The reliability and validity of the Dutch versions of the Acute Recovery Stress Scale and the Short Recovery and Stress Scale

Liset Schuitert

Master Thesis – Talent Development and Creativity

S3076784

June 2021

Department of Psychology

University of Groningen

Examiner/Daily supervisor: Ruud den Hartigh

Second evaluator: Yannick Hill

Abstract

To satisfy the need for a reliable, valid, and time-efficient instrument to measure recovery and stress, the Acute Recovery and Stress Scale (ARSS) and the Short Recovery and Stress Scale (SRSS) were developed. The aim of the present study is to measure the reliability and validity of both instruments in Dutch. A total of 82 Dutch and Flemish athletes filled out the questionnaires from a variety of individual and team sports. The Dutch versions showed good internal consistency for both the ARSS (α = .78 - .83) and SRSS (α = .78), and a satisfactory model fit was found for six out of eight subscales of the ARSS. A poor fit was found for the subscales *Muscular Stress* and *Lack of Activation*. For the ARSS, the item-total correlations ranged between .54 and .89, which reflects an acceptable construct validity, and which means the subscales of the questionnaire can be used independently. Finally, for both recovery and stress, a positive relationship was found with the total rate of recovery, reflecting satisfactory criterion validity. In addition, a low correlation was found between RPE and the stress dimension of the ARSS and SRSS. Taken together, these first insights suggest that the ARSS and SRSS are reliable and valid tools to use in a sport setting. However, further research is needed with the use of repeated measures for further validation and usage in practical settings.

Keywords: load, elite athletes, training, monitoring, self-report measures

Introduction

During the Olympics of 2016 in Rio de Janeiro, Femke Heemskerk, a Dutch elite swimmer, performed below her potential. The Olympic games were a high deception, and she was very frustrated about her performances. She concluded that in the last phase of her preparation, the level of stress was too high in combination with suboptimal recovery. Also, she felt lots of pressure from people around her and did not get the support she needed. The time she had was too short to work on her swimming skills and to take enough rest. This example illustrates that performance may decrease when recovery is suboptimal, in combination with excessive training load. In addition, psychosocial load such as stress in private life or social support from others, should also be considered. Taken together, when recovery and stress levels of an athlete are not balanced, the level of performances decreases (Kellmann et al., 2018; Heidari et al., 2018). Hence, the recovery-stress state of an athlete is of high importance.

Regular training is an important factor in the performance and success of athletes. In the last decade, a lot of research has been conducted on recovery and stress and how to improve this in the training cycle of athletes (Kellmann et al., 2018; Hausswirth & Mujika, 2013; Heidari et al., 2018). To reach optimal performance, a good balance between recovery and stress (i.e., intensity and duration of training) is necessary (Hausswirth & Mujika, 2013). However, this balance is very subtle. If recovery is not good enough, there is a higher risk of incurring an injury. In addition, performance progress is suboptimal when the recovery period is too long and stress is too low (Hausswirth & Mujika, 2013; Heidari et al., 2018; Garcia Secades, et al., 2017; Kenttä & Hassmén, 1998). Therefore, it is important to monitor recovery and stress processes.

Currently, there are a lot of instruments which measure recovery and stress, but an instrument which is time-efficient, captures all dimensions and is valid is missing (Kellmann et al., 2018; Heidari et al., 2018). Therefore, there is a need for a short questionnaire that measures both recovery and stress in a more time-efficient, and multidimensional way to gain more insight in recovery and stress processes of athletes. In this master thesis I will make a first step in the validation of a new instrument focused on measuring recovery and stress.

Recovery and stress in sports

Recovery can be seen as a multidimensional (e.g., physiological, psychological) restorative process relative to time (Kellmann et al., 2018; Heidari et al., 2018). Besides this, it can be further expanded by different kinds of modalities of recovery, for example psychological recovery strategies or regeneration which is a natural process of restoring broken organs or missing cells in humans (Kellmann et al., 2018). Recovery is highly important in the training and match cycle of an athlete, yet coaches tend to underestimate the need for recovery (Kellmann et al., 2018). A possibility to make recovery more important is to include recovery as an integral component of the training schedule (Bompa & Haff, 2009). This can be done through the implementation of strategies like measuring the rate of perceived exertion and implementing self-report questionnaires that measures recovery. For instance, an optimal balance for an athlete can be one active recovery day and one complete recovery day per microcycle (Bompa & Haff, 2009). A microcycle is the smallest unit of time in a recurrent schedule (Bompa & Haff, 2009).

