Original paper

Validation of the Acute Recovery and Stress Scale (ARSS) and the Short Recovery and Stress Scale (SRSS) in three English-speaking regions

Sarah Kölling[1,2], Paul Schaffran[1], Adam Bibbey[3], Michael Drew[4], Ben Raysmith[4,5], Anu Nässi[1], & Michael Kellmann[1,6]

[1]*Faculty of Sport Science, Ruhr University Bochum, Bochum, Germany,* [2]*Department of Sport Science, Stellenbosch University, Stellenbosch, South Africa,* [3]*Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom,* [4]*Australian Institute of Sport, Belconnen, ACT, Australia,* [5]*Department of Medical and Health Sciences, Linköping University, Linköping, Sweden,* [6]*School of Human Movement and Nutrition Sciences, The University of Queensland, St Lucia, Australia.*

Correspondence:

Sarah Kölling, Faculty of Sport Science, Ruhr University Bochum, Gesundheitscampus Nord 10, D-44801 Bochum, Germany. E-Mail: sarah.koelling@rub.de

ORCiD: https://orcid.org/0000-0002-6415-424X

Paul Schaffran, paul.schaffran@rub.de

Adam Bibbey, abibbey@brookes.ac.uk

Michael Drew, michael.drew@ausport.gov.au

Ben Raysmith, ben.raysmith@athletics.org.au

Anu Nässi, anu.naessi@rub.de

Michael Kellmann, michael.kellmann@rub.de

Word count: 5.349

Tables: 7

Figures: 0

Disclosure statement:

The authors report no conflict of interest.

Abstract

The Acute Recovery and Stress Scale (ARSS) and the Short Recovery and Stress Scale (SRSS) are well-established monitoring tools in German-speaking countries. This study aimed at validating them for English-speaking populations. Overall, 996 athletes (536 males, 24.9 ± 9.1 years) of Australia/New Zealand (n = 380), the United Kingdom (n = 316), and North America (n = 300) participated. The 32-item ARSS consists of eight scales. These scales constitute the eight items of the SRSS with the corresponding ARSS items as descriptors. Confirmatory factor analysis (CFA), internal consistency (α), and discriminatory power of the items (r_{it}) were calculated for the total and subsamples separately. Satisfactory discriminatory power (r_{it} > .30) for all ARSS and SRSS items and acceptable internal consistency (α > .70) was achieved. CFA indicated good fit indices for the total sample and subsamples, and strong measurement invariance was found across subsamples and gender. Correlations between corresponding scales and items (r_s = .68 - .78) support theoretical congruency as well as independent usage of both questionnaires. Construct validity of both tools is shown through hypothesis-conforming correlations with the Recovery-Stress Questionnaire for Athletes. Future studies may apply the ARSS and SRSS as monitoring tools in English-speaking regions worldwide.

Keywords: *Monitoring, training, psychology, assessment, measurement invariance*

Introduction

Training monitoring can be considered an established routine in high-performance settings, as the necessity to balance training load and recovery has been recognised to ensure optimal preparation and athletes' health in the long-term (Bourdon et al., 2017; Kellmann et al., 2018). The ideal monitoring process is characterised by a multi-methodological approach and involves the assessment of various markers including performance tests, blood-borne parameters and self-report measures. As subjective questionnaires are cost effective and practical, they are commonly used (Saw, Main, & Gastin, 2015b). Moreover, it has been shown that questionnaires provide useful data which seem to be superior to objective parameters (Meeusen et al., 2013; Saw, Main, & Gastin, 2016). In the field of overtraining research, changes in mood are associated with the overtraining syndrome, and, due to its long-term genesis and gradual development, mood disturbances seem to predict non-functional processes more reliably than physiologic parameters (Meeusen & De Pauw, 2018; Meeusen et al., 2013). In their systematic review, Saw et al. (2016) highlight that self-report measures reflect acute and chronic training loads with greater sensitivity and consistency than objective measures.

Among self-report measures, a variety of tools is available assessing training response via internal training load, mood, and the recovery-stress state (Saw, Kellmann, Main, & Gastin, 2017). The available tools can be classified in terms of the underlying constructs, such as one- or multidimensional, and whether stressors, in terms of influencing factors such as training load, or resulting symptoms, in terms of outcome variables such as well-being or mood, are assessed (Saw et al., 2016). Using the *Organizational Stressor Indicator for Sport Performers* (OSI-SP; Arnold, Fletcher, & Daniels, 2013), frequency, intensity and duration of sport-specific stressors can be identified. As it covers the experience of athletes over the past month, this scale may be applicable for specific occasions within a season. The *Perceived Stress Scale* (PSS; Cohen, Kamarck, & Mermelstein, 1983) measures the degree (i.e., frequency) to which situations during the last month are appraised as stressful (i.e., unpredictable, uncontrollable, overloading). Yet, the time frame is not suitable for daily monitoring and it is a global scale that does not cover sport-specific situations. Therefore, the target group or application area of the questionnaire needs to be considered before selecting and interpreting data. Another example is the *Profile of Mood States* (POMS; McNair, Lorr, & Droppleman, 1992). The mood dimensions *Tension*, *Depression*, *Anger*, *Vigour*, *Fatigue*, and *Confusion* have been shown to predict short-term performance and to reflect increases/decreases of training workload in sports (Andreato et al., 2014; Beedie, Terry, & Lane, 2000). However, it has been criticised that clinical issues are overemphasized while recovery aspects are disregarded and sport-specific inferences cannot be drawn explicitly (Mäetsu, Jürimäe, & Jürimäe, 2005; Nässi, Ferrauti, Meyer, Pfeiffer, & Kellmann, 2017b). On the other hand, the *Session-Rating of Perceived Exertion* (sRPE) is a common one-dimensional scale to determine the internal training load which is established in research and practices from the perspective of exercise and training science (Foster et al., 2001; Foster, Rodriguez-Marroyo, & de Koning, 2017). For instance, training responses of an entire team can be compared and individual adaptations to the training programme may be derived. One problem, as pointed out by Meeusen et al. (2013) in terms of the development of the overtraining syndrome, is that psychological signs of maladaptation (over time) remain unnoticed. Another well-known instrument is the *Recovery-Stress Questionnaire for Athletes* (RESTQ-Sport; Kellmann & Kallus, 2016) that assesses sport-specific and general components of recovery and stress multidimensionally. With 76 items, however, it is rather time-consuming and these aspects are rated retrospectively (in the past three days/nights). It is not applicable to reflect the acute recovery-stress state and short-term fluctuations thereof in settings of high-frequency measurements. For further discussion of advantages and disadvantages of current tools available for the training monitoring refer to Nässi et al. (2017a) and Saw et al. (2017).

Overall, choosing an appropriate tool depends on several aspects, such as the purpose of the measurement (e.g., long-term monitoring), the feasibility in the training context (e.g., daily use, every 1 to 4 weeks, key time points), and the dimensions that align with the measurement intention (e.g., stressors, symptoms, emotions, mood, recovery, stress) based on theoretical foundations (Saw et al., 2017). One major requirement is the economical implementation to reduce training interruptions and the burden that is put on the athlete, as this may affect the stakeholders' compliance. At the same time, questionnaires should be developed following scientific standards and empirical evidence. The shortening of existing questionnaires is a practical approach, but only as long as psychometric properties (i.e., validity and reliability) are still fulfilled (Horvath & Röthlin, 2018). As there was a paucity in sport science and practice concerning the latter aspect, two instruments were designed which are practical and at the same time approach the multidimensional facets of recovery and stress.

While several definitions and research areas exist around the term stress (e.g., external stressors vs. stress responses, interaction between environment and individual), the tools presented in the current paper were developed based on the assumption of the 'scissors' model' (Kellmann, 2002). One main pillar is the subjective perception of an external demand (i.e., objective load) resulting in individual responses which may differ between individuals (Lazarus, 1991). In the context of sports, this implies that a specific training session will be perceived differently from a team of athletes depending on their current condition (e.g., health status). In relation to the transactional stress theory (Lazarus & Folkman, 1984), the scissors' model describes the stress level of an individual in the aftermath of appraisal and coping processes to deal with the demand. Primary (i.e., situation as stressor) and secondary (i.e., capacity to deal with the stressor) appraisals are important components of the recovery-stress balance, yet other un-specific external and/or internal aspects (e.g., muscle fatigue) play an important role as well. Thus, within this framework, the conceptualisation of stress is also related to the terms 'fatigue' and 'strain' (in analogy to material sciences and physics), because cognitive evaluations following primary appraisal (e.g., threat) do not necessarily need to take place to cause an increased stress level (Kallus, 2016). The second pillar postulates that recovery and stress processes are interrelated, but recovery is not merely defined by the absence of stress (Senttä & Hassmén, 1998). An important implication is, therefore, that a high stress state is manageable within the individual's stress capacity as long as the recovery demands are met appropriately (Kellmann, 2002). For instance, an athlete can be highly stressed (e.g., due to high training load) and highly recovered (e.g., due to personal accomplishments in his/her sport) at the same time. This further implies that a state of acute high stress and/or acute need for recovery can be usually tolerated, whereas chronic stress and underrecovery become maladaptive in the long-term. Consequently, the process over time is of paramount importance, as the accumulation of demands and/or the exhaustion of resources outside of the individual's limits of capacity are postulated to be the critical factors in this framework (Kellmann, 2002). Moreover, recovery is considered to be a restorative process (i.e., relative to time) comprising several facets such as physiological and psychological mechanisms, and social activities (Kellmann et al., 2018). The differentiation of recovery facets is important, as non-sport stressors influence the athlete's biopsychosocial state that may not be captured through the assessment of the 'bare' training load (Kellmann et al., 2018).

The German *Acute Recovery and Stress Scale* (ARSS; Hitzschke et al., 2016; Kellmann, Kölling, & Hitzschke, 2016) consists of a list of 32 adjectives which are summarised into four recovery and four stress scales representing physical, mental, emotional and overall dimensions. These were developed through an extensive exploratory development process in several phases including exploratory and confirmatory factor analyses among different athletic populations. Based on these scales, the *Short Recovery and Stress Scale* (SRSS; Hitzschke et al., 2015; Kellmann et al., 2016) was derived to allow for frequent measurements. One of the SRSS's special features is that its eight items are further explained by the single adjectives of the ARSS, but those adjectives are not answered separately. The usefulness and validity of both tools has been demonstrated in different settings (Hitzschke et al., 2017; Kölling et al., 2015; Pelka et al., 2017).

To provide these promising tools for a broader target group, and as there was no established scale comparable to the concept of the SRSS in areas of international sport science, they have been translated into English. Initial indications of their validity have been reported by Nässi, Ferrauti, Meyer, Pfeiffer, and Kellmann (2017a). However, those results were retrieved from only 267 participants, who were mainly located in Australia (65%). The applicability of the ARSS and the SRSS in the English-speaking areas of North America has been disregarded so far. As the issue of different language styles needs to be addressed (Kachru, Kachru, & Nelson, 2006; Trudgill & Hannah, 2017), the present study aimed at replicating the construct validity and presenting further evidence of the validity and reliability of the English ARSS and SRSS among a larger sample size as well as separately among three English-speaking areas covering Australia/New Zealand, the United Kingdom, and North America. Moreover, measurement invariance of the ARSS will be tested via multigroup confirmatory factor analysis to assess differences in the latent constructs between those regions as well as between gender.

Method

Participants

The overall sample consisted of 996 English-speaking athletes (*n* = 536 males, *n* = 460 females) with an average of 24.9 ± 9.1 years. At the time of participation, 82.5% reported to compete regularly on a regional level or higher, 12.3% were recreational and 5.1% collegiate athletes. The average training duration was 10.7 ± 13.4 hours per week. The majority was engaged in team sports played with balls (53.2%), the remaining athletes (46.8%) participated in individual sports (e.g., athletics, triathlon, martial arts). The subsamples can be divided into Australia/New Zealand (ANZ, *n* = 380), United Kingdom (UK, *n* = 316), and North America including USA and Canada (NA, *n* = 300). As this was a new data collection, data of the study from Nässi et al. (2017a) were not included. Gender and level distributions of the subsamples are depicted in Table 1.

Measures

Acute Recovery and Stress Scale (ARSS). All participants completed the 32-item list of the ARSS that resulted from the translation process and initial analyses by Nässi et al. (2017a). Each expression (describing a different state of recovery and stress, e.g., “rested”, “tired”) is answered on a Likert-type rating scale ranging from 0 (*does not apply at all*) to 6 (*fully applies*) to indicate the extend that most closely applies to the individual ‘right now’. Four adjectives each are then grouped as mean score. By this means, eight scales can be calculated representing the *Recovery* dimension with *Physical Performance Capability*, *Mental Performance Capability*, *Emotional Balance*, *Overall Recovery,* and the *Stress* dimension with *Muscular Stress*, *Lack of Activation*, *Negative Emotional State,* and *Overall Stress*. Acceptable to good scale homogeneity with Cronbach’s α ranging from .76 to .86 has been reported by Nässi et al. (2017a), which also resembles findings of the German original (α = .76 to .90; Hitzschke et al., 2016).

Short Recovery and Stress Scale (SRSS). The SRSS was answered by 865 participants. Following the original by Kellmann et al. (2016), the eight scales of the ARSS are rated as single items on the same 7-point rating scale ranging from 0 (*does not apply at all*) to 6 (*fully applies*) in relation to their highest ever state. The corresponding adjectives served as descriptors for each item. In the case of *Emotional Balance,* six adjectives were displayed (see below). While the *Short Recovery Scale* is represented by the items *Physical Performance Capability*, *Mental Performance Capability*, *Emotional Balance*, and *Overall Recovery,* and the *Short Stress Scale* is represented by the items *Muscular Stress*, *Lack of Activation*, *Negative Emotional State*, and *Overall Stress,* there is no further aggregation to mean or overall scores. The German *Short Recovery Scale* (α = .70) and the *Short Stress Scale* (α = .76; Hitzschke et al., 2015) as well as the first English version of the *Short Recovery Scale* (α = .74) and the *Short*

Stress Scale (α = .78; Nässi et al., 2017a) demonstrated acceptable scale homogeneity. For copies of the questionnaires as well as detailed descriptions of the translation and validation process of the English ARSS and SRSS see Kellmann and Kölling (2019).

Recovery-Stress Questionnaire for Athletes (RESTQ-Sport-76). In order to assess the construct and convergent validity of both questionnaires, 530 participants answered the RESTQ-Sport-76 by Kellmann and Kallus (2016) in addition to the ARSS and SRSS. Regarding the past three days and nights, 76 statements are rated in terms of the frequency of stress perceptions and recovery activities on a Likert-type rating scale ranging from 0 (*never*) to 6 (*always*). These are then summarised to 19 non-sport and sport-specific scales. Information regarding validity and reliability can be found in the manual (Kellmann & Kallus, 2016).

Procedure

In advance of the data collection, the procedure was approved by the local ethics committee. The participants were recruited in the different regions via online links by sport institutions, clubs, and sport associations in the above indicated regions. Before starting the survey, participants were informed about the purpose of the study as well as the voluntary nature and anonymity of participation. After actively providing their consent, athletes answered the questionnaires. In order to perform separate analyses for each language region, the data was collected until the target of 300 completed questionnaires was reached in each group. Considering required sample sizes to achieve robust estimations of the models for the confirmatory factor analyses (see below), 160 participants in each subsample would have been sufficient based on a ratio of 5:1 for each parameter to be estimated (Kline, 2011). The higher number was chosen in the light of representativeness, thus aiming to reach a diverse database that covers different types of sport as well as fairly equal age and gender distributions. To ensure minimal missing data in the ARSS items, this questionnaire was answered first followed by the SRSS and the RESTQ-Sport.

Data analysis

The statistical analyses were carried out with IBM's statistical package SPSS Version 25. On the item level, item-total correlations were calculated to determine the discriminatory power (r_{it}). Cronbach's α was used to analyse the internal consistency of each ARSS scale and of the *Short Recovery Scale* as well as the *Short Stress Scale*. Spearman correlations were applied to assess the inter-correlation within the ARSS and the SRSS and between. Spearman correlations were also used to test convergent validity with the 19 scales of the RESTQ-Sport-76. Results of the inter-correlations and correlations with the RESTQ-Sport-76 will be reported for the total sample only, while results for each subsample can be found in Kellmann and Kölling (2019). Confirmatory factor analysis (CFA) was performed with R (Lavaan package version 0.6-3 by Rosseel, 2012; semTools package version 0.5-1 by Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2018) based on linear structural equation modelling. To deal with the lack of multivariate normal distribution, robust maximum likelihood estimators were applied. In a first step, three different CFAs were performed separately for the *Recovery* and *Stress* dimension, respectively: (a) in the first-order CFA, four items each are assessed by one of the four latent factors between which covariance relations are specified; (b) in the hierarchical CFA, a higher order latent factor (i.e., *Recovery*, *Stress*) is related to the four latent factors and no residual correlations are specified between the four factors; (c) in the bifactor CFA, all items simultaneously load on one general factor (i.e., *Recovery*, *Stress*) and on one of the four corresponding factors. Based on theoretical considerations and modification indices, the final model, in a second step, was assessed and used for further analyses. The final *Recovery* and *Stress* models were then analysed in each group. The following fit indices were evaluated regarding the critical values as recommended by Beauducel and Wittmann (2005), Byrne

(2001) as well as Hu and Bentler (1999): χ^2 and the corresponding *p*-value, Comparative Fit Index (CFI, > .90, Beauducel & Wittmann, 2005), Standardized Root Mean Residual (SRMR, < .10, Beauducel & Wittmann, 2005), Root Mean Square Error of Approximation (RMSEA, < .08, Hu & Bentler, 1999), with its Lower Limit of the 90 %-Confidence Interval (LO90) and Upper Limit of the 90 %-Confidence Interval (HI90). In order to analyse whether the structure of the measurement model is invariant across the groups (i.e., whether the ARSS assesses the same construct), multigroup CFA was conducted (Cheung & Rensvold, 2002). Invariance was then tested in terms of (a) configural invariance (i.e., the construct is measured by the same number of factors and associations of items in all groups), (b) metric invariance (weak invariance; i.e., factor loadings are equal across groups), and (c) scalar invariance (strong invariance; i.e., observed indicators show equal intercepts when regressed on the latent factors). Measurement invariance was evaluated by change of the fit indices following recommendations by Cheung and Rensvold (2002) for changes of CFI (ΔCFI ≤ -.01) and by Chen (2007) for changes of RMSEA (ΔRMSEA < .015) and SRMR (ΔSRMR < .010), as both papers advice against relying on the χ^2-Difference Test.

Results

As a result of the first validation study by Nässi et al. (2017a), the scale *Emotional Balance* needed further modification. In consideration of different language styles and cultural contexts, two more items were added to strengthen the solidity and practical comprehensiveness of *Emotional Balance*. Before running the analyses, a preliminary step was to make a final decision of the most suitable four items. Based on the CFA statistics, the adjectives “satisfied” and “balanced” were replaced by “pleased” and “stable”, respectively. These were then included in the following analyses[1].

[1] Please note that the six adjectives served as descriptors for the SRSS item *Emotional Balance*, as the data was collected simultaneously.

Tables 2 and 3 show the descriptive statistics (*M* ± *SD*), item-total correlations, and Cronbach's α for the ARSS and the SRSS for the total sample as well as for each subsample (i.e., ANZ, UK, NA), respectively. Satisfactory discriminatory power (r_{it} > .30) was found for the total sample (ARSS r_{it} = .54 - .78, SRSS r_{it} = .34 - .76) and within the subsamples. Cronbach's α was also ranging between acceptable and good values for the total sample within the ARSS (α = .78 - .88) as well as for the *Short Recovery Scale* (α = .84) and the *Short Stress Scale* (α = .78), which is also reflected in the subsamples (for ARSS see Table 2). Specifically, the *Short Recovery Scale* showed good internal consistency in the subsamples (ANZ α = .84, UK α = .81, NA α = .87), while values of the *Short Stress Scale* were slightly lower (ANZ α = .76, UK α = .76, NA α = .83). Overall, item mean (ARSS *M* = 1.50 - 4.35, SRSS *M* = 2.10 - 3.98) and item dispersion (ARSS *SD* = 1.30 - 1.80, SRSS *SD* = 1.23 - 1.64) were satisfactory for the total sample. On a descriptive level, the North American subsample showed the highest recovery and the lowest stress ratings as well as slightly higher item-total correlations and Cronbach's α (Table 2 for ARSS, Table 3 for SRSS).

Table 4 depicts the correlations within the ARSS scales and within the SRSS items as well as the correlation coefficients between the corresponding scales/items of the two questionnaires. Within the *Recovery* dimension of the ARSS, Spearman's rho (r_s) ranged between .50 and .72, and between .29 and .71 within the *Stress* dimension. Between these dimensions, correlations ranged between r_s = -.14 and -.64. A similar pattern was found for the SRSS item inter-correlations, although the coefficients were slightly weaker ranging from r_s = .51 to .66 for the *Short Recovery Scale* and from r_s = .21 to .65 for the *Short Stress Scale*. Between both scales correlations ranged between r_s = -.12 and -.66. Furthermore, correlations between corresponding ARSS scales and SRSS items were moderate to high (r_s = .68 - .78).

Results of the CFA are displayed in Table 5. In a first step, alternative models to the first-order CFA were calculated (i.e., hierarchical & bifactor models). For the *Recovery* dimension, satisfactory fit indices were found for all three models in the total sample. However, as the bifactor model indicated a Heywood case for ANZ, and a solution could not be found for the hierarchical model in NA (results not presented here), the first-order model was retained for further analyses. Furthermore, for the *Stress* dimension, only the first-order model converged in the total sample (and only the hierarchical model could be calculated for UK and NA revealing poor model fit). Thus, the first-order model was also retained for further analyses among the *Stress* dimension. In a second step, two modifications were approved through covariations of the measurement errors between "strong" and "physically capable" of *Physical Performance Capability* as well as between "attentive" and "receptive" of *Mental Performance Capability*, which were also applied for the subsamples (see final model in Table 5). The model fit for the *Stress* dimension was improved by two modifications among *Muscular Stress* (i.e., covariations of the measurement errors between "muscle exhaustion" and "muscle fatigue" as well as between "muscle soreness" and "muscle stiffness") and among *Lack of Activation* (i.e., covariations of the measurement errors between "unmotivated" and "unenthusiastic"). The final *Recovery* as well as the *Stress* model showed good fit for each subsample, although the RMSEA was just outside the threshold in the North American subsample.

Table 6 presents the results of the multigroup CFAs between male and female participants as well as between the regional subsamples. Strong measurement invariance was present in both group comparisons for the *Recovery* as well as the *Stress* dimension. Thus, the form of the models as well as the factor loadings are invariant across groups, and prerequisites for latent mean comparisons are fulfilled.

