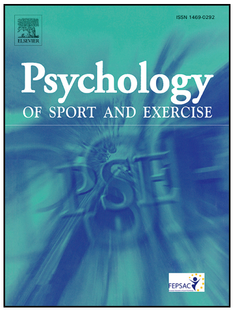
Journal Pre-proof




Factorial validity and measurement invariance of the Athlete Burnout Questionnaire (ABQ)

Michael C. Grugan, Luke F. Olsson, Robert Vaughan, Daniel.J. Madigan, Andrew P. Hill

PII: S1469-0292(24)00049-9

DOI: https://doi.org/10.1016/j.psychsport.2024.102638

Reference: PSYSPO 102638

To appear in: *Psychology of Sport & Exercise*

Received Date: 27 March 2023

Revised Date: 20 March 2024

Accepted Date: 4 April 2024

Please cite this article as: Grugan, M.C., Olsson, L.F., Vaughan, R., Madigan, D.J., Hill, A.P., Factorial validity and measurement invariance of the Athlete Burnout Questionnaire (ABQ), *Psychology of Sport & Exercise* (2024), doi: https://doi.org/10.1016/j.psychsport.2024.102638.

**Factorial Validity and Measurement Invariance of the Athlete Burnout Questionnaire (ABQ)**

Michael C. Grugan[1], Luke F. Olsson[2], Robert Vaughan[3], Daniel, J. Madigan[4], and Andrew P. Hill[4,5]

[1]Department of Psychology, Northumbria University, Newcastle, UK

[2]School of Sport, Rehabilitation, and Exercise Sciences, University of Essex, Colchester, UK

[3]School of Human and Social Sciences, University of West London, London, UK

[4]School of Science, Technology, and Health, York St John University, York, UK

[5]Graudate Department of Kinesiology, University of Toronto, Toronto, Canada

**Author Note**

Correspondence concerning this article should be addressed to Michael C. Grugan, Department of Psychology, Northumbria University, Newcastle, Northumberland Building, College Lane, NE1 8SG, UK. E-mail: michael.grugan@northumbria.ac.uk

**Declarations of interest**: None; **Funding**: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors; **Availability of data**: The authors do not have permission to share data. **Availability of code and other material**: The code and material in the current study are available from the corresponding author on reasonable request;


**Authors contributions**: The study was designed and conceptualised by all authors; Material preparation, data collection, and analysis were performed by all authors. The first draft of the manuscript was written by Michael C. Grugan. All authors reviewed and edited previous versions of the manuscript and approved the final manuscript.

Factorial Validity and Measurement Invariance of the Athlete Burnout Questionnaire (ABQ)

**Abstract**

The Athlete Burnout Questionnaire (ABQ) is the gold standard measure for burnout in athletes. However, previous assessments of factorial validity have: (a) tested overly restrictive measurement models; (b) provided mixed support for factorial validity; and (c) not been applied to assess measurement invariance across gender, sport type, or age. To address these issues, we used ABQ data provided by 914 athletes ($M_{age}$ = 21.75 years, $SD$ = 8.79) and examined factorial validity using confirmatory factor analysis (CFA) and exploratory structural equation modelling (ESEM) techniques. We also examined measurement invariance of the ABQ data across reported gender (female, male), sport type (individual, team), and age (≤ 18 years, > 18 years) groups. The analyses revealed that an ESEM model provided superior fit over the corresponding CFA model. In terms of measurement invariance, support was provided for the equivalence of the ABQ across each group. This means that researchers using the ABQ can collect data across these groups and examine potential differences with confidence that the ABQ is approximately invariant. In all, we provide evidence that the majority of ABQ items are key target construct indicators and the burnout construct (as measured by the ABQ) has the same structure and meaning to different athlete groups.

*Keywords*: athlete burnout; psychometrics; factorial validity; measurement invariance

## Introduction

Participation in sport is often an enjoyable, worthwhile, and highly rewarding experience (Kilpatrick et al., 2005). While this positive way of thinking about sport is familiar for many athletes, for others it is far removed from their current thoughts and feelings toward sport. That is, some athletes see sport as an endeavour that is overtaxing, unrewarding, and unenjoyable (Creswell & Eklund, 2006; Gustafsson et al., 2008). This cynical view of sport and its associated value is typified by athletes who experience high levels of athlete burnout (Eklund & DeFreese, 2020). Over the last two decades, researchers have dedicated considerable effort to better understand what athlete burnout is, how it differs from other experiential states, and what factors are most likely to underpin its development (Eklund & DeFreese, 2020; Gustafsson et al., 2017; Pacewicz et al., 2019). This line of research has been made possible by the development of conceptual models of athlete burnout (e.g., Coakley, 1992; Raedeke, 1997; Smith, 1986) and associated domain-specific measures (e.g., Eades, 1990; Isoard-Gautheur et al., 2018; Raedeke & Smith, 2001).

### Athlete Burnout

The most widely adopted model of athlete burnout was developed by Raedeke and Smith (Raedeke, 1997; Raedeke & Smith, 2001; 2009). As described by Raedeke and Smith, athlete burnout is a psychological syndrome characterised by a constellation of symptoms: *reduced sense of accomplishment*, RSA; *physical and emotional exhaustion,* EXH; and *sport devaluation,* SD. The first burnout symptom, RSA, reflects a sense of low accomplishment and personal inadequacy in sport. The second burnout symptom, EXH, reflects the perceived depletion of physical and emotional resources resulting from sport training and competition. The third burnout symptom, SD, reflects the development of a diminished and cynical view toward the benefits of sport participation (Raedeke & Smith, 2001). This conceptualisation of burnout is specific to athletes and the demands they face in sport (Eklund & DeFreese, 2020).

Athlete burnout is often characterised as an extreme and persistent form of sport disillusionment (Madigan et al., 2019). In this regard, it is unsurprising that burnout has been found to contribute toward diminished physical and psychological well-being among athletes. Researchers have found that burnout is positively associated with negative outcomes (e.g., depressed mood, psychological stress, and negative affect) and negatively associated with adaptive outcomes (e.g., coping skills, hope, perceived control, and optimism) among athletes (Gustafsson et al., 2017). Additionally, researchers have hypothesised that burnout is likely to give rise to long term performance impairment, illicit substance use, and sleep dysfunction (Eklund & DeFreese, 2020). In these regards, high levels of athlete burnout may confer vulnerability to a host of negative and damaging outcomes for athletes in sport.

While it is difficult to know exactly how many athletes are affected by burnout and potentially vulnerable to burnout-induced problems, some estimates suggest that up to 10% of athletes may experience meaningful levels of burnout (Gustafsson et al., 2007; Raedeke & Smith, 2009). In addition to evidence on worrying rates of burnout in athletes, researchers have found that average levels of athlete burnout have increased over the last two decades. Specifically, adopting a cross-temporal meta-analytical framework, Madigan et al. (2022) traced average levels of reported athlete burnout from 1997 to 2019. In this review of 91 studies ($N$ = 21,012), Madigan and colleagues found that mean levels of RSA and SD have increased over time. This evidence is important as it suggests that athletes are now at greater risk of burnout symptoms and more susceptible to its negative consequences. In this regard, the study of athlete burnout is extremely important.