Next to an optimal balance, athletes can actively contribute to the recovery process. Although recovery is not yet considered as an important component in the training process, Kellmann et al. (2018) found that a positive perception of recovery will lead to the ideal recovery routine. An athlete can anticipate on this, by addressing the correct psychological and physiological mechanisms which are needed to recover as effectively as possible from training (Kellmann et al., 2018).

On the other side, stress is also an important factor in the recovery process of athletes. With regard to stress, a stressor on the system is the external and internal load an athlete experiences during a training session. Stress can be seen as a collection of demands (e.g., training load, life stressors, injuries) that has a negative influence on the homeostasis of metabolic and physiological processes (Kellmann et al., 2018). In particular, trainers and coaches are convinced that the training result depends on the amount and type of the stimulus and that gaining insight into this cause-and-effect relationship between training dose and response needs to be clear to prescribe exercise training (Lambert, 2012). That is, if the stimulus is high, there is a high chance that an athlete needs more recovery time then when the stimulus is low. To further elaborate on this, it is important to quantify the training load precisely (Sharma and Mujika, 2018). The external training load is the objective measure of the energy and work an athlete puts in a training session or competition (Lambert, 2012). On the other hand, the internal load is characterized by a disturbance in homeostasis of the metabolic and physiological processes that takes place during a training session (Lambert, 2012).

Taken together, there is not always an optimal balance between recovery and stress. Different consequences can occur when there are suboptimal recovery and stress processes. When the recovery-stress balance is not adequate, there will be a disruption in the psychological or physiological imbalance which can lead to overreaching (OR) (Brink, Visscher, Coutts and Lemmink, 2010). OR can be specified by a brief period of decreased sport-specific performance and is in line with symptoms as poor concentration, disturbed mood, increased fatigue and changing sleeping and eating patterns. Next to this, OR can be divided into functional overreaching (FOR) or non-functional overreaching (NFOR) which will be determined by the criteria of what the duration of decreased performance is and the severity of the symptoms

(Brink et al., 2010; Meeusen et al., 2013). If an athlete experiences reduced short-term performance, then it is FOR whereas NFOR is more for conditions where reduced performance takes longer and represents more symptoms (Brink et al., 2010).

If it is the case that an athlete has insufficient individualized and systematic recovery and does not take action, the status of overreaching can lead to different detrimental situations (Kellmann et al., 2018). This can be specified as an overtraining syndrome and a characteristic of this syndrome is that when it gets worse, it will become an under-recovery syndrome (Kellmann et al., 2018). An under-recovery syndrome is the effect of a recovery-stress state which is not in balance. There can be an imbalance between recovery aspects and daily life rules and demands, and this will represent the reaction to common stress (e.g., media, friends, family) as a far-reaching state of insufficient recovery (Kellmann & Kölling, 2019). On the other hand, Meeusen et al. (2013) are talking more about negative hormonal and psychological alterations and a diminished performance. The general rule here is: if you apply an intervention strategy early, the recovery period from overtraining will be shorter (Kellmann & Kölling, 2019).

A potential consequence of overtraining are injuries. Stress can increase the risk of injuries through higher physical fatigue, increased muscle tension, lower muscular efficacy, and a reduced motor coordination and flexibility (Hanson, McCullagh & Tonymon, 1992; Codonhato et al., 2018). Physical (and mental) exhaustion can be caused by the high-demand competitions and high training loads that require maximum levels of effort and performance from an athlete (Fletcher & Sarkar, 2012; Codonhato et al., 2018). Recovery can have a positive influence on stress because an athlete takes time to slow down and it possibly decreases injury risk sideways (Codonhato et al., 2018).

In conclusion, an optimal balance between recovery and stress leads to optimal achievements and peak performances. Important to note is that this is only reachable when training stress and the following recovery is balanced and when athletes can recover after a training and competition (Kellmann et al., 2018). To better facilitate that athletes do not have maladaptive responses, it is important to measure and monitor both recovery and stress in a reliable and valid way.

Current instruments

There are different instruments with which recovery and stress in sport can be assessed, ranging from one item tests to long questionnaires. Popular instruments are: Borg's Rating of Perceived Exhaustion (Borg, 1998), Delayed-Onset Muscle Soreness (Ohnhaus & Adler, 1975), Profile of Mood States (McNair, Lorr, & Droppleman, 1992) and Recovery-Stress Questionnaire for Athletes (Kellmann & Kallus, 2001, 2016). The descriptive information of these instruments, which have their advantages and disadvantages, can be found in Table 1.

Borg's Rating of Perceived Exhaustion (RPE; Borg, 1998) is a one-dimensional scale with a range from 6 to 20 (very, very easy to very, very hard) that assesses the intensity of the training and the exhaustion an athlete experiences at a certain time. Nowadays, a ten-point scale is also available which is frequently used in practice (Borg, 1998). An advantage of this instrument is that it is an economical, short, and practical tool. Another way to measure recovery and stress is using the Delayed-Onset Muscle Soreness (DOMS; Ohnhaus & Adler, 1975) in which the outcome represents the pain index. This is done through a mark on a 10-cm line, with a range from no pain (left endpoint) to extreme pain (right endpoint). The distance (in mm or cm) between the mark and the left endpoint is the pain index (Ohnhaus & Adler, 1975). Although these instruments are easy to administer, they carry the disadvantage that they are one-dimensional.

The third instrument is the Profile of Mood States (POMS; McNair et al., 1992). The POMS measures the mood state of a person through a 65-item questionnaire (McNair et al., 1992) with the advantage that it is a multidimensional tool. A disadvantage of this instrument is that it is not explicitly developed as a sport-specific tool, so the dimensions do not sufficiently represent the recovery aspect (Mäetsu, Jürimäe & Jürimäe, 2005). This instrument is developed for use in the clinical-psychological context, but it is also used in sport-scientific practice and in research about overtraining (e.g., Bresciani et al., 2011; Raglin, Morgan, & O'Connor, 1991).

Nowadays, a frequently used questionnaire to report and investigate recovery and stress in sports is the Recovery-Stress Questionnaire for Athletes (REST-Q-Sport; Kellmann, 2011, 2016). This self-report questionnaire measures recovery and perceived stress and gives insight in the recovery strategies an athlete can use. The focus of the REST-Q is on the mental and physical impact of training and performance stress, and to facilitate strategies for improvement of recovery (Kellmann & Kallus, 2001). An advantage of this instrument is that it is a reliable and valid tool, but on the other side, it is very time-consuming.

As stated above, the REST-Q is a valid, but long questionnaire. Therefore, a shorter, but valid questionnaire which measures the same aspects as the REST-Q was developed: The Acute Recovery and Stress Scale (ARSS; Hitzschke et al., 2015; Kellmann, Kölling, & Hitzschke, 2016), and the Short Recovery and Stress Scale (SRSS; Hitzschke et al., 2015; Kellmann et al., 2016). The mentioned questionnaires are both trying to assess the emotional, mental, and physiological aspects of recovery and stress.

Table 1

Descriptives of the current instruments

	Type of study	Advantage	Disadvantage
RPE	Experimental studies. Individual training load	Economical, short, practical and valid instrument for assessing training intensity	One-dimensional, it is unclear what the underlying reasons are for a change in effort across training sessions for the same training load. So therefore, it might be difficult to gain practical intervention measures based on the estimations only
DOMS	Experimental studies	Economical, quick, good sensitivity for change in pain perception, high retest reliability	One-dimensional, restricted to the muscular component of stress
POMS	Overtraining studies. Sport-scientific practice	Multidimensional, holistic approach	Not a sport-specific tool, no training prescription and intervention is available, no concrete recommendations for performance control and the domains do not cover the recovery aspect enough
REST-Q	Monitoring of training. Prevention of injury and overtraining	Early indicator of overtraining symptoms, can differentiate between two groups, reliable and valid	Long questionnaire with a lot of items, participants need to remember feelings and experiences for three days which causes unsureness to the results

Note. RPE = Rating of Perceived Exertion; DOMS = Delayed-Onset Muscle Soreness; POMS = Profile of Mood States; REST-Q = Recovery-Stress Questionnaire for Athletes

Source. Kellmann and Kölling (2019).

Acute Recovery Stress Scale and Short Recovery Stress Scale

As mentioned above, there are various instruments to measure recovery and stress in athletes. Nevertheless, a short, valid, and time-efficient instrument, which also assesses all dimensions of recovery and stress adequately, is missing. The final version of the ARSS consists of 32 adjectives which are grouped in eight scales and two dimensions, which are recovery and stress (Kellmann & Kölling, 2019).

The SRSS consists of eight items divided over mental, physical, emotional, and overall levels and thereby also measures the recovery-stress state in a multidimensional way (Kellmann & Kölling, 2019). These eight items are in accordance with the eight scales of the ARSS. In this way, the SRSS is much shorter than the ARSS. The items of the SRSS are grouped into the Short Recovery Scale (SRS), and the Short Stress Scale (SSS) which both consists of four items that depict the underlying adjectives from the ARSS (Kellmann & Kölling, 2019).

For determining the validity, the construct and convergent validity are based on the scales of the ARSS and SRSS. In total, there are eight different scales. The scale *'Physical Performance Capability'* gives information about recovery adaptations when regeneration strategies are used, whereas the scale *'Mental Performance Capability'* is about concentration, being receptive and attentive, and feeling alert (Kellmann & Kölling, 2019). The third scale is *'Emotional Balance'*, and this is about being in a good mood, being stable and having control. In addition, the scale *'Overall Recovery'* gives information about the physical and mental recovery, and it shows the reaction to recovery and stress related stimuli (Kellmann & Kölling, 2019). The above-mentioned scales are all related to the recovery aspect of the ARSS and SRSS.

The stress aspect of the ARSS and SRSS consists of the scale *'Muscle Stress'* which is about stress reduction, the effects of recovery and preceding stress. Second, the scale *'Lack of Activation'* gives insight into feelings of amotivation and lack of energy. It shows a reaction to stress factors on the longer term (Kellmann & Kölling, 2019). In addition, the scale *'Negative Emotional State'* is about feeling currently emotionally stressed by certain demands, whereas the last scale *'Overall Stress'* is valuable for a more universal aspect of stress and is about being tired, exhausted, and overloaded (Kellmann & Kölling, 2019).

Currently, different studies have been conducted to test the reliability and validity of the ARSS and the SRSS. In Germany, Hitzschke et al. (2015) did research during a training camp of the German junior national field hockey team. They found a satisfactory internal consistency for the ARSS ($\alpha = .84$ and $\alpha = .96$). For the development of the ARSS, Kellmann & Kölling (2019) used Exploratory and Confirmatory Factor Analyses to test the questionnaire. For both factors, the model fit was acceptable for recovery (Root Mean Square Error of Approximation (RMSEA) = .07, Comparative Fit Index (CFI) = .97, Standardized Root Mean Square Residual (SRMR) = .04) and for stress (RMSEA = .09, CFI = .94, SRMR = .05). The construct and convergent validity were determined by the relationship with the current instruments – DOMS, POMS and REST-Q – under both recovery and stress situations (Kellmann & Kölling, 2019). In addition, there was a moderate inter-correlation and a satisfactory homogeneity range.

Another study is the research of Nässi et al. (2017) for the validation procedure of the English version. They introduced the instrument in England with 267 athletes from different types of sport. They found a satisfactory internal consistency for both the ARSS and the SRSS (α of .74 - .89). Second, there was a good model fit for the scales of the ARSS, which is in line with the results of the German version. In addition, Nässi et al. (2017) found promising results for the construct and convergent validity.

Besides the above-mentioned promising results, all scales of the ARSS detected the loads of a hockey training camp (Kölling et al., 2015) and the scales *'Muscular Stress'* and *'Overall Stress'* detected differences in load of a cycling training camp (Hammes et al., 2016). It has been shown that the SRSS was sensitive to high-resistance micro-cycle strength training (Raeder et al., 2016), to high-intensity interval training (Wiewelhove et al., 2016) and to short sprint protocols as well (Pelka et al., 2017). The results so far indicate that the scales are reliable and valid in both the English and German language, but currently the psychometric properties in the Dutch language are unknown.

Current study

The current study examines some psychometric properties (i.e., reliability and validity) of the ARSS and the SRSS in Dutch. The research question is: What is the reliability and validity of the Dutch versions of the Acute Recovery Stress Scale and the Short Recovery and Stress Scale. In order to answer this question, I will first determine Cronbach's alpha to measure the reliability. Moreover, with the use of a confirmatory factor analysis, I will test the model fit. The overall aim of the tests for validity is to test whether the ARSS and SRSS measures recovery and stress appropriately (construct validity). In order to further check the validity, I will investigate the relationship between recovery and stress and the Rating of Perceived Exertion and Total rate of Recovery from the past training (convergent validity). The Total rate of Recovery will be measured by multiplying the time since last training times the Total Quality Recovery Scale (TQR). My hypothesis is that when your recovery and stress processes are not optimal, you will experience the training relatively heavier. Earlier research concluded that all athletes recover in a different way (Kellmann et al., 2018). An athlete will reach adequate balance between the perceived recovery and stress when reaching peak performance and avoiding overtraining (or under-recovery) (Kellmann, 2010). When there is an imbalance between the recovery and the stress state, it means that this is associated with an impaired performance, an increased injury risk, and physical recovery (Kenttä & Hassmén, 1998; Van der Does, Brink, Otter, Visscher, & Lemmink, 2017).

Besides looking at recovery and stress in general, it might be interesting to investigate if athletes who score low on the recovery dimensions experience more stress on the stress dimensions (and vice versa). The corresponding hypothesis is that athletes who have higher stress (exertion/internal load) in the week before, or athletes who have higher stress relative to recovery, score higher on stress and lower on recovery scales of the SRSS.

The last type of validity is the criterion validity. To assess this type of validity, I will use the concurrent validity and I will test it with the correlation between the ARSS and SRSS and to test if both measures can estimate individual performance of athletes on different tests (i.e., the ARSS and SRSS) at the same time. The difference with determining the relationship with the RPE and TQR is that in this case, the whole dimension of recovery and stress is considered. Also, the time since the last training plays a role in this case. So, I will look to the whole test and not specifically to one subscale or one measure.

Method

Participants

Of the 450 contacted athletes, 194 athletes filled out the questionnaires and 42% of the athletes was 16 years or older and completed the whole questionnaire. Participants were recruited through a collaboration with NOC*NSF and with Sport Vlaanderen. A total of 82 Dutch-speaking and Flemish-speaking athletes (47 male; 34 female; 1 other) completed an online version of the ARSS questionnaire and the SRSS questionnaire in their own time after a training session. The following sports were represented: hockey (17), badminton (15), korfball (8), soccer (7), horse riding (5), judo (5), weight lifting (5), archery (3), handball (3), powerlifting (3), athletics (2), canoeing (2), cycling (2), triathlon (2), curling (1), gymnastics (1), rowing (1), softball (1), taekwondo (1), waterpolo (1), other (3) in the levels Olympic (17), Europe's top (18), Dutch top (17) and regional (30).

Measures

The ARSS consisted of a list of 32 adjectives which aimed to measure elements of recovery and stress. The participants had to rate how much each statement applied to them now, with responses ranging from 0 (does not apply at all) to 6 (fully applies). The statements can be grouped into eight different scales, of which four are related to recovery and the other four are related to stress. The recovery-related scales are *Emotional Balance* (four items), *Mental Performance Capability* (four items), *Physical Performance Capability* (four items), and *Overall Recovery* (four items) and the stress-related scales are *Muscular Stress* (four items), *Negative Emotional State* (four items), *Lack of Activation* (four items), and *Overall Stress* (four items).

The SRSS is complementary to the ARSS. The instruction for the SRSS is similar to the instruction of the ARSS, because the SRSS is a compact version of the ARSS (Hitzschke et al., 2015; Kellmann & Kölling, 2019). This means that the items of the ARSS are taken together under a subscale. Each item in the SRSS is supported with concrete examples that match the items of the ARSS. For example, the item *Mental Performance Capability* is supported by the underlying adjectives attentive, receptive, mentally alert, and concentrated. Hence, the ARSS does not reveal the final categorization, but only the adjectives, whereas the SRSS provides athletes a complete overview of what the instrument is measuring because every subscale is supported with examples that are equivalent to the items of the ARSS (Kellmann & Kölling, 2019; Nässi et al., 2017).

Procedure

Before the questionnaires were filled out by the athletes, I first conducted a translation procedure (from English to Dutch). Specifically, a parallel back-translation procedure (Vallerand, 1989) was used to translate the English versions of the ARSS and the SRSS. This translation procedure was done by experts and academic staff of the Center for Human Movement Science of the University Medical Center Groningen, and the Department of Psychology at the University of Groningen. Besides, two sport psychologists, two sport scientists and two applied sport scientists from rowing and football federations were involved. Both the expert coaches and the staff of the university individually translated the items of the English versions of the ARSS and the SRSS. More information about the translation procedure can be found in Appendix A.

After the translation procedure, agreements and disagreements were analyzed and items that were the same in at least three out of six translations were regarded as items with satisfactory agreement. In case of no agreement, items were discussed during a meeting until there was consensus. In this discussion, the item was checked on: a) the translation, which was closest to English, b) the German equivalent, c) the use in sport and d) the use in the REST-Q-Sport questionnaire (Kellmann & Kallus, 2001). When agreement of the Dutch version was official, this result was back translated by a near-native English individual. After this, a comparison was made between the original English version and the new English version. Any difference was discussed and changed in the Dutch version if this was necessary.

For collecting the data, the authors contacted NOC*NSF and Sport Vlaanderen and distributed the survey links, and the manager of both organizations provided the link of the questionnaires to the athletes. The athletes filled out the questionnaires through Qualtrics after a training session. They could do this via computer or their phone. Information regarding the aim of the study was provided to each participant and informed consent was collected prior to the questionnaires. Being specifically, the question about the RPE and time since training are about the last training. In addition, the question about the TQR relates to this moment. The factor load was computed by multiplying the duration of the last training session by the RPE with the outcome of the session-RPE (sRPE) (Foster et al., 2001). The participants provided sociodemographic information, next to answering questions from the questionnaires. The procedures were approved by the ethics committee of the University of Groningen. The number of the ethical request is PSY-1920-S-0513. The time to fill out the ARSS was between four and five minutes. In addition, filling out the SRSS took between 40 and 60 seconds.

Data analysis

The data-analysis was done using the program SPSS 25. For reliability, I determined the internal consistency of the scales using Cronbach's alpha. If the value exceeded $\alpha = .70$, it was considered as satisfactory (DeVellis, 1991; Saw et al., 2017).

For the validity, a confirmatory factor analysis (CFA) is used to measure different models. The CFA indicators consisted of several model indices with the aim of describing the underlying factor structure. The first model was the goodness-of-fit-index (GFI) which determines the fit between the observed covariance matrix and the hypothesized model. This index should reach indexes above .90 and preferably above .95 (Kline, 2005). After, the comparative-fit-index (CFI) was used, which is looking at the discrepancy between the hypothesized model and the data to analyze the model fit while also considered issues of sample size. This value was satisfactory if it is close to .95 (Hu & Bentler, 1999). The third fit index was the root-mean-square error of approximation (RMSEA), which is looking at how far a hypothesized model, with perfect chosen parameter estimates, is from an excellent model. If this value is below .05 it will be considered as an excellent fit, whereas values up to .08 are reasonable, and a value above .10 is indicated as a poor fit (Kline, 2005). The last fit index was the standardized root-mean-square residual (SRMR), that determines the difference between the model and the observed correlation. A value below .08 is satisfactory (Hu & Bentler, 1999).

Besides the factor analysis, another aim of the validity is to test whether the ARSS and SRSS measures recovery and stress appropriately (construct validity). In order to check the validity, I will first start with the convergent validity and investigate what the relationship is between recovery and stress and the RPE and TQR of the past training. I will do this by using Spearman correlations to determine if the TQR was related to the recovery scales of the ARSS and SRSS and if the RPE scores are related to the stress scales of the ARSS and SRSS. If this is the case and there is a high correlation, it is then appropriate that the ARSS and SRSS can detect both current recovery and preceding stress levels in an athlete. A correlation can be considered as perfect ($r = 1$), almost perfect ($r > .9$), very large ($.7 < r < .9$), large ($.5 < r < .7$), moderate ($.3 < r < .5$), small ($.1 < r < .3$) and trivial ($r < .1$) (Hopkins, 2002).

In addition, I will test the discriminant validity with the use of both the recovery and stress dimensions and see whether athletes who score low on the recovery dimensions experience more stress on the stress dimensions and vice versa. I will do this by using a bivariate Pearson correlation and to see if the ARSS and SRSS can predict future scores with regard to recovery and stress. If the value is above ≥ .80 it can be considered as valid. In addition, a scatterplot is made to make the relationship between recovery and stress visual.

The last type of validity is the criterion validity. To assess this type of validity, I will use the concurrent validity and I will test it with the correlation between the ARSS and SRSS, and to test if both measures can estimate individual performance of athletes on different tests (i.e., the ARSS and SRSS) at the same time. The difference with determining the relationship with the RPE and TQR is that in this case, the whole dimension of recovery and stress is considered. So, I will look to the whole test and not specifically to one subscale or one measure.

Results

After applying a translation procedure, no items were added or removed from the ARSS. In total, 82 athletes (47 male; 34 female; 1 other) with a mean age of 21.1 years (*SD* = 6.47) were included in this study. The participants trained on average 17.51 hours per week (*SD* = 7.77). The mean RPE of the last training was 13.04 (*SD* = 2.41) with a range of 6 to 19. The average training duration per week in hours is 17.5 (*SD* = 7.8) with a mean of 1.94 (*SD* = 1.0) for the duration of the last training.

The descriptive statistics (*M* ± *SD*) for the ARSS are shown in Table 2 and for the SRSS in Table 3.

Table 2

Descriptive statistics and item-total correlations of the ARSS

		M	*SD*	α	r_{it}			*M*	*SD*	α	r_{it}
Recovery dimension	**PPC**	3.78	1.23	.83		Stress dimension	**MS**	2.00	1.49	.83	
	PPC-item 1	3.98	1.11		.82		MS-item 1	2.39	1.55		.83
	PPC-item 2	4.11	1.18		.78		MS-item 2	2.39	1.42		.88
	PPC-item 3	3.63	1.30		.85		MS-item 3	1.54	1.56		.87
	PPC-item 4	3.41	1.32		.89		MS-item 4	1.68	1.42		.82

MPC	3.48	1.43	.78		**LA**	1.64	1.50	.81	
MPC-item 1	3.70	1.39		.74	LA-item 1	1.32	1.62		.79
MPC-item 2	2.80	1.64		.54	LA-item 2	1.89	1.52		.74
MPC-item 3	3.66	1.37		.81	LA-item 3	1.28	1.44		.84
MPC-item 4	3.78	1.30		.75	LA-item 4	2.06	1.43		.72
EB	4.10	1.26	.81		**NES**	1.70	1.54	.83	
EB-item 1	4.46	1.20		.82	NES-item 1	1.38	1.52		.84
EB-item 2	4.16	1.12		.75	NES-item 2	1.95	1.49		.79
EB-item 3	4.16	1.30		.82	NES-item 3	1.52	1.53		.86
EB-item 4	3.63	1.41		.80	NES-item 4	1.94	1.61		.86
OR	3.66	1.32	.83		**OS**	1.85	1.39	.82	
OR-item 1	4.22	1.30		.81	OS-item 1	2.72	1.30		.82
OR-item 2	3.40	1.31		.85	OS-item 2	1.51	1.53		.81
OR-item 3	3.34	1.34		.85	OS-item 3	1.45	1.30		.74
OR-item 4	3.67	1.30		.86	OS-item 4	1.73	1.46		.88

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 of the SRSS

		M	*SD*			*M*	*SD*
Short Recovery Scale	PPC	3.93	1.18	Short Stress Scale	MS	2.20	1.44
	MPC	3.82	1.34		LA	1.49	1.43
	EB	3.66	1.35		NES	1.78	1.63
	OR	3.77	1.30		OS	1.93	1.30

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.

Reliability ARSS

For the ARSS, scale homogeneity levels (α = .78 - .83) are all above the minimum acceptable level of α = .70 (DeVellis, 1991). Moreover, item-total correlations varied between *r* = .54 and *r* = .89, and the ARSS showed good homogeneity for the dimensions of recovery

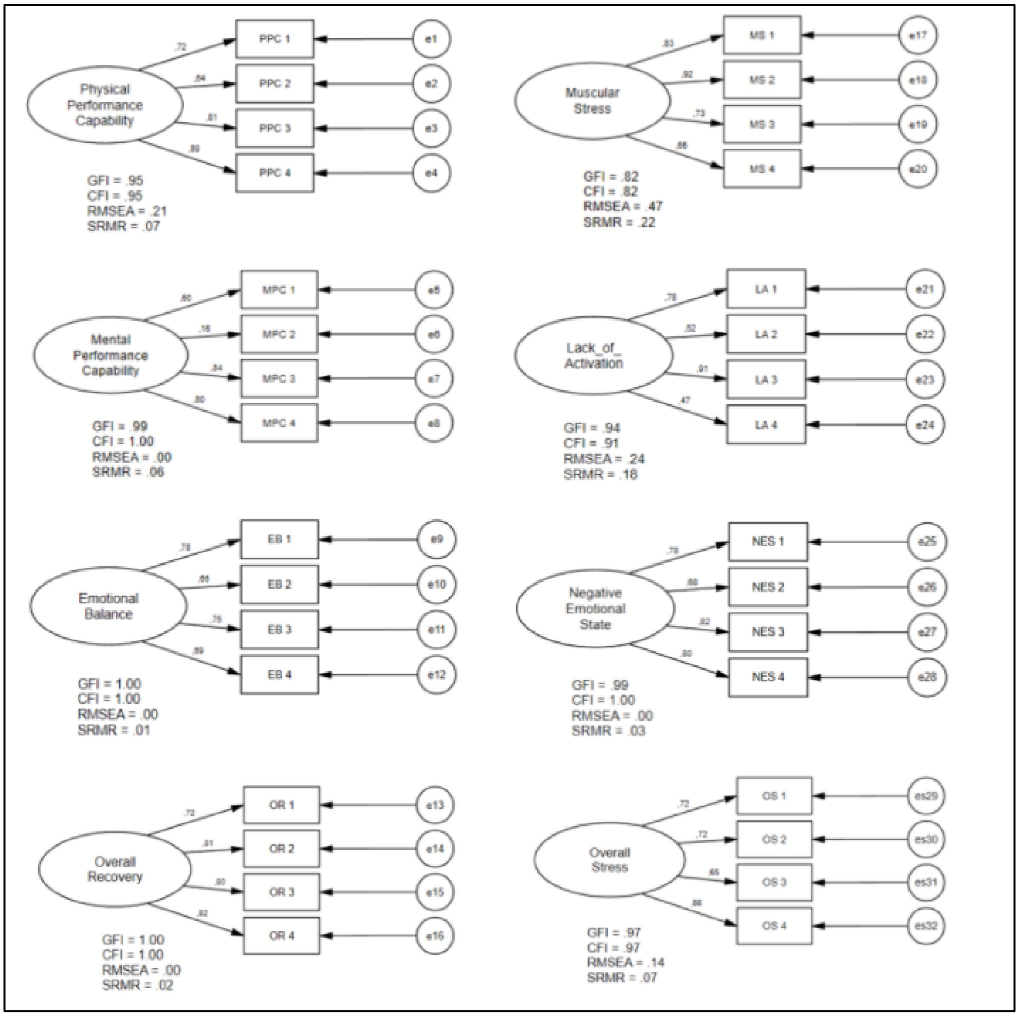
(α = .89) and stress (α = .82). In Table 2, the item-total correlations for the ARSS can be found. To further deepen this, for ARSS_recovery, item 4 of the scale *Physical Performance Capability* has the highest value (r_{it} = .89) whereas item 2 of the scale *Mental Performance Capability* has the lowest value (r_{it} = .54). For ARSS_stress, item 2 of the scale *Mental Performance Capability* and item 4 of the scale *Overall Stress* have the highest value (r_{it} = .88) whereas item 4 of the scale *Lack of Activation* has the lowest value (r_{it} = .72).

Final structure of the ARSS

After conducting a CFA, the final four items of each scale represent the model fit, reaching almost all the cut-off values for a minimum of three domains (i.e., GFI, CFI, RMSEA, and SRMR). This means that the scales meet at least three factors of the CFA. The model fit was acceptable (GFI ≥ .95 and CFI ≥ .95) for the scales *Physical Performance Capability, Mental Performance Capability, Emotional Balance, Overall Recovery, Negative Emotional State* and *Overall Stress*. The above-mentioned scales showed good model fit for at least three out of four measurement indices, but further improvements are needed for both *Muscular Stress* and *Lack of Activation* (see Figure 1). *Muscular Stress* showed poor results for all four measurements and *Lack of Activation* showed good fit for both GFI and CFI, but poor fit for both RMSEA and SRMR.

Figure 1

Measurement models for the scales of the ARSS as developed by CFA.



Note. Fit-indices data are presented in line to the corresponding model. Modified indices presented in parentheses where error correlations were allowed based on modification indexes. GFI = goodness-of-fit-index; CFI = comparative-fit-index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual.

Final structure of the SRSS

Regarding the homogeneity, the SRSS showed good homogeneity for both recovery and stress (α = .78 and α = .78). For recovery, item-total correlations for the SRSS vary between *r* = .73 and *r* = .83, and for stress it varies between *r* = .67 and *r* = .85. To further deepen this, for

SRSS_recovery, the scale *Mental Performance Capability* has the highest value (r = .83) whereas *Overall Recovery* has the lowest value (r = .73). For SRSS_stress, the scale *Overall Stress* has the highest value (r = .85), and *Muscular Stress* has the lowest value (r = .67).

Convergent validity

In Table 4, the correlation between recovery and preceding load of the recovery and stress scales of the ARSS and SRSS is shown. When looking at the table, the largest correlation for the ARSS is found between the factor TQR and the ARSS scale *Overall Recovery* (r_s = .79) whereas the largest correlation for the SRSS is found between the factor TQR and the SRSS scale *Muscular Stress* (r_s= - .63). For recovery there was a positive association with the TQR, whereas there was a negative association between the TQR and the dimension stress.

To further examine convergent validity, I examined the relationship between recovery and stress dimensions/items and the RPE of the past training. When looking at the results, the correlation between ARSS_recovery and the TQR is .65 and for the SRSS_recