606–1614. https://doi.org/10.7752/jpes.2018.03236

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Leeson, H., & Fletcher, R. B. (2005). Longitudinal stability of the Group Environment Questionnaire with elite female athletes. *Group Dynamics: Theory, Research, and Practice*, *9*(3), 147–160. https://doi.org/10.1037/1089-2699.9.3.147

Leo, F. M., González-Ponce, I., Sánchez-Oliva, D., Pulido, J. J., & García-Calvo, T. (2015). Adaptation and validation in Spanish of the Group Environment Questionnaire (GEQ) with professional football players. *Psicothema*, *27*(3), 261–268. https://doi.org/10.7334/psicothema2014.247

Li, F., & Harmer, P. (1996). Confirmatory factor analysis of the Group Environment Questionnaire with an intercollegiate sample. *Journal of Sport & Exercise Psychology*, *18*(1), 49–63. https://doi.org/10.1123/jsep.18.1.49

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774–798). Wiley.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics. A festschrift for Roderick P. McDonald* (pp. 275–340). Erlbaum.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257–284. https://doi.org/10.1037/a0032773

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Martin, L. J., Carron, A. V., Eys, M. A., & Loughead, T. (2013). Validation of the child sport cohesion questionnaire. *Measurement in Physical Education & Exercise Science*, *17* (2), 105–119. https://doi.org/10.1080/1091367X.2013.761023

Martin, L. J., Carron, A. V., Eys, M. A., & Loughead, T. M. (2012). Development of a cohesion inventory for children's sport teams. *Group Dynamics: Theory, Research, & Practice*, *16*(1), 68–79. https://doi.org/10.1037/a0024691

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical & Statistical Psychology*, *23*(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Morin, A. J. S., & Asparouhov, T. (2018). *Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA*. Substantive Methodological Synergy Research Laboratory.

Morin, A. J. S., Maïano, C., White, R. L., Owen, K. B., Tracey, D., Mascret, N., & Lonsdale, C. (2016). English validation of the short form of the Physical Self-Inventory (PSI-S). *Psychology of Sport & Exercise*, *27*(1), 180–194. https://doi.org/10.1016/j.psychsport.2016.08.016

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395e436). Information Age Publishing, Inc.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (pp. 1044–1073). Wiley.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide. Eighth Edition*. Muthén & Muthén.

Myers, N. D., Park, S. E., Lefevor, G. T., Dietz, S., Prilleltensky, I., & Prado, G. J. (2016). Measuring multidimensional subjective well-being with the I COPPE scale in a Hispanic sample. *Measurement in Physical Education & Exercise Science*, *20*(4), 230–243. https://doi.org/10.1080/1091367X.2015.1111220

Ntoumanis, N., & Aggelonidis, Y. (2004). A psychometric evaluation of the Group Environment Questionnaire in a sample of elite and regional level Greek volleyball players. *European Physical Education Review*, *10*(3), 261–278. https://doi.org/10.1177/1356336X04047126

Ohlert, J. (2012). Kohäsionsfragebogen für Individual-und Teamsport-Leistungssport (KIT-L): A German-language instrument for measuring group cohesion in individual and team sports. *International Journal of Sport & Exercise Psychology*, *10*(1), 39–51. https://doi.org/10.1080/1612197X.2012.645129

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education & Exercise Science*, *19*(1), 12–21. https://doi.org/10.1080/1091367X.2014.952370

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Satorra, A., & Bentler, E. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.

Schutz, R. W., Eom, H. J., Smoll, F. L., & Smith, R. E. (1994). Examination of the factorial validity of the Group Environment Questionnaire. *Research Quarterly for Exercise & Sport*, *65*(3), 226–236. https://doi.org/10.1080/02701367.1994.10607623

Steca, P., Pala, A. N., Greco, A., Monzani, D., & D'Addario, M. (2013). A psychometric evaluation of the Group Environment Questionnaire in a sample of professional basketball and soccer players. *Perceptual & Motor Skills*, *116*(1), 262–271. https://doi.org/10.2466/25.08.30.PMS.116.1.262-271

Stenling, A., Ivarsson, A., Johnson, U., & Lindwall, M. (2015). Bayesian structural equation modeling in sport and exercise psychology. *Journal of Sport & Exercise Psychology*, *37*(4), 410–420. https://doi.org/10.1123/jsep.2014-0330

Westre, K. R., & Weiss, M. R. (1991). The relationship between perceived coaching behaviors and group cohesion in high school football teams. *The Sport Psychologist*, *5*(1), 41–54. https://doi.org/10.1123/tsp.5.1.41

Whitton, S. M., & Fletcher, R. B. (2014). The Group Environment Questionnaire: A multilevel confirmatory factor analysis. *Small Group Research*, *45*(1), 68–88. https://doi.org/10.1177/1046496413511121

Widmeyer, W. N., Brawley, L. R., & Carron, A. V. (1990). The effects of group size in sport. *Journal of Sport & Exercise Psychology*, *12*(2), 177–190. https://doi.org/10.1123/jsep.12.2.177