**Conceptualising and Measuring Athlete Burnout**

The model of athlete burnout proposed by Raedeke and Smith (Raedeke, 1997; Raedeke & Smith, 2001; 2009) was developed based on Maslach and Jackson's (1981) conceptualisation of burnout in care and human service workers. One major distinction made between sport and human service institutions is that athletes do not provide a service to others

(Eklund and DeFreese, 2020). This meant that the three symptom dimensions of athlete burnout included a focus on *sport participation* and *sport performance* rather than *work with others*. This is why RSA captures low accomplishment and personal inadequacy *in sport performance* rather than *in one's work with others*. Similarly, SD captures cynical attitudes toward *sport participation* rather than *the recipients of one's healthcare or service*. Another important distinction made was that sport is a context in which physical demands are an obvious source of psychosocial stress (Eklund & DeFreese, 2020). This is why EXH includes a focus on both physical and emotional exhaustion.

To operationalise athlete burnout Raedeke and Smith (2001) developed the Athlete Burnout Questionnaire (ABQ). In line with the conceptual model of athlete burnout outlined above, the ABQ contains three separate 5-item subscales that measure RSA ("*I am not achieving much in sport*"), EXH ("*I am exhausted by the mental and physical demands of sport*"), and SD ("*I'm not into sport like I used to be*"). The process of arriving at this 15-item scale was iterative and involved multiple forms of robust evaluation (e.g., statistical tests of factorial validity and feedback from expert panels on content validity, suitability, and readability). While other measures of athlete burnout have since been developed (e.g., Isoard-Gautheur et al. 2018), the ABQ remains the most widely used and continues to be considered the gold standard measure for burnout in athletes (Eklund & DeFreese, 2020).

One reason for the dominance of the ABQ is that it has a strong conceptual grounding in a well-established model of burnout (Maslach & Jackson, 1981). However, just as importantly, the ABQ has typically performed well under psychometric scrutiny. This is evident in research showing support for convergent and discriminant validity, test-retest reliability, and scale internal consistency (Creswell & Eklund, 2006; Raedeke & Smith, 2001, 2009). While this may be the case, assessment of psychometric properties is, of course, an ongoing process and there are two aspects of validity evidence that require further examination – factorial validity (i.e., the extent to which ABQ items measure the intended burnout construct) and measurement invariance (i.e., the extent to which the burnout construct has the same structure and meaning to different athlete groups).

**Factorial Validity of the ABQ**

In the original scale validation study, as well as all subsequent studies examining the factorial validity of the ABQ, researchers have employed confirmatory factor analysis (CFA). This approach allows researchers to examine the relationships between indicators (e.g., ABQ items) and latent constructs (e.g., symptom dimensions of athlete burnout) in a pre-determined factor structure. In adopting this approach, some researchers have found reasonable support for the original three-factor ABQ. This is evident in studies where the CFA model specification for the ABQ provides acceptable fit to ABQ data (e.g., DeFreese & Smith, 2013; Raedeke & Smith, 2001; Ruser et al., 2020). Despite this support, it is important to note that other researchers have often found that the same CFA model specification provides either marginal or suboptimal fit to ABQ data (e.g., Appleby et al., 2022; Barcza-Renner et al., 2016; Casanova et al., 2023). To better understand why this mixed pattern of support exists, it is important to reflect on limitations of the CFA modelling technique.

While CFA is a strong technique with several modelling capabilities, it is not without limitation (Asparouhov & Muthén, 2009). One major limitation is that indicators in CFA are assumed to be pure indicators of the construct they are developed to measure (zero cross-loadings are permitted). The problem with this assumption is that indicators are imperfect and rarely provide a reflection of a single construct (Morin et al., 2020). Most multidimensional measures include indicators that can be expected (logically, theoretically, or empirically) to present construct-relevant associations with more than one factor (Morin, 2023). If these cross-loadings are incorrectly set to zero, the misspecification can result in poor model fit and biased factor correlations (Tóth-Király et al., 2017). This is problematic as researchers may then reject the model under examination in error and call into question whether the constructs being examined are conceptually distinct (Steenkamp & Maydeu-Olivares, 2022).

In terms of the ABQ, it is realistic that some items will present significant, and even reasonably large, cross-loadings. For example, RSA and SD are frequently the two most strongly correlated symptom dimensions of burnout ($r$ = .47 to .74; Raedeke and Smith, 2009). This overlap partly reflects the fact that the two burnout symptoms are attitudinal in nature and characterised by negative feelings (e.g., cynicism and dissatisfaction; Raedeke and Smith, 2009). We therefore expect RSA items to cross-load on the SD factor (and vice versa). Even if these cross-loadings are only small (e.g., $\lambda \leq .10$), ignoring them could undermine model fit and the discriminant validity of factors (Tóth-Király et al., 2017). This is important to acknowledge given that CFA sometimes provides poor model fit for the ABQ (e.g., Appleby et al., 2022; Barcza-Renner et al., 2016; Casanova et al., 2023) and very high factor correlations – especially between RSA and SD ($r$ > .85; Lower-Hoppe et al., 2022). Even though these results may be a function of the overly restrictive CFA model specification, it is easy to assume the evidence reflects issues with the ABQ and that modifications are required (e.g., removal of items; Casanova et al., 2023; Lower-Hoppe et al., 2022).

One technique that overcomes many of the limitations of CFA is exploratory structural equation modelling (ESEM). Unlike the CFA approach, indicators in ESEM are permitted to load on all factors, allowing for complex structure (i.e., cross-loadings). However, in line with CFA, ESEM enables researchers to examine an *a-priori* factor structure (using target rotation), obtain overall tests of model fit, and examine standard errors for individual parameter estimates. The key benefit of ESEM is that it helps researchers to achieve a more accurate representation of latent factors and factor correlations (Morin et al., 2020). In addition, ESEM helps researchers to better identify potentially problematic indicators. While cross-loadings are to be expected, they should be easy to explain (based on theory, logic, or empirical evidence) and weaker than corresponding target loadings (Morin et al., 2020). This technique is therefore needed to provide a more comprehensive and accurate assessment of the ABQ and identify any items that may need to be revised or removed.

**Measurement Invariance of the ABQ**

In the field of sport psychology, researchers often assume that measures (and their indicators) behave in the same manner for athletes from different groups. The problem with this assumption is that some measures function differently across groups (Wells, 2021). An illustrative example of such group-based differences is evident in the measurement of depression. Some depression measures include indicators that may have less relevance to depression in men than women (e.g., items measuring frequency of "*crying*" as a depressive symptom; Kim & Yoon, 2011). This means it is possible that a sample of men and women who are equally depressed could score different total depression scores when using such measures. This problem can lead to evidence of group differences that are a product of measurement non-invariance as opposed to true group differences in depression. To avoid such issues, it is important that researchers establish evidence that a construct has the same structure and meaning across different groups, and responses are not confounded by features of the respondents (i.e., evidence of measurement invariance, Putnick & Bornstein, 2016)

The potential for measurement non-invariance to interfere with group comparisons using ABQ data is an area that requires further examination. While numerous tests of measurement invariance have been conducted on translated versions of the ABQ (e.g., Isoard-Gautheur et al., 2010; Liu et al., 2022; Zhang et al., 2016), the psychometric properties of these scales are specific to the context and language in which the scales were examined. In terms of the original (English, 15-item ABQ) version of the ABQ, only one study dedicated to examining measurement invariance has been published. Lonsdale et al., (2006) tested the original ABQ and found evidence to support invariance across groups who differed in their method of reporting ABQ scores (online *versus* paper-and-pencil). More recently, Casanova et al. (2023) tested the ABQ for invariance but examined a shortened version with some slight wording amendments. In doing so, Casanova et al. found evidence to support invariance across groups who differed in their athlete category (athlete *versus* dancer), class standing (lower-class *versus* upper-class) and scholarship status (scholarship *versus* no scholarship). While this evidence is important, we still do not know if the original ABQ is invariant across other commonly examined groups in sport included in most samples.

When it comes to the ABQ, there has been encouragement from researchers to examine measurement invariance across gender and other sport populations (Gustafsson et al., 2007). In addition to gender, other sport populations that are of substantive importance are sport type and age. These three variables are important for at least two major reasons. The first reason is that researchers have found evidence of differences in ABQ scores across each of these variables. For example, Dubuc-Charbonneau et al. (2014) found evidence that female athletes reported higher EXH than male athletes, and Cremades and Wiggins (2008) found evidence that individual sport athletes reported higher RSA than team sport athletes. To properly evaluate such results, it is important to identify whether the measurement properties of the ABQ generalise across these groups (Marsh et al., 2016). If they do not, existing evidence on differences in athlete burnout may be invalid.

The second reason that gender, sport type, and age are important to examine is that researchers often collect and study ABQ data using heterogenous samples (e.g., multiple genders, sport types, and age ranges). This practice relies on the assumption that the underlying factors are measuring the same construct in the same way across these different groups. If we find that the athlete burnout construct (as measured by the ABQ) has a different structure or meaning to different groups, researchers may have to rethink how they design future ABQ research. To deal with non-invariance, it may be necessary to collect more homogenous samples or statistically control for variables such as age, gender, and sport type in the planned analysis.

## The Present Study

In line with the evidence presented above, there is a need to further examine the factorial validity of the ABQ using ESEM and measurement invariance approaches. To address these requirements, we conducted both single-group and multi-group analyses. In the single group-analyses, we examined the ABQ using both CFA and ESEM techniques. In the multi-group analyses, we examined measurement invariance of the ABQ across meaningful groups defined based on reported gender, sport type, and age. We hypothesised that: (a) ESEM would provide better model fit for the ABQ (than a corresponding CFA model specification); and (b) the ABQ would be invariant across identified groups.

# Method

## Participants

Three independent ABQ data sets were utilised in the present study. The three data sets have not been used previously in any published research. Data set one consisted of 575 adult athletes from a range of team and individual sports in the UK ($M_{age}$ = 24.81 years, *SD* = 9.85, age range = 18–59), data set two consisted of 182 adolescent athletes from a range of team and individual sports in the UK ($M_{age}$ = 17.10 years, *SD* = .54, age range = 16–17), and data set three consisted of 157 male footballers from youth and young adult teams in the UK ($M_{age}$ = 15.96 years, *SD* = 1.20, age range = 13–19). The data collected in each independent data set was approved under institutional ethical approval. In all cases, approved consent procedures were followed, and appropriate permissions were gained prior to inviting athletes to complete the voluntary paper-and-pencil study questionnaire.

In tests of factorial validity and measurement invariance, large sample sizes are required to obtain accurate parameter estimates and achieve adequate power (Hu et al., 2023). One method that researchers often use to obtain an appropriately large sample size for tests of invariance is to combine extant data sets into one large, pooled data set (van Dijk et at al., 2022). This practice is common in tests of measurement invariance in sport and exercise psychology (e.g., Grugan et al., 2021; Vlachopoulos, 2008). In adopting this approach, we pooled the ABQ data. In the combined data set of 914 athletes ($M_{age}$ = 21.75 years, *SD* = 8.79, age range = 13–59), meaningful groups were coded based on their reported gender ($n_1$ = 377 female athletes, $n_2$ = 532 male athletes), sport type ($n_1$ = 344 individual sport athletes, $n_2$ = 570 team sport athletes), and age ($n_1 = 416 \leq 18$ years, $n_2 = 498 > 18$ years).

We evaluated the appropriateness of this data set in relation to power-related guidelines and considerations in tests of measurement invariance. Based on a simulation study conducted by Sass et al. (2014), an overall sample size of 600 (300 per group) provides adequate power for detecting large non-invariance using stringent cut-off values under WLMSV estimation and various modelling conditions (average rejection rates for ΔCFI, ΔTLI, ΔRMSEA ≥ 80%). In this regard, our combined data set is reasonable for testing measurement invariance across the groups of interest (i.e., gender, sport type, and age).

**Measure**

**Athlete Burnout**. The Athlete Burnout Questionnaire (ABQ; Raedeke & Smith, 2001) was used to measure athlete burnout. This ABQ includes three 5-item subscales: RSA, EXH, and SD. All participants were instructed to think about their current sport involvement and rate how often they experienced the feelings identified in each item using a 5-point (1 = *almost never* to 5 = *almost always*) Likert scale.

**Data Analysis**

In the present study we examined both single-group and multi-group measurement models using WLSMV estimation for categorical variables in M*plus* 8.1 (Muthén & Muthén, 1998-2017). When compared to ML estimation, WLSMV is slightly less efficient at handling missing data (Asparouhov & Muthén, 2010). However, this was not considered an issue due to the extremely low level of missing data at the item level (< 1% for all items). In addition to screening for missing values, we also checked for imputation errors, re-coded the two reversed scored ABQ items (RSA1 and RSA14), and calculated item statistics and scale reliability estimates. To assess scale reliability, we computed McDonald's omega (ω) estimates for each of the three ABQ subscales.

**Single-group Measurement Models**. In line with previous research, we initially adopted a CFA approach to examine the factorial validity of the ABQ. In this model, indicators were constrained to load on first-order target factors only and all latent factors were specified to covary. However, as highlighted in the introduction, one issue with the CFA specification is that it can be highly restrictive. To address this issue, we also adopted an ESEM approach to examine the factorial validity of the ABQ. In this model, ABQ items were permitted to load on all first-order factors and all latent factors were specified to covary.

In line with previous psychometric research, we used multiple indices to evaluate overall model fit: $\chi^2$, CFI, TLI, RMSEA, and SRMR. However, as the $\chi^2$ is oversensitive to sample size and minor model misspecifications, we predominantly focused on the alternative model fit indices (e.g., CFI, TLI, and RMSEA). We considered models meeting the following criteria to reflect at least adequate fit: > .90 CFI, TLI, < .08 RMSEA, 90% CI < .05 to < .08, < .08 SRMR (Marsh et al., 2004). When evaluating the standardised factor loadings in each model, we considered the magnitude of the estimates (≥ .30 was considered meaningful), degree of cross-loading (the number of indicators loading meaningfully on more than one factor), and solution interpretability (Morin et al., 2020).

**Multi-group Measurement Models**. The first step in this process involved exploring the suitability of combining the three independent data sets into one large, pooled data set. This was achieved using the alignment methodology (Asparouhov & Muthén, 2023) to identify the percentage of approximately invariant parameters (i.e., factor loadings and response thresholds) across the groups of interest (in this case, the groups represent each independent data sets). When approximate invariance holds for > 80% of the parameters, the alignment methodology can be used to reliably compare latent means. We therefore used this threshold to identify whether it is reasonable to combine the three independent data sets.

The second step in this process involved testing the ABQ for measurement invariance across important athlete groups. We tested the following sequential measurement invariance models, as outlined by Morin (2023): *configural* (i.e., equality of measurement model, number of factors, indicators, and indicators-to-factors associations); *weak* (equality of factor loadings across groups); strong invariance (equality of response thresholds); *strict* (equality of the indicator uniquenesses); *latent variance-covariance* (equality of the factor variances and covariances); and *latent means* (equality of factor means). The Mplus syntax for these models was developed using De Beer and Morin's (2022) ESEM invariance syntax generator.

We examined measurement invariance across the coded gender, sport type, and age groups. In each assessment, the first stage involved examining the overall fit of each model. In the second stage, we examined changes between nested models using both the Mplus DIFFTEST function (MD$\Delta\chi^2$) and changes in the following alternative fit indices: CFI, RMSEA, and SRMR. While the chi-square difference test (MD$\Delta\chi^2$) provides a test of exact invariance, the change in alternative fit indices provide a test of approximate invariance (Millsap, 2005). In line with common recommendations, we relied predominantly on changes in the alternative fit indices and used the following criteria to identify measurement non-invariance: ($\Delta$CFI > -.002, $\Delta$RMSEA > +.010, $\Delta$SRMR > .010; Sass et al., 2014). These cut-off values are more stringent than traditional cut-off values used in tests or measurement invariance (e.g., Cheung & Rensvold, 2002) and are more appropriate for the proposed model specification (ESEM) and estimation method (WLSMV).

We supplemented the traditional tests of measurement invariance outlined above with the alignment methodology. We used this approach to: (a) explore the percentage of approximately invariant parameters across the groups of interest; and (b) compare latent means across groups in cases where evidence of approximate measurement invariance is satisfied. With fixed alignment, the factor means in each reference group (female athletes, individual sport athletes, and adolescent athletes) are fixed to 0. The methodology produces factor means for the non-reference group and identifies whether the estimates are statistically different from the reference group at the 5% significance level.

# Results

## Scale Reliability Estimates

The scale reliability estimates are reported in Table 2. For each of the three athlete burnout subscales, estimates for McDonald's omega were all acceptable (ω = .77 to .85).

## Single-group Measurement Models

**CFA**. The CFA model provided poor fit to the data ($\chi^2$ = 2109.99, *df* = 87, *p* < .001, CFI = .853, TLI =.822, RMSEA = .160 [.154, .160], and SRMR =.086). However, all factor loadings were significant (*p* < .001) and meaningful (λ ≥ .54). The standardised CFA factor correlations were positive, significant, and moderate-to-large in magnitude (*r* = .49 to .82). See Table 1 for model fit statistics and Table 2 for standardised factor loadings.

**ESEM**. The ESEM model provided good model fit ($\chi^2$ = 561.01, *df* = 63, *p* < .001, CFI = .964, TLI =.940, RMSEA = .093 [.086, .100], and SRMR =.027). In terms of parameter estimates, defined and interpretable factors for RSA, EXH, and SD were evident. This was reflected in the high percentage of meaningful target factor loadings (93% of cases) and small percentage of meaningful non-target factor loadings (13% of cases). While most items demonstrated a clean pattern of factor loadings across the three factors, four items were flagged due to meaningful cross-loading. These were items SD3 (target λ = .40; non-target λ = .41), RA7 (target λ = .30; non-target λ = .39), RA13 (target λ = .34; non-target λ = .42), and SD15 (target λ = .23; non-target λ = .34). In comparison to the CFA model, the standardised factor correlations were smaller in magnitude, yet still positive and significant (*r* = .25 to .66). See Table 1 for model fit statistics and Table 2 for standardised factor loadings.

## Multi-group Measurement Models

**Samples.** The ABQ was assessed for measurement invariance across the independent data sets using the alignment methodology. The approach identified that 92% of factor loadings and 95% of response thresholds were approximately invariant. This means that it was reasonable to combine the three independent data sets into one large, pooled data set.

**Gender**. For gender, the six increasingly restrictive models provided good fit. While all the $\chi^2$ difference tests (MD$\Delta\chi^2$) were significant (meaning exact invariance was not supported in any model comparison), changes in the alternative fit indices were below the specified cut-off values for four (out of five) of the nested model comparisons (meaning approximate invariance was supported in these cases). The evidence provides support for the equality of factor loadings, response thresholds, indicator uniquenesses, and factor variances and covariances. However, the equality of factor means was not fully supported. While the change in RMSEA and SRMR was below the identified cut-off values, the change in CFI ($\Delta$CFI = -.005) was not. This evidence suggests that there may be differences in latent burnout scores between the male and female groups.

The results from the alignment methodology were consistent with this evidence. We found that 96% of factor loadings and 100% of response thresholds were approximately invariant. The alignment methodology also provided support for differences in factor means. While there were no significant differences in levels of RSA, we found that: (a) levels of EXH were higher for the male group ($M$ = + .16, $p < .05$); and (b) levels of SD were lower for the male group ($M$ = + .16, $p < .05$).

**Sport Type**. For sport type, the six increasingly restrictive models provided good fit. While all the $\chi^2$ difference tests (MD$\Delta\chi^2$) were significant (meaning exact invariance was not supported in any model comparison), changes in the alternative fit indices were below the specified cut-off values for four (out of five) of the nested model comparisons (meaning approximate invariance was supported in these cases). The evidence provides support for the equality of factor loadings, response thresholds, indicator uniquenesses, and factor variances and covariances. However, the equality of factor means was not fully supported. While the change in SRMR was below the identified cut-off value, the change in CFI ($\Delta$CFI = -.005) and RMSEA was not ($\Delta$RMSEA = +.010). This evidence suggests that there may be differences in latent burnout scores between the individual and team sport groups.

The results from the alignment methodology were consistent with this evidence. We found that 84% of factor loadings and 98% of response thresholds were approximately invariant. The alignment methodology also provided support for differences in factor means. We found that: (a) levels of RSA were lower for the team sport group ($M$ = - .28, $p < .05$); (b) levels of EXH were higher for the team sport group ($M$ = + .23, $p < .05$); and (c) levels of SD were lower for the team sport group ($M$ = - .33, $p < .05$).

**Age**. For age, the six increasingly restrictive models provided good fit. While all the $\chi^2$ difference tests (MD$\Delta\chi^2$) were significant (meaning exact invariance was not supported in any model comparison), changes in the alternative fit indices were below the specified cut-off values for three (out of five) of the nested model comparisons (meaning approximate invariance was supported in these cases). The evidence provides support for the equality of factor loadings, response thresholds, and indicator uniquenesses. However, the equality of factor variances and covariances was not fully supported, nor was the equality of factor means. In the two final nested models, the change in CFI exceeded the identified cut-off value ($\Delta$CFI = -.004 and -.010, respectively). This evidence suggests that there may be differences in the structural parameters of the ABQ (e.g., variances, covariances, and/or latent means).

The results from the alignment methodology were consistent with this evidence. We found that 84% of factor loadings and 93% of response thresholds were approximately invariant. The alignment methodology also provided support for differences in factor means. While there were no significant differences in levels of RSA, we found that: (a) levels of EXH were lower for the adult group ($M$ = - .27, $p < .05$); and (b) levels of SD were higher for the adult group ($M$ = + .16, $p < .05$). The model fit statistics for all invariance models are reported in Table 1.

## Discussion

The first aim of the present study was to examine the factorial validity of the ABQ using both CFA and ESEM techniques. In line with our first hypothesis, we found that the

ESEM model provided superior fit over the more restrictive CFA model. In the ESEM model, while there was clear evidence of three distinct and discernable factors for the three symptom dimensions of the ABQ, some model misspecification was evident based on cross-loading. The second aim was to examine measurement invariance of the ABQ across important groups defined based on their reported gender (female, male), sport type (individual sport, team sport), and age (≤ 18 years, > 18 years). In line with our second hypothesis, we found evidence that supports the invariance of ABQ measurement properties required to make valid latent mean comparisons across these athlete groups.

**Single-group Measurement Models**

While some studies have found support for the original ABQ under a CFA model specification (DeFreese & Smith, 2013; Raedeke & Smith, 2001; Ruser et al., 2020), there are studies in which the same model provides either marginal or suboptimal fit to ABQ data (Appleby et al., 2022; Barcza-Renner et al., 2016; Lonsdale et al., 2006). In the present study, when specifying the ABQ under a CFA model specification, we found evidence of poor model fit. In the CFA model, individual factor loadings did not reveal any signs of potential model misspecification. In all cases, ABQ items provided significant and meaningful loadings on their target factor. However, in line with previous research, potential model misspecification was evident in the strong positive factor correlation between RSA and SD ($r$ > .80; see also Lower-Hoppe et al., 2022). Rather than assuming this result reflects poor discriminant validity, it is important to acknowledge the limitations of CFA (zero cross-loadings permitted). Given that RSA and SD are highly correlated (Raedeke and Smith, 2009), cross-loading items can logically be anticipated. By fixing these cross-loadings to zero, it is unsurprising that CFA often results in poor model fit and very high factor correlations.

To provide a more accurate representation of the latent factors and factor correlations in the ABQ, and learn more about potentially problematic items, it may be important to adopt an ESEM model specification. In doing so, we found that the ESEM model (which permits cross-loadings) provided better fit and lower factor correlations relative to the corresponding CFA model. For example, improved support for the distinction between RSA and SD was provided ($r$ = .66). This pattern of results is consistent with wider psychometric research in sport showing that ESEM outperforms CFA (e.g., Hill et al., 2016; Grugan et al., 2021; Myers et al., 2011). In addition to good model fit, support for the ABQ was evident in the pattern of ESEM factor loadings. That is, in all except one case (item SD15), items provided a meaningful loading on their target factor. The ESEM evidence therefore provides support for the factorial validity of the ABQ in that the empirical evidence (i.e., factor loadings and factor correlations) closely matches the corresponding conceptual model of athlete burnout.

While there was evidence to support the factorial validity of the ABQ, it is important to highlight that a small number of RSA and SD items were flagged due to meaningful cross-loading. In the case of RSA, we found that item RA7 ("*I am not performing up to my ability in sport*") made a meaningful non-target loading on the SD factor, while item RA13 ("*It seems that no matter what I do, I don't perform as well as I should*") made a meaningful non-target loading on the EXH factor. These cross-loadings reflect the presence of construct-relevant associations with non-target factors and can be explained based on theory and logic (Morin et al., 2020). For example, in interviewing athletes with elevated ABQ scores, Gustafsson et al. (2008) found evidence that a lack of personal accomplishment is sometimes an immediate precursor to experiences of SD and EXH. In this regard, it makes sense that items capturing persistent yet futile attempts to perform to an expected standard might share a meaningful association with SD and EXH.

The issue of cross-loading was also evident for two SD items. Item SD3 ("*The effort I spend in sport would be better off spent doing other things*") made a meaningful non-target loading on the EXH factor, while item SD15 ("*I have negative feelings in sport*") made a meaningful non-target loading on the RSA factor. This evidence of construct-relevant association can also be logically explained. For example, Gustafsson et al. (2008) found that some athletes with elevated ABQ scores referenced feelings of *dissatisfaction* with their personal performance. This might explain why item SD15 – which references "*negative feelings*" – emerged as a better indicator of RSA than SD. In the case of item SD3, it is conceivable that doubts about sporting participation (a key characteristic of SD) and feelings of lethargy and needing a break from sport (key characteristics of EXH) represent somewhat similar experiential states (Creswell & Eklund, 2006). This similarity would explain why this item made a comparably strong loading on these two factors.

The key point to emphasize is that the majority ABQ items appear to be key target construct indicators. This means that researchers can be confident in using item scores to measure each of the three latent athlete burnout symptoms. However, there are two important caveats to acknowledge. The first caveat is that item SD15 ("*I have negative feelings in sport*") failed to make a meaningful target factor loading. In reviewing previous research, we found that SD15 has previously been flagged for issues with clarity and been removed from the ABQ following factor analysis (Casanova et al., 2023; Isoard-Gautheur et al., 2010). If issues with this item persist, it may be advisable to replace "*negative feelings*" with a type of negative feeling more characteristic of SD (e.g., "*unenthusiastic*, "*cynical*", or "*pessimistic*"; Creswell & Eklund, 2006).

The second caveat to consider when evaluating the support provided for the ABQ is that we found a small number of items that are not distinct enough to distinguish between the target factor and other symptoms of burnout. In line with previous research (e.g., Casanova et al., 2023), we found that this issue was particularly apparent for items designed to measure RSA and SD (RA7, RA13, and SD3). The two RSA items capture a sense of reduced accomplishment (e.g., "*I am not achieving much in sport*"). However, these items could be refined to include the troubling sense of inadequacy or self-imposed verdict of failure that defines this symptom (e.g. "*It's painful to say, but I'm not performing up to my ability in sport*"). The SD item could also be refined by incorporating a deeper sense of contempt toward sport (e.g., "*I feel like the time I spend in my sport is being wasted*"). These are the types of change that may be required to: (a) ensure that each item is a better predictor of the target construct; and (b) help achieve a clearer interpretation between the factors.

**Multi-group Measurement Models**

An important aim in the present study was to examine measurement invariance of the ABQ. This type of assessment is important as many researchers are interested in whether ABQ scores differ across groups of athletes (e.g., Dubuc-Charbonneau et al., 2014; Cremades and Wiggins, 2008). An important prerequisite to this type of research is evidence for equality of factor loadings and response thresholds across groups. Only when this level of measurement invariance is supported can researchers confidently conclude that group differences in latent factor means reflect true group differences (Han et al., 2019). Without this evidence, it is impossible to rule out that such differences arise due to measurement non-invariance (Morin et al., 2011). Even though evidence of measurement invariance is clearly essential, many measures we use in sport and exercise psychology have not been properly examined for their measurement invariance across athlete groups (Pacewicz et al., 2022).

In terms of previous research on the original ABQ, the only study examining the original version of the ABQ (English, 15-item) for measurement invariance focussed on methods of data collection (Lonsdale et al., 2006). To build on this study and carry out research called for by burnout researchers (Gustafsson et al., 2007), we examined measurement invariance across reported gender (female, male), sport type (individual sport, team sport), and age (≤ 18 years, > 18 years) groups. In doing so, we found consistent evidence that the three-factor ABQ operates equivalently across the specified groups. This was evident in that: (a) each increasingly restrictive invariance model (configural, weak, strong, and strict) provided good fit to the data; (b) differences in the alternative fit measures between nested models provided evidence for the approximate invariance of factor loadings, response thresholds, and indicator uniquenesses; and (c) the percentage of approximately invariant parameters (i.e., factor loadings and response thresholds) in the ABQ across gender, sport type, and age groups was very high. These findings suggest that we can reliably make ABQ-based comparisons across these groups.

In all three tests of measurement invariance (gender, sport type, and age) we found evidence for differences in the latent means of the ABQ factors. For example, in line with previous research, we found that levels of RSA were lower for team sport athletes in comparison to individual sport athletes (e.g., Cremades and Wiggins, 2008), and levels of SD were higher in adult athletes in comparison to adolescent athletes (e.g., Madigan et al., 2022). We also found some differences that were inconsistent with previous research, such as levels of EXH being higher for male athletes rather than for female athletes (Dubuc-Charbonneau et al., 2014). Given the limited and inconsistent evidence that exists pertaining to differences in athlete burnout across gender, sport type, and age, more research in this area is clearly required. The evidence reported in the present study is important in this regard as it provides evidence of the necessary invariance required to make valid comparisons between latent ABQ factor means when studying these variables.

**Limitations and Future Research**

While the present study has several notable strengths, there are important limitations that warrant consideration. The first point is that we only examined first-order ABQ models in our tests of factorial validity and measurement invariance. This approach is consistent with the original ABQ specification (Raedeke & Smith, 2001). However, as researchers often examine burnout as a global construct, it may be important to examine the applicability of a hierarchical ABQ structure using both single and multi-group analyses. An additional point to highlight is that we used one pooled sample of data to examine factorial validity. Researchers may wish to establish whether the support we found (in addition to the potential areas of model misspecification we identified) is stable across multiple independent samples. This is because the stability of model parameters across samples is an important test of model applicability (Fabrigar & Wegener, 2012).

It is also important for researchers to examine the ABQ for measurement invariance in groups beyond those tested in the present study (e.g., elite *versus* non-elite athletes). One assessment that is required is measurement invariance across measurement occasions. With researchers using the ABQ to examine changes in burnout following intervention (e.g., Langan et al., 2015) or significant life events (e.g., COVID-19; Woods et al., 2022), this research is a priority. When testing for measurement invariance, if researchers find evidence of non-invariance (e.g., differences in fit measures between nested models that exceed specific cut-off values), they may look to estimate the magnitude of the misfit. This can be achieved by computing effect size measures of non-invariance that inform researchers about the degree of non-invariance and its practical importance (Gunn et al., 2020).

**Conclusion**

The major finding in the present study is that the athlete burnout construct (as measured using the ABQ) is approximately invariant across the reported gender (female, male), sport type (individual sport, team sport), and age (≤ 18 years, > 18 years) groups. Researchers using the ABQ can collect data across such groups and examine potential differences with confidence that the measure is acceptably invariant. We also found evidence that ABQ data is best modelled using an ESEM specification. While we found support for the factorial validity of the ABQ using this technique, it is important to highlight that some potentially problematic items were identified. We have therefore suggested areas of refinement to help to solve these issues and improve the alignment between the ABQ and the associated theoretical framework of athlete burnout.

## References


Appleby, R., Davis, P. A., Davis, L., Stenling, A., & Vickery, W. (2022). Preliminary psychometric validation of the Teammate Burnout Questionnaire. *Frontiers in Psychology*, *8*(1), 1–10.

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

Asparouhov, T., & Muthén, B. (2010). *Weighted least squares estimation with missing data*. Statmodel.

Asparouhov, T., & Muthén, B. (2023). Multiple group alignment for exploratory and structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(2), 169–191.

Barcza-Renner, K., Eklund, R. C., Morin, A. J. S., & Habeeb, C. M. (2016). Controlling coaching behaviors and athlete burnout: Investigating the mediating roles of perfectionism and motivation. *Journal of Sport and Exercise Psychology*, *38*(1), 30–44.

Casanova, M. P., Reeves, A. J., & Baker, R. T. (2023). Psychometric Properties of a Modified Athlete Burnout Questionnaire in the Collegiate Athletics Setting. *Journal of Sport Rehabilitation*, *32*(5), 581–589.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255.

Coakley, J. (1992). Burnout among adolescent athletes: A personal failure or social problem? *Sociology of Sport Journal*, *9*(3), 271–285.

Cremades, J. G., & Wiggins, M. S. (2008). Direction and intensity of trait anxiety as predictors of burnout among collegiate athletes. *Athletic Insight: The Online Journal of Sport Psychology*, *10*(2), 1–15.

Creswell, S. L., & Eklund, R. C. (2006). The convergent and discriminant validity of burnout measures in sport: A multi-trait/multi-method analysis. *Journal of Sports Sciences*, *24*(2), 209–220.

De Beer, L.T., & Morin, A.J.S (2022). *(B)ESEM invariance syntax generator for Mplus*. Retrieved from https://www.surveyhost.co.za/b_esem/

DeFreese, J. D., & Smith, A. L. (2013). Areas of worklife and the athlete burnout-engagement relationship. *Journal of Applied Sport Psychology*, *25*(2), 180–196.

Dubuc-Charbonneau, N., Durand-Bush, N., & Forneris, T. (2014). Exploring Levels of Student-Athlete Burnout at Two Canadian Universities. *Canadian Journal of Higher Education*, *44*(2), 135–151.

Eades, A. M. (1990). *An investigation of burnout of intercollegiate athletes: The development of the Eades Athlete Burnout Inventory* [Unpublished master's thesis]. University of California.

Eklund, R. C., & DeFreese, J. D. (2020). Athlete burnout. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (Vol. 2, 4th ed., pp. 1220–1240). Wiley.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory Factor Analysis*. Oxford University Press.

Grugan, M. C., Hill, A. P., Mallinson-Howard, S. H., Donachie, T. C., Olsson, L. F., Madigan, D. J., & Vaughan, R. S. (2021). Development and initial validation of the Perfectionistic Climate Questionnaire-Sport (PCQ-S). *Psychology of Sport and Exercise*, *56*(1), 101997.

Gunn, H. J., Grimm, K. J., & Edwards, M. C. (2020). Evaluation of six effect size measures of measurement non-invariance for continuous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(4), 503–514.

Gustafsson, H., DeFreese, J. D., & Madigan, D. J. (2017). Athlete burnout: Review and recommendations. *Current Opinion in Psychology*, *16*(1), 109–113.

Gustafsson, H., Hassmén, P., Kenttä, G., & Johansson, M. (2008). A qualitative analysis of burnout in elite Swedish athletes. *Psychology of Sport and Exercise*, *9*(6), 800–816.

Gustafsson, H., Kenttä, G., Hassmén, P., & Lundqvist, C. (2007). Prevalence of Burnout in Competitive Adolescent Athletes. *Sport Psychologist*, *21*(1), 21–37.

Han, K., Colarelli, S. M., & Weed, N. C. (2019). Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. *Psychological Assessment*, *31*(12), 1481–1496.

Hill, A. P., Appleton, P. R., & Mallinson, S. H. (2016). Development and initial validation of the Performance Perfectionism Scale for Sport (PPS-S). *Journal of Psychoeducational Assessment*, *34*(7), 653–669.

Hu, C., Pellegrini, E. K., & Chung, G. W. (2023). Measurement Equivalence/Invariance Across Groups, Time, and Test Formats. In L. R. Ford, & T. A. Scandura (Eds.), *The Sage Handbook of Survey Development and Application* (pp. 170–182). Sage.

Isoard-Gautheur, S., Martinent, G., Guillet-Descas, E., Trouilloud, D., Cece, V., & Mette, A. (2018). Development and evaluation of the psychometric properties of a new measure of athlete burnout: The Athlete Burnout Scale. *International Journal of Stress Management*, *25*(1), 108–123.

Isoard-Gautheur, S., Oger, M., Guillet, E., & Martin-Krumm, C. (2010). Validation of a French version of the Athlete Burnout Questionnaire (ABQ): In competitive sport and physical education context. *European Journal of Psychological Assessment*, *26*(3), 203–211.

Kilpatrick, M., Hebert, E., & Bartholomew, J. (2005). College students' motivation for physical activity: Differentiating men's and women's motives for sport participation and exercise. *Journal of American College Health*, *54*(2), 87–94.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, *18*(2), 212–228.

Langan, E., Toner, J., Blake, C., & Lonsdale, C. (2015). Testing the effects of a self-determination theory-based intervention with youth Gaelic football coaches on athlete motivation and burnout. *The Sport Psychologist*, *29*(4), 293–301.

Liu, H., Wang, X., Wu, D. H., Zou, Y. D., Jiang, X. B., Gao, Z. Q., ... & Liu, J. D. (2022). Psychometric properties of the Chinese translated Athlete Burnout Questionnaire: Evidence from Chinese collegiate athletes and elite athletes. *Frontiers in Psychology*, *13*(1), 823400.

Lonsdale, C., Hodge, K., & Rose, E. A. (2006). Pixels vs. Paper: Comparing Online and Traditional Survey Methods in Sport Psychology. *Journal of Sport & Exercise Psychology*, *28*(1), 100–108.

Lower-Hoppe, L. M., Lee, W., Brgoch, S. M., Ryder, A., & Lowe, C. (2022). Making the connection: Examining the antecedents and consequences of athlete burnout among collegiate sport club athletes. *Journal of Sport Behavior*, *45*(4), 108–131.

Madigan, D. J., Gustafsson, H., Smith, A., Raedeke, T., & Hill, A. P. (2019). The BASES expert statement on burnout in sport. *The Sport and Exercise Scientist, 61(1)*, 6–7.

Madigan, D. J., Olsson, L. F., Hill, A. P., & Curran, T. (2022). Athlete burnout symptoms are increasing: A cross-temporal meta-analysis of average levels from 1997 to 2019. *Journal of Sport and Exercise Psychology*, *44*(3), 153–168.

Marsh, H. W., Hau, K. T., & Wen, Z. L. (2004). In search of golden rules: comment on approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu

& Bentler (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341.

Marsh, H. W., Parker, P. D., Morin, A. J. S. (2016). Invariance testing across samples and time: Cohort-sequence analysis of perceived body composition. In N. Ntoumanis, & N. D. Myers (Eds.), *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists* (pp. 101–130). Wiley.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, *2*(2), 99–113.

Millsap, R. E. (2005). Four Unresolved Problems in Studies of Factorial Invariance. In A. Maydeu-Olivares, & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 153–171). Routledge.

Morin, A. J. S. (2023). Exploratory structural equation modeling. In R. H. Hoyle (Ed)., *Handbook of Structural Equation Modeling* (2nd ed., pp. 1220–1240). Guilford.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modelling (ESEM), and bifactor-ESEM. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (Vol. 2, 4th ed., pp. 1044–1073). Wiley.

Morin, A. J., Moullec, G., Maiano, C., Layet, L., Just, J. L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Revue d'Epidemiologie et de Sante Publique*, *59*(5), 327–340.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Myers, N. D., Chase, M. A., Pierce, S. W., & Martin, E. (2011). Coaching efficacy and exploratory structural equation modeling: A substantive-methodological synergy. *Journal of Sport and Exercise Psychology*, *33*(6), 779–806.

Pacewicz, C. E., Hill, C. R., Lee, S., Myers, N. D., Prilleltensky, I., McMahon, A., ... & Brincks, A. M. (2022). Testing measurement invariance in physical education and exercise science: A tutorial using the well-being self-efficacy scale. *Measurement in Physical Education and Exercise Science*, *26*(2), 165–177.

Pacewicz, C. E., Mellano, K. T., & Smith, A. L. (2019). A meta-analytic review of the relationship between social constructs and athlete burnout. *Psychology of Sport and Exercise*, *43*, 155–164.

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*(1), 71–90.

Raedeke, T. D. (1997). Is athlete burnout more than just stress? A sport commitment perspective. *Journal of Sport and Exercise Psychology*, *19*(4), 396–417.

Raedeke, T. D., & Smith, A. L. (2001). Development and preliminary validation of an athlete burnout measure. *Journal of Sport and Exercise Psychology*, *23*(4), 281–306.

Raedeke, T. D., & Smith, A. L. (2009). *The Athlete Burnout Questionnaire Test Manual*. Fitness Information Technology.

Ruser, J. B., Yukhymenko-Lescroart, M. A., Gilbert, J. N., Gilbert, W., & Moore, S. D. (2020). Gratitude, coach–athlete relationships, and burnout in collegiate student-athletes. *Journal of Clinical Sport Psychology*, *15*(1), 37–53.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 167–180.

Smith, R. E. (1986). Toward a cognitive-affective model of athletic burnout. *Journal of Sport and Exercise Psychology, 8*(1), 36–50.

Steenkamp, J. B. E., & Maydeu-Olivares, A. (2023). Unrestricted factor analysis: a powerful alternative to confirmatory factor analysis. *Journal of the Academy of Marketing Science*, *51*(1), 86–113.

Tóth-Király, I., Bõthe, B., Rigó, A., & Orosz, G. (2017). An illustration of the exploratory structural equation modeling (ESEM) framework on the passion scale. *Frontiers in Psychology*, *8*(1), 1–15.

van Dijk, W., Schatschneider, C., Al Otaiba, S., & Hart, S. A. (2022). Assessing measurement invariance across multiple groups: When is fit good enough? *Educational and Psychological Measurement*, *82*(3), 482–505.

Vlachopoulos, S. P. (2008). The basic psychological needs in exercise scale: measurement invariance over gender. Structural Equation Modeling: A Multidisciplinary Journal, *15*(1), 114–135.

Wells, C. S. (2021). *Assessing measurement invariance for applied research*. Cambridge University Press.

Woods, S., Dunne, S., Gallagher, P., & Harney, S. (2022). Is a pandemic as good as a rest? Comparing athlete burnout and stress before and after the suspension of organised team sport due to Covid-19 restrictions, and investigating the impact of athletes' responses to this period. *Psychology of Sport and Exercise*, *60*(1), 102168.

Zhang, C. Q., Si, G., Chung, P. K., & Gucciardi, D. F. (2016). Mindfulness and burnout in elite junior athletes: The mediating role of experiential avoidance. *Journal of Applied Sport Psychology*, *28*(4), 437-451.

**Table 1**. *Goodness of fit statistics for CFA, ESEM, and ESEM invariance measurement models.*

| Model | WLSMV $\chi^2$ (*df*) | CFI | TLI | RMSEA | RMSEA 90% CI | SRMR | MD$\Delta\chi^2$ | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| ABQ Factorial Validity ($N$ = 914) | | | | | | | | | | |
| CFA | 2109.99*** (87) | .853 | .822 | .160 | [.154, .165] | .086 | – – | – – | – – | – – |
| ESEM | 561.01*** (63) | .964 | .940 | .093 | [.086, .100] | .027 | – – | – – | – – | – – |
| ABQ Gender Invariance ($n_1$ = 377 female athletes, $n_2$ = 532 male athletes, $N$ = 909) | | | | | | | | | | |
| Configural | 589.81*** (126) | .968 | .946 | .090 | [.083, .097] | .028 | – – | – – | – – | – – |
| Weak | 609.73*** (162) | .969 | .959 | .078 | [.071, .085] | .036 | 132.62*** (36) | +.001 | -.012 | +.008 |
| Strong | 640.54*** (204) | .969 | .969 | .069 | [.063, .075] | .037 | 97.57*** (42) | .000 | -.009 | +.001 |
| Strict | 631.72*** (219) | .971 | .972 | .064 | [.059, .070] | .039 | 34.67** (15) | +.002 | -.005 | +.002 |
| Latent Var-Cov | 455.72*** (225) | .984 | .985 | .047 | [.041, .054] | .044 | 14.37* (6) | +.013 | -.017 | +.005 |
| Latent Means | 526.50*** (228) | .979 | .981 | .054 | [.048, .060] | .045 | 26.99*** (3) | -.005 | +.007 | +.001 |
| ABQ Sport Type Invariance ($n_1$ = 344 individual sport athletes, $n_2$ = 570 team sport athletes, $N$ = 914) | | | | | | | | | | |
| Configural | 598.09*** (126) | .966 | .943 | .091 | [.083, .098] | .029 | – – | – – | – – | – – |
| Weak | 596.37*** (162) | .969 | .959 | .077 | [.070, .083] | .036 | 122.48*** (36) | +.003 | -.014 | +.007 |
| Strong | 594.21*** (204) | .972 | .971 | .065 | [.059, .071] | .036 | 68.51** (42) | +.003 | -.012 | .000 |
| Strict | 620.96*** (219) | .971 | .972 | .063 | [.058, .069] | .040 | 56.10*** (15) | -.001 | -.002 | +.004 |
| Latent Var-Cov | 451.34*** (225) | .984 | .985 | .047 | [.041, .053] | .044 | 15.34* (6) | +.013 | -.016 | +.004 |
| Latent Means | 570.22*** (228) | .975 | .977 | .057 | [.051, .063] | .045 | 39.61*** (3) | -.009 | +.010 | +.001 |
| ABQ Age Invariance ($n_1$ = 416 ≤ 18 years, $n_2$ = 498 > 18 years, $N$ = 914) | | | | | | | | | | |
| Configural | 664.36*** (126) | .964 | .940 | .097 | [.090, .104] | .031 | – – | – – | – – | – – |
| Weak | 623.12*** (162) | .969 | .960 | .079 | [.072, .086] | .035 | – – | +.005 | -.018 | +.004 |
| Strong | 681.54*** (204) | .968 | .967 | .072 | [.066, .078] | .038 | 127.19*** (42) | -.001 | -.007 | +.003 |
| Strict | 703.82*** (219) | .968 | .969 | .070 | [.064, .075] | .039 | 47.64*** (15) | .000 | -.002 | +.001 |
| Latent Var-Cov | 773.42*** (225) | .964 | .966 | .073 | [.067, .079] | .067 | 71.03*** (6) | -.004 | +.003 | +.028 |
| Latent Means | 922.37*** (228) | .954 | .958 | .082 | [.076, .087] | .069 | 55.42*** (3) | -.010 | +.009 | +.002 |

*Note*. WLSMV = Robust variance-adjusted weighted least squares estimation; *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 square residual; MD$\Delta\chi^2$ = Mplus DIFFTEST function; The MD$\Delta\chi^2$ failed to compute for the weak invariance model testing age invariance. CFA= Confirmatory factor analysis; ESEM = Exploratory structural equation modelling; ESEM models were estimated with target oblique rotation; ***$p$ < .001, **$p$ < .01, *$p$ < .05.

**Table 2**. *Item statistics, scale reliability estimates, and standardized factor loadings for CFA and ESEM models*

| | | | | CFA | ESEM | | |
|---|---|---|---|---|---|---|---|
| Item | *M* | *s* | ω | Target Factor Loading | RA Factor Loading | EXH Factor Loading | SD Factor Loading |
| RSA1 | 2.79 | 1.13 | | **.54$^{***}$ (.03)** | **.83$^{***}$ (.03)** | -.22$^{***}$ (.02) | -.06 (.03) |
| RSA5 | 2.45 | 1.21 | | **.79$^{***}$ (.02)** | **.53$^{***}$ (.03)** | .05$^{*}$ (.02) | .29$^{***}$ (.03) |
| RSA7 | 2.90 | 1.18 | .77 | **.75$^{***}$ (.02)** | **.30$^{***}$ (.03)** | .14$^{***}$ (.03) | .39$^{***}$ (.04) |
| RSA13 | 2.51 | 1.11 | | **.73$^{***}$ (.02)** | **.34$^{***}$ (.04)** | .42$^{***}$ (.03) | .13$^{***}$ (.04) |
| RSA14 | 2.80 | 1.07 | | **.56$^{***}$ (.03)** | **.90$^{***}$ (.03)** | -.05$^{*}$ (.02) | -.21$^{***}$ (.03) |
| EXH2 | 2.53 | 1.05 | | **.62$^{***}$ (.02)** | -.10$^{***}$ (.03) | **.84$^{***}$ (.02)** | -.16$^{***}$ (.03) |
| EXH4 | 2.37 | 1.02 | | **.77$^{***}$ (.02)** | -.01 (.03) | **.83$^{***}$ (.02)** | -.04 (.03) |
| EXH8 | 2.26 | 1.10 | .85 | **.84$^{***}$ (.02)** | .08$^{*}$ (.03) | **.69$^{***}$ (.02)** | .15$^{***}$ (.03) |
| EXH10 | 2.42 | 1.09 | | **.83$^{***}$ (.01)** | .08$^{**}$ (.03) | **.72$^{***}$ (.02)** | .10$^{**}$ (.03) |
| EXH12 | 2.23 | 1.11 | | **.81$^{***}$ (.02)** | .06 (.03) | **.74$^{***}$ (.02)** | .07$^{*}$ (.03) |
| SD3 | 2.02 | 1.10 | | **.66$^{***}$ (.03)** | -.03 (.04) | .41$^{***}$ (.03) | **.40$^{***}$ (.04)** |
| SD6 | 2.47 | 1.36 | | **.86$^{***}$ (.01)** | -.03 (.02) | -.08$^{***}$ (.02) | **.95$^{***}$ (.02)** |
| SD9 | 2.44 | 1.39 | .84 | **.85$^{***}$ (.01)** | .18$^{***}$ (.03) | -.06$^{**}$ (.02) | **.75$^{***}$ (.02)** |
| SD11 | 2.75 | 1.37 | | **.79$^{***}$ (.02)** | .03 (.03) | -.11$^{***}$ (.02) | **.84$^{***}$ (.03)** |
| SD15 | 1.99 | 1.13 | | **.68$^{***}$ (.02)** | .34$^{***}$ (.04) | .26$^{***}$ (.03) | .23$^{***}$ (.04) |
| | | | | Factor Correlations F1 | | .49$^{***}$ (.03) | .82$^{***}$ (.02) |
| | | | | F2 | .25$^{***}$ (.03) | | .56$^{***}$ (.03) |
| | | | | F3 | .66$^{***}$ (.02) | .43$^{***}$ (.03) | |

*Note*. RSA = Reduced sense of accomplishment; EXH = Physical and emotional exhaustion; SD = Sport devaluation; Bold typeface denotes meaningful loading (≥ .30) on target factor; Underlined typeface denotes meaningful cross-loading (≥ .30) on non-target factor; Standard errors reported in parentheses; Values below the diagonal (see factor correlations) are ESEM factor correlations; Values above the diagonal (see factor correlations) are CFA factor correlations; *N* = 914; ****p* < .001; ***p* < .01; **p* < .05.

**Highlights**

- The Athlete Burnout Questionnaire (ABQ) is considered the gold standard measure for burnout in athletes.
- The aim of the present study was to further validate the ABQ using Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modelling (ESEM) techniques.
- An ESEM model specification provided superior fit over a corresponding CFA model specification.
- Support for approximate measurement invariance (i.e., equality of factor loadings, response thresholds, and indicator uniquenesses) was supported across gender (female, male), sport type (individual, team), and age (≤ 18 years, > 18 years) groups.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

The authors declare that they have no conflict of interest.