Table 7 presents the correlations between RESTQ-Sport-76 and the ARSS and SRSS for the total sample. Considering hypothesis-relevant relations between the different questionnaires (i.e., positive correlations with similar dimensions and negative correlations with opposite dimensions), a congruent pattern for the ARSS as well as the SRSS was found. For example, the ARSS's scale *Physical Performance Capability* showed higher correlation coefficients with the RESTQ-Sport-76 scales *Physical Recovery* (r_s = .63) and *Being in Shape* (r_s = .67), whereas

ARSS's *Muscular Stress* showed almost no correlation with *Personal Accomplishment* (r_s = -.03) or *Self-Regulation* (r_s = .01). Overall, correlations were small to moderate, while coefficients with the SRSS were consistently lower. With the exception of three correlations, coefficients did not exceed values above .7 (i.e., ARSS *Emotional Balance* with *General Well-being* [r_s = .72], ARSS *Negative Emotional State* with *General Stress* [r_s = .70] and *Emotional Stress* [r_s = .74]).

Discussion

The aim of the present study was to validate the English version of two established questionnaires which assess the current recovery-stress state of athletes. Following initial analyses by Nässi et al. (2017a), further evidence was now provided via reliability and confirmatory factor analyses among a large sample of athletes as well as for subsamples representing three common English-speaking regions (i.e., Australia/New Zealand, the United Kingdom, North America).

Overall, reliability of the ARSS and the SRSS was confirmed via high discriminatory power on the item level as well as via internal consistency on the scale level. These results were consistent not only for the total sample but also for each subsample. Furthermore, these results strengthen the initial validation study (Nässi et al., 2017a) and they are comparable to the German original questionnaires (Kellmann et al., 2016). Regarding the factorial structure of the ARSS, results of the CFA are also in line with previous findings. The RMSEA's threshold of the North American sample was the only fit index above the recommended .08 in both models. However, suggested cut-off criteria are frequently discussed, so that they should be considered as general guidelines rather than as definite rules (Worthington & Whittaker, 2006). Further, analyses of measurement invariance indicated that (a) the same factorial structure, (b) the same item loadings, and (c) the same item intercepts can be assumed across the subsamples as well as between the genders. In terms of the state-oriented measure, (a) and (b) can be considered as the relevant aspects of measurement invariance. Thus, the general applicability can be assumed for all language regions. Moreover, concerns regarding the construct validity among the North American subsample which have been raised by Kellmann and Kölling (2019) can be considered as overcome. This is also supported by findings of the descriptive statistics and reliability analyses, as these were stronger for the North American subsample. Kellmann and Kölling (2019) reported test-retest analyses and multiple Cronbach's α calculations for repeated measurements within an intervention study that applied the ARSS in a subgroup of North American athletes (n = 109). Increased internal consistency values indicate that participants become accustomed to the questionnaire over time which supports an introductory measurement when applying the ARSS in practice. In addition, correlations between repeated measures are stronger for comparable starting conditions and short time intervals (Kellmann & Kölling, 2019).

Correlational analyses with the RESTQ-Sport-76 underline the construct and convergent validity of the ARSS and the SRSS, as comparable patterns to those of Nässi et al. (2017a) as well as Kellmann et al. (2016) were found. The slightly smaller coefficients for the SRSS items can probably be explained with the answering mode. The RESTQ-Sport-76 scales are generated through a mean value which leads to more numerical increments, while the SRSS is judged on a single rating scale so that only integers will be analysed. The ARSS scales, on the other hand, are generated by mean scores and therefore the same rule applies as for the RESTQ-Sport-76.

At the same time, the difference in the gradation of the scales is consequently also present between the ARSS and the SRSS. Therefore, the ARSS provides more detailed information of the current recovery-stress state of a person, while the SRSS surpasses with its briefness and economical application. As the correlations between the instruments did not exceed .78, this supports their content-wise correspondence on the one hand and their uniqueness on the other hand. Depending on the research question or the application purpose, the ARSS and the SRSS can be considered as independent as well as supplementary monitoring tools. In accordance with the manuals, it is recommended to apply the ARSS before the first use of the SRSS in order to familiarise the participant with the construct (Kellmann et al., 2016; Kellmann & Kölling, 2019). It seems worthwhile, though, to investigate the applicability of the SRSS without prior familiarisation. Due to the combination of the SRSS's shortness and multidimensionality, this tool specifically meets requirements for frequent assessments in applied sport scientific and coaching settings (Horvath & Röthlin, 2018). For example, using the briefer SRSS may help avoid the regression to the mean in long-term measurements. However, the compliance of the athletes needs to be further ensured by explaining to them the purpose and benefits of (psychometric) monitoring and the confidentiality of their data handling (Kellmann & Beckmann, 2003; Saw, Main, & Gastin, 2015a).

In general, there is a broad consensus regarding the necessity to validate psychometric assessments and tests which have been translated and adapted to cultural contexts and practical guidelines for this process are available (Gudmundsson, 2009; International Test Commission, 2018; Lenz, Gómez Soler, Dell'Aquilla, & Martinez Uribe, 2017). Following those guidelines, the procedure for both ARSS and SRSS has been documented by Nässi et al. (2017a) and Kellmann and Kölling (2019). Although the preparation and statistical validation processes were meticulously planned accordingly, another issue in terms of cross-cultural differences within the English language became apparent. Due to many varieties of English (e.g., variations in UK, ANZ, North America, Asia, and Africa; Kachru et al., 2006), it seems reasonable to consider these within the validation process and to analyse respective psychometric properties. Thus, the composition of the North American subsample merits critical discussion. In this subsample, U.S. American as well as Canadian participants were combined based on practicability considerations. Because of the regional vicinity, the common understanding of the items was assumed to be closer related within athletes of the same continent. Thus, discrepancies based on the British versus American English spelling cannot be ruled out and measurement invariance should be tested among those different populations. Moreover, the competition levels were differently distributed between the subsamples. ANZ had the highest amount of international competing athletes, while these were underrepresented in the NA subsample.

The ARSS and SRSS were developed to assess different dimensions of recovery and stress that represent the acute state of an athlete. According to the 'scissors' model', states of recovery and stress need to be considered on a continuum that is particularly individual (Kellmann, 2002). In view of that, the scales should be able to reflect the dynamic processes and the interaction between them in response to the individual circumstances and environmental demands. During the monitoring process, individual and situation-specific ranges of adequate recovery-stress states may be identified to support the training and regeneration management. In that sense, parallels to Hanin's (2000) *Individual Zones of Optimal Functioning* become apparent when addressing the optimisation of the recovery-stress balance. Furthermore, as Meeusen et al. (2013) differentiate between functional and non-functional overreaching and the gradual development of the OTS, continuing monitoring of the mood state is of considerate importance. At this point, it seems noteworthy to underline that the labels *Emotional Balance* and *Negative Emotional State* actually represent the global evaluation of the individual's current mood (e.g., "feeling down"). Since emotions and mood are conceptually different, the broader sensations (i.e., mood) are assessed, while these may follow an emotion that is caused by a specific situation or object (Lane & Terry, 2000). Consequently, the English ARSS and the SRSS still need to be applied in long-term monitoring research and longitudinal studies. The sensitivity to change of the scales that is documented for the original (Kellmann et al., 2016) requires replication. Also, the underlying constructs should be measured in different situations of stress and recovery in sport to analyse trans-situational stability. In addition, the accumulation of stress states and simultaneous compensation through adequate recovery, as depicted in the 'scissors' model' (Kellmann, 2002), may be investigated in future studies by assessing performance and physiological markers that have been identified as surrogate parameters of athletes' need for recovery. Moreover, other aspects of recovery such as periodization (Mujika, Halson, Burke, Balagué, & Farrow, 2018) or the implementation in stress management interventions (Rumbold, Fletcher, & Kevin, 2012) may be explored in the future. Differential perceptions of recovery and stress may be further examined via mixed methods designs applying the approach described by Rumbold, Fletcher, and Daniels (2018) from the organisational perspective.

Conclusion

Taken together, the present study provides valuable scientific innovation among monitoring research as well as applied sport science. In accordance with the German original, the English ARSS and SRSS emerged as two valid and reliable tools which were developed on a sound theoretical background and are suitable for economical applications in individual as well as group settings. The quality criteria of psychometric test construction were replicated among a large sample of the target population. The unique approach of shortening the ARSS to the SRSS is now generalised to a broader athletic population to address the necessity of providing short scales to increase athletes' compliance (Horvarth & Röthlin, 2018; Saw et al., 2015a, 2017). Moreover, to the best of our knowledge, this was the first study within sport psychology and sport science that explicitly considered different English-speaking regions in the validation process and provides proof of strong measurement invariance.

Future research can now build upon these findings to establish the ARSS and SRSS in English-speaking regions worldwide. For example, focus should be laid on the investigation of (a) their applicability in field settings (e.g., training monitoring during preparation phases), (b) the sensitivity to change in laboratory studies (e.g., recovery interventions), and (c) on testing and probably also on adapting the tools among adolescent athletes.

References

Andreato, L. V., de Moraes, S. M. F., Del Conti Esteves, J. V., Miranda, M. L., Pastório, J. J., Pastório, E. J., . . . Franchini, E. (2014). Psychological, physiological, performance and perceptive responses to Brazilian jiu-jitsu combats. *Kinesiology, 46*(1), 44-52.

Arnold, R., Fletcher, D., & Daniels, K. (2013). Development and validation of the Organizational Stressor Indicator for Sport Performers (OSI-SP). *Journal of Sport & Exercise Psychology, 35*, 180-196. doi:10.1123/jsep.35.2.180

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*(1), 41-75.

Beedie, C. J., Terry, P. C., & Lane, A. M. (2000). The Profile of Mood States and atheltic performance: Two meta-analyses. *Journal of Applied Sport Psychology, 12*, 49-68.

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P. B., Kellmann, M., Varley, M. C., . . . Cable, N. T. (2017). Monitoring athlete training loads: Consensus statement. *International Journal of Sports Physiology and Performance, 12*(S2), 161-170. doi: 10.1123/IJSPP.2017-0208

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS. Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.

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

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 385-396. doi:10.2307/2136404

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., . . . Dodge, C. (2001). A new approach to monitoring exercise training. *Journal of Strength Conditioning Research, 15*, 109-115.

Foster, C., Rodriguez-Marroyo, J. A., & de Koning, J. J. (2017). Monitoring training loads: The past, the present, and the future. *International Journal of Sports Physiology and Performance*, *12*(S2), 2-8. doi:10.1123/IJSPP.2016-0388

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology, 61*(2), 29-45. doi: 10.1027/1901-2276.61.2.29

Hanin, Y. L. (2000). Individual Zones of Optimal Functioning model: Emotion-performance relationships in sport. In Y. L. Haning (Ed.), *Emotions in sport* (pp. 65-89). Champaign, IL: Human Kinetics.

Hitzschke, B., Holst, T., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2016). Entwicklung des Akutmaßes zur Erfassung von Erholung und Beanspruchung im Sport [Development of the Acute Recovery and Stress Scale]. *Diagnostica, 62*, 212-226. doi: 10.1026/0012-1924/a000155

Hitzschke, B., Kölling, S., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2015). Entwicklung der Kurzskala zur Erfassung von Erholung und Beanspruchung im Sport (KEB) [Development of the Short Recovery and Stress Scale for sports (SRSS)]. *Zeitschrift für Sportpsychologie, 22*, 146-161. doi: 10.1026/1612-5010/a000150

Hitzschke, B., Wiewelhove, T., Raeder, C., Ferrauti, A., Meyer, T., Pfeiffer, M., . . . Kölling, S. (2017). Evaluation of psychological measures for the assessment of recovery and stress during a shock-microcycle in strength and high-intensity interval training. *Performance Enhancement & Health, 5*, 147-157. doi: 10.1016/j.peh.2017.08.001

Horvath, S., & Röthlin, P. (2018). How to improve athletes' return of investment: Shortening questionnaires in the applied sport psychology setting. *Journal of Applied Sport Psychology, 30*, 241-248. doi: 10.1080/10413200.2017.1382020

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-1. Retrieved from https://CRAN.R-project.org/package=semTools

International Test Commission, ITC. (2018). ITC guidelines for translating and adapting tests (second edition). *International Journal of Testing, 18*, 101-134. doi: 10.1080/15305058.2017.1398166

Kachru, B. B., Kachru, Y., & Nelson, C. L. (Eds.). (2006). *The handbook of world Englishes*. Oxford: Blackwell Publishing.

Kallus, K. W. (2016). Stress and recovery: An overview. In K. W. Kallus & M. Kellmann (Eds.), *The Recovery Stress Questionnaires: User manual* (pp. 27-48). Frankfurt a.M.: Pearson Assessment & Information GmbH.

Kellmann, M. (2002). Underrecovery and overtraining: Different concepts - similar impact? In M. Kellmann (Ed.), *Enhancing recovery: Preventing underperformance in athletes* (pp. 3-24). Champaign, IL: Human Kinetics.

Kellmann, M., & Beckmann, J. (2003). Research and intervention in sport psychology: New perspectives on an inherent conflict. *International Journal of Sport and Exercise Psychology, 1*, 13-26.

Kellmann, M., Bertollo, M., Bosquet, L., Brink, M. S., Coutts, A. J., Duffield, R., . . . Beckmann, J. (2018). Recovery and performance in sport: Consensus statement. *International Journal of Sports Physiology and Performance, 13*(2), 240-245. doi: 10.1123/ijspp.2017-0759

Kellmann, M., & Kallus, K. W. (2016). The Recovery-Stress Questionnaire for Athletes. In K. W. Kallus & M. Kellmann (Eds.), *The Recovery-Stress Questionnaires: User manual* (pp. 89-134). Frankfurt: Pearson.

Kellmann, M., & Kölling, S. (2019). *Recovery and stress in sport: A manual for testing and assessment*. Abingdon: Routledge.

Kellmann, M., Kölling, S., & Hitzschke, B. (2016). *Das Akutmaß und die Kurzskala zur Erfassung von Erholung und Beanspruchung im Sport - Manual [The Acute and the Short Recovery and Stress Scale for Sports - manual]*. Hellenthal: Sportverlag Strauß.

Kenttä, G., & Hassmén, P. (1998). Overtraining and recovery. A conceptual model. *Sports Medicine, 26*, 1-16.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kölling, S., Hitzschke, B., Holst, T., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2015). Validity of the Acute Recovery and Stress Scale - Training monitoring of the German Junior National Field Hockey Team. *International Journal of Sport Science and Coaching, 10*, 529-542.

Lane, A. M., & Terry, P. C. (2000). The nature of mood: Development of a conceptual model with a focus on depression. *Journal of Applied Sport Psychology, 12*(1), 16-33. doi:10.1080/10413200008404211

Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer.

Lenz, A. S., Gómez Soler, I., Dell'Aquilla, J., & Martinez Uribe, P. (2017). Translation and cross-cultural adaptation of assessments for use in counseling research. *Measurement and Evaluation in Counseling and Development, 50*, 224-231. doi: 10.1080/07481756.2017.1320947

Mäetsu, J., Jürimäe, J., & Jürimäe, T. (2005). Monitoring of performance and training in rowing. *Sports Medicine, 35*, 597-617.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Revised manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Meeusen, R., & De Pauw, K. (2018). Overtraining - what do we know? In M. Kellmann & J. Beckmann (Eds.), *Sport, recovery, and performance. Interdisciplinary insights* (pp. 51-62). Abingdon: Routledge.

Meeusen, R., Duclos, M., Foster, C., Fry, A., Glesson, M., Nieman, D., . . . Urhausen, A. (2013). Prevention, diagnosis and treatment of the overtraining syndrome: Joint consensus statement of the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM). *Medicine and Science in Sports and Exercise, 45*, 186-205.

Mujika, I., Halson, S. L., Burke, L. M., Balagué, G., & Farrow, D. (2018). An integrated, multifactorial approach to periodization for optimal performance in individual and team sports. *International Journal of Sports Physiology and Performance, 13*, 538-561. doi:10.1123/ijspp.2018-0093

Nässi, A., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017a). Development of two short measures for recovery and stress in sport. *European Journal of Sport Science, 17*, 894-903. doi:10.1080/17461391.2017.1318180

Nässi, A., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017b). Psychological tools used for monitoring training responses of athletes. *Performance Enhancement & Health, 5*, 125-133. doi:10.1016/j.peh.2017.05.001

Pelka, M., Kölling, S., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017). Acute effects of psychological relaxation techniques between two physical tasks. *Journal of Sports Sciences, 35*, 216-223. doi: 10.1080/02640414.2016.1161208

Rumbold, J. L., Fletcher, D., & Daniels, K. (2018). Using a mixed method audit to inform organizational stress management interventions in sport. *Psychology of Sport and Exercise, 35*, 27-38. doi:10.1016/j.psychsport.2017.10.010

Rumbold, J. L., Fletcher, D., & Kevin, D. (2012). A systematic review of stress management interventions with sport performers. *Sport, Exercise, and Performance Psychology, 1*, 173-193. doi:10.1037/a0026628

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2). doi:10.18637/jss.v048.i02

Saw, A. E., Kellmann, M., Main, L. C., & Gastin, P. B. (2017). Athlete self-report measures in research an practice: Considerations for the discerning reader and fastidious practioner. *International Journal of Sports Physiology and Performance, 12*(Suppl. 2), S2127-S2135.

Saw, A. E., Main, L. C., & Gastin, P. B. (2015a). Monitoring athletes through self-report: Factors influencing implementation. *Journal of Sports Science and Medicine, 14*, 137-146.

Saw, A. E., Main, L. C., & Gastin, P. B. (2015b). Role of self-report measure in athlete preparation. *Journal of Strength Conditioning Research, 29*, 685-691.

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-report measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine, 50*, 281-291. doi: 10.1136/bjsports-2015-094758

Trudgill, J., & Hannah, J. (Eds.). (2017). *International English. A guide to varieties of English around the world* (6th ed.). London: Routledge.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research. A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806-838. doi:10.1177/0011000006288127

Table 1. Description of the subsamples.

	ANZ	UK	NA
Number (*n*)	380	316	300
Male/female (n [%])	213 (56.1%)/167 (43.9%)	158 (50.0%)/158 (50.0%)	165 (55.0%)/135 (45.0%)
Age (years)	25.5 ± 9.9	26.0 ± 9.8	23.0 ± 6.5
Competition level			
International level	180 (47.4%)	91 (28.8%)	49 (16.3%)
National level	121 (31.8%)	77 (24.4%)	165 (55.0%)
Regional level	44 (11.6%)	62 (19.6%)	33 (11.0%)
Collegiate athletes	2 (0.5%)	23 (7.3%)	26 (8.7%)
Recreational athletes	33 (8.7%)	63 (19.9%)	27 (9.0%)

Table 2. Descriptive statistics and item-total correlations of the ARSS for the total sample and subsamples

		Total (N = 996)				ANZ (n = 379)				UK (n = 316)				NA (n = 300)			
		M	*SD*	r_{it}	α	*M*	*SD*	r_{it}	α	*M*	*SD*	r_{it}	α	*M*	*SD*	r_{it}	α
Recovery Dimension	**PPC**	**3.80**	**1.14**		.85	**3.78**	**1.02**		.82	**3.58**	**1.14**		.84	**4.06**	**1.22**		.87
	PPC item 1	4.02	1.30	.71		4.03	1.22	.61		3.85	1.33	.71		4.20	1.35	.79	
	PPC item 2	4.35	1.35	.63		4.34	1.22	.62		4.13	1.43	.59		4.61	1.38	.65	
	PPC item 3	3.52	1.38	.68		3.44	1.26	.63		3.29	1.39	.66		3.88	1.46	.72	
	PPC item 4	3.31	1.44	.74		3.34	1.33	.73		3.05	1.43	.72		3.55	1.55	.76	
	MPC	**3.85**	**1.11**		.84	**3.82**	**1.01**		.81	**3.56**	**1.10**		.80	**4.22**	**1.16**		.87
	MPC item 1	3.93	1.38	.63		3.86	1.27	.58		3.65	1.42	.56		4.30	1.41	.70	
	MPC item 2	3.79	1.31	.65		3.73	1.22	.60		3.54	1.31	.62		4.12	1.35	.68	
	MPC item 3	3.95	1.36	.69		3.98	1.25	.66		3.58	1.41	.63		4.32	1.35	.76	
	MPC item 4	3.75	1.37	.71		3.71	1.28	.69		3.46	1.41	.64		4.13	1.34	.78	
	EB	**3.91**	**1.20**		.87	**3.98**	**1.07**		.83	**3.64**	**1.26**		.87	**4.10**	**1.25**		.88
	EB item 1	4.02	1.42	.77		4.12	1.32	.69		3.80	1.48	.79		4.11	1.48	.81	
	EB item 2	3.95	1.36	.72		3.98	1.27	.65		3.66	1.39	.73		4.21	1.40	.75	
	EB item 3	4.22	1.37	.73		4.23	1.26	.69		3.95	1.48	.74		4.47	1.32	.75	
	EB item 4	3.46	1.54	.66		3.59	1.41	.63		3.16	1.60	.65		3.61	1.58	.69	
	OR	**3.31**	**1.21**		.82	**3.35**	**1.10**		.80	**3.10**	**1.22**		.83	**3.48**	**1.31**		.83
	OR item 1	3.86	1.52	.58		3.83	1.37	.58		3.71	1.59	.61		4.05	1.61	.56	
	OR item 2	3.09	1.55	.64		3.11	1.45	.59		2.86	1.54	.67		3.30	1.65	.64	
	OR item 3	2.98	1.47	.68		3.08	1.40	.65		2.77	1.43	.69		3.10	1.59	.68	
	OR item 4	3.30	1.48	.67		3.37	1.36	.60		3.05	1.45	.67		3.47	1.62	.73	
Stress Dimension	**MS**	**2.62**	**1.42**		.88	**2.80**	**1.34**		.87	**2.62**	**1.37**		.86	**2.39**	**1.54**		.90
	MS item 1	2.58	1.67	.69		2.72	1.54	.67		2.58	1.66	.64		2.39	1.81	.75	
	MS item 2	2.66	1.64	.78		2.88	1.57	.78		2.66	1.61	.72		2.37	1.71	.83	
	MS item 3	2.58	1.70	.78		2.79	1.59	.76		2.54	1.69	.76		2.36	1.81	.82	
	MS item 4	2.66	1.64	.69		2.82	1.60	.71		2.69	1.60	.69		2.42	1.72	.67	
	LA	**2.03**	**1.30**		.81	**2.08**	**1.22**		.77	**2.18**	**1.32**		.81	**1.81**	**1.35**		.85
	LA item 1	1.58	1.64	.59		1.57	1.59	.54		1.77	1.69	.56		1.39	1.65	.66	
	LA item 2	2.33	1.62	.63		2.43	1.61	.60		2.44	1.63	.61		2.08	1.59	.68	
	LA item 3	1.80	1.63	.66		1.86	1.59	.60		1.94	1.67	.69		1.58	1.62	.69	
	LA item 4	2.41	1.63	.63		2.46	1.55	.56		2.57	1.65	.63		2.18	1.67	.71	
	NES	**1.89**	**1.28**		.78	**1.86**	**1.17**		.75	**2.13**	**1.36**		.78	**1.67**	**1.29**		.80
	NES item 1	1.50	1.60	.59		1.45	1.49	.56		1.80	1.73	.59		1.26	1.55	.61	
	NES item 2	2.54	1.80	.58		2.45	1.70	.53		2.85	1.87	.61		2.33	1.81	.59	
	NES item 3	1.62	1.51	.63		1.64	1.40	.63		1.76	1.61	.58		1.46	1.51	.67	
	NES item 4	1.89	1.69	.54		1.90	1.56	.48		2.13	1.84	.54		1.63	1.65	.57	
	OS	**2.55**	**1.40**		.86	**2.61**	**1.28**		.83	**2.64**	**1.39**		.86	**2.37**	**1.52**		.89
	OS item 1	3.29	1.71	.68		3.41	1.59	.65		3.34	1.69	.65		3.09	1.86	.74	
	OS item 2	2.36	1.66	.77		2.45	1.62	.74		2.44	1.64	.75		2.17	1.70	.81	
	OS item 3	2.27	1.63	.64		2.28	1.51	.53		2.47	1.67	.71		2.05	1.71	.69	
	OS item 4	2.24	1.63	.75		2.27	1.59	.70		2.30	1.58	.74		2.14	1.73	.83	

Note: ARSS = Acute Recovery and Stress Scale; PPC = *Physical Performance Capability*; MPC = *Mental Performance Capability*; EB = *Emotional Balance*; OR = *Overall Recovery*; MS = *Muscular Stress*; LA = *Lack of Activation*; NES = *Negative Emotional State*; OS = *Overall Stress*.

Table 3. Descriptive statistics and item-total correlations of the SRSS for the total sample and subsamples

		Total (N = 865)			ANZ (n = 368)			UK (n = 304)			NA (n = 186)		
		M	***SD***	r_{it}	***M***	***SD***	r_{it}	***M***	***SD***	r_{it}	***M***	***SD***	r_{it}
Short Recovery Scale	PPC	3.90	1.23	.70	3.88	1.17	.69	3.76	1.17	.63	4.16	1.39	.77
	MPC	3.98	1.27	.73	3.89	1.23	.72	3.81	1.29	.69	4.42	1.25	.77
	EB	3.88	1.36	.67	3.92	1.28	.67	3.62	1.39	.60	4.20	1.37	.73
	OR	3.65	1.27	.64	3.69	1.23	.62	3.45	1.22	.63	3.89	1.37	.66
Short Stress Scale	MS	2.76	1.56	.34	2.84	1.52	.32	2.80	1.49	.31	2.55	1.72	.41
	LA	2.23	1.51	.66	2.29	1.50	.60	2.34	1.46	.67	1.90	1.57	.76
	NES	2.10	1.64	.61	2.05	1.58	.61	2.33	1.65	.56	1.84	1.70	.67
	OS	2.51	1.53	.76	2.48	1.49	.73	2.65	1.41	.75	2.36	1.76	.80

Note: SRSS = Short Recovery and Stress Scale; PPC = *Physical Performance Capability*; MPC = *Mental Performance Capability*; EB = *Emotional Balance*; OR = *Overall Recovery*; MS = *Muscular Stress*; LA = *Lack of Activation*; NES = *Negative Emotional State*; OS = *Overall Stress*.

Table 4. Spearman correlations within the ARSS scales (N = 996) and the SRSS items (N = 865) and between corresponding scales/items

Scale/Item	Between ARSS & SRSS	upper data matrix: ARSS							
		PPC	MPC	EB	OR	MS	LA	NES	OS
Physical Performance Capability	**.70**		**.72**	**.70**	.59	-.24	-.60	-.43	-.50
Mental Performance Capability	.68	.64		.69	.50	-.16	-.56	-.43	-.43
Emotional Balance	**.75**	.52	.66		.50	-.14	-.56	-.64	-.47
Overall Recovery	**.71**	.62	.52	.51		-.54	-.41	-.35	-.57
Muscular Stress	**.72**	-.28	-.14	-.12	-.44		.36	.29	.61
Lack of Activation	**.73**	-.49	-.53	-.48	-.40	.29		.67	**.71**
Negative Emotional State	**.78**	-.35	-.50	-.66	-.34	.21	.63		.62
Overall Stress	**.74**	-.46	-.49	-.52	-.49	.44	.65	.65	
		lower data matrix: SRSS							

Note: ARSS = Acute Recovery and Stress Scale; SRSS = Short Recovery and Stress Scale; PPC = *Physical Performance Capability*; MPC = *Mental Performance Capability*; EB = *Emotional Balance*; OR = *Overall Recovery*; MS = *Muscular Stress*; LA = *Lack of Activation*; NES = *Negative Emotional State*; OS = *Overall Stress*; all correlations are significant with $p < .001$; $r_s \geq .70$ are bolded.

Table 5. Results of the confirmatory factor analysis of the ARSS for the total sample and subsamples

		χ^2	*df*	*p*	CFI	SRMR	RMSEA	LO90	HI90
Recovery dimension									
Total sample (*N* = 993)	First-order model	453.45	98	< .001	.949	.038	.069	.063	.076
	Hierarchical model	464.77	100	< .001	.947	.040	.070	.063	.076
	Bifactor model	317.31	88	< .001	.968	.031	.058	.051	.065
	Final model	393.02	96	< .001	.958	.036	.064	.057	.070
ANZ (*n* = 379)	Final model	212.80	96	< .001	.951	.048	.062	.051	.074
UK (*n* = 316)	Final model	195.89	96	< .001	.955	.046	.065	.052	.078
NA (*n* = 298)	Final model	248.36	96	< .001	.942	.046	.081	.069	.094
Male sample (*n* = 535)	Final model	230.15	96	< .001	.964	.036	.059	.050	.069
Female sample (*n* = 458)	Final model	259.09	96	< .001	.949	.045	.067	.057	.077
Stress dimension									
Total sample (*N* = 993)	First-order model	606.30	98	< .001	.924	.055	.083	.077	.090
	Hierarchical model	n.a.	-/-	-/-	n.a.	n.a.	n.a.	-/-	-/-
	Bifactor model	n.a.	-/-	-/-	n.a.	n.a.	n.a.	-/-	-/-
	Final model	467.34	95	< .001	.945	.050	.072	.066	.079
ANZ (*n* = 379)	Final model	246.47	95	< .001	.939	.062	.073	.062	.084
UK (*n* = 316)	Final model	226.12	95	< .001	.941	.052	.073	.061	.085
NA (*n* = 298)	Final model	272.55	95	< .001	.926	.059	.092	.079	.105
Male sample (*n* = 535)	Final model	295.05	95	< .001	.945	.053	.073	.063	.082
Female sample (*n* = 458)	Final model	287.31	95	< .001	.938	.057	.074	.065	.084

Note: ARSS = Acute Recovery and Stress Scale; ANZ = Australian/New Zealand; UK = United Kingdom; NA = North America (USA. Canada); CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Error of Approximation; LO90 = Lower limit of the 90%-confidence interval; HI90 = Upper limit of the 90%-confidence interval; n.a. = not available.

Table 6. Results of multigroup confirmatory factor analyses for gender and regional subsamples

Measurement invariance		χ^2	*df*	*p*	CFI	SRMR	RMSEA	LO90	HI90	ΔCFI	ΔSRMR	ΔRMSEA
Recovery Dimension												
Gender	Configural	487.65	192	< .001	.958	.038	.063	.056	.070	-/-	-/-	-/-
	Metric	505.35	204	< .001	.957	.042	.061	.055	.068	-.001	.004	-.002
	Scalar	559.06	216	< .001	.952	.044	.063	.057	.07	-.005	.002	.002
Region	Configural	656.85	288	< .001	.949	.044	.069	.062	.076	-/-	-/-	-/-
	Metric	69.67	360	< .001	.948	.05	.067	.06	.074	-.001	.006	-.002
	Scalar	734.53	336	< .001	.946	.051	.066	.059	.072	-.002	.001	-.001
Stress Dimension												
Gender	Configural	582.64	190	< .001	.942	.052	.065	.073	.067	-/-	-/-	-/-
	Metric	597.18	202	< .001	.942	.054	.071	.064	.078	.000	.002	.006
	Scalar	644.13	214	< .001	.938	.055	.071	.065	.078	-.004	.001	.000
Region	Configural	747.46	285	< .001	.934	.055	.079	.072	.086	-/-	-/-	-/-
	Metric	807.14	309	< .001	.93	.064	.078	.072	.085	-.004	.009	-.001
	Scalar	838.74	333	< .001	.93	.065	.075	.069	.082	.000	.001	-.003

Note: CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Error of Approximation; LO90 = Lower limit of the 90%-confidence interval; HI90 = Upper limit of the 90%-confidence interval.

619

Table 7. Spearman correlations between the RESTQ-Sport-76 and the ARSS (upper lines) and SRSS (lower lines)

ARSS/ SRSS	RESTQ-Sport-76																		
	Overall Stress							*Overall Recovery*					*Sport-specific Stress*			*Sport-specific Recovery*			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
PPC	-.44[a]	-.41[a]	-.34[a]	-.35[a]	-.34[a]	-.50[a]	-.49[a]	.50[a]	.31[a]	.63[a]	.53[a]	.36[a]	-.27[a]	-.37[a]	-.32[a]	.67[a]	.43[a]	.58[a]	.47[a]
	-.34[a]	-.34[a]	-.29[a]	-.31[a]	-.35[a]	-.46[a]	-.46[a]	.42[a]	.29[a]	.55[a]	.46[a]	.29[a]	-.24[a]	-.31[a]	-.38[a]	.60[a]	.36[a]	.46[a]	.36[a]
MPC	-.45[a]	-.41[a]	-.38[a]	-.33[a]	-.35[a]	-.52[a]	-.44[a]	.50[a]	.27[a]	.55[a]	.50[a]	.36[a]	-.27[a]	-.40[a]	-.21[a]	.51[a]	.39[a]	.50[a]	.37[a]
	-.50[a]	-.47[a]	-.40[a]	-.40[a]	-.40[a]	-.57[a]	-.46[a]	.48[a]	.33[a]	.49[a]	.50[a]	.36[a]	-.30[a]	-.37[a]	-.23[a]	.49[a]	.36[a]	.46[a]	.33[a]
EB	-.61[a]	-.58[a]	-.46[a]	-.45[a]	-.36[a]	-.52[a]	-.50[a]	.54[a]	.44[a]	.67[a]	**.72**[a]	.45[a]	-.31[a]	-.43[a]	-.24[a]	.56[a]	.45[a]	.54[a]	.36[a]
	-.62[a]	-.57[a]	-.46[a]	-.48[a]	-.40[a]	-.53[a]	-.47[a]	.47[a]	.37[a]	.59[a]	.65[a]	.42[a]	-.29[a]	-.41[a]	-.21[a]	.50[a]	.38[a]	.45[a]	.29[a]
OR	-.30[a]	-.31[a]	-.24[a]	-.35[a]	-.43[a]	-.36[a]	-.49[a]	.27[a]	.22[a]	.48[a]	.34[a]	.34[a]	-.30[a]	-.28[a]	-.48[a]	.43[a]	.19[a]	.30[a]	.15[a]
	-.29[a]	-.31[a]	-.24[a]	-.36[a]	-.40[a]	-.38[a]	-.46[a]	.31[a]	.19[a]	.50[a]	.38[a]	.35[a]	-.28[a]	-.29[a]	-.43[a]	.49[a]	.27[a]	.35[a]	.24[a]
MS	.18[a]	.21[a]	.19[a]	.23[a]	.36[a]	.27[a]	.40[a]	-.12[b]	-.06[d]	-.26[a]	-.11[b]	-.17[a]	.27[a]	.25[a]	.57[a]	-.24[a]	-.03[d]	-.11[c]	.01[d]
	.14[b]	.18[a]	.18[a]	.24[a]	.34[a]	.25[a]	.39[a]	-.16[a]	-.03[d]	-.26[a]	-.10[c]	-.19[a]	.28[a]	.23[a]	.57[a]	-.26[a]	-.08[d]	-.13[b]	-.06[d]
LA	.55[a]	.53[a]	.49[a]	.45[a]	.46[a]	.62[a]	.58[a]	-.44[a]	-.26[a]	-.55[a]	-.50[a]	-.39[a]	.35[a]	.51[a]	.38[a]	-.56[a]	-.33[a]	-.46[a]	-.34[a]
	.53[a]	.49[a]	.45[a]	.41[a]	.42[a]	.59[a]	.54[a]	-.42[a]	-.27[a]	-.51[a]	-.48[a]	-.37[a]	.36[a]	.50[a]	.32[a]	-.52[a]	-.29[a]	-.44[a]	-.36[a]
NES	**.70**[a]	**.74**[a]	.66[a]	.61[a]	.47[a]	.58[a]	.60[a]	-.39[a]	-.37[a]	-.57[a]	-.61[a]	-.48[a]	.43[a]	.55[a]	.35[a]	-.51[a]	-.36[a]	-.42[a]	-.24[a]
	.68[a]	.66[a]	.60[a]	.52[a]	.42[a]	.54[a]	.56[a]	-.41[a]	-.41[a]	-.56[a]	-.62[a]	-.44[a]	.36[a]	.52[a]	.25[a]	-.50[a]	-.37[a]	-.46[a]	-.30[a]
OS	.48[a]	.49[a]	.43[a]	.49[a]	.58[a]	.56[a]	.64[a]	-.31[a]	-.23[a]	-.51[a]	-.39[a]	-.38[a]	.45[a]	.52[a]	.52[a]	-.51[a]	-.22[a]	-.35[a]	-.19[a]
	.53[a]	.53[a]	.46[a]	.51[a]	.53[a]	.54[a]	.57[a]	-.32[a]	-.25[a]	-.51[a]	-.43[a]	-.42[a]	.41[a]	.49[a]	.39[a]	-.49[a]	-.28[a]	-.40[a]	-.25[a]

Note: [a] = $p < .001$; [b] = $p < .01$; [c] = $p < .05$; [d] = non-significant; $r_s \geq .70$ are bolded; ARSS = Acute Recovery and Stress Scale; SRSS = Short Recovery and Stress Scale; RESTQ-Sport-76 = Recovery-Stress Questionnaire for Athletes; PPC = *Physical Performance Capability*; MPC = *Mental Performance Capability*; EB = *Emotional Balance*; OR = *Overall Recovery* MS = *Muscular Stress*; LA = *Lack of Activation*; NES = *Negative Emotional State*; OS = *Overall Stress*; 1 = *General Stress*; 2 = *Emotional Stress*; 3 = *Social Stress*; 4 = *Conflicts/Pressure*; 5 = *Fatigue*; 6 = *Lack of Energy*; 7 = *Physical Complaints*; 8 = *Success*; 9 = *Social Recovery*; 10 = *Physical Recovery*; 11 = *General Well-being*; 12 = *Sleep Quality* 13 = *Disturbed Breaks*; 14 = *Emotional Exhaustion*; 15 = *Injury*; 16 = *Being in Shape*; 17 = *Personal Accomplishment*; 18 = *Self-Efficacy*; 19 = *Self-Regulation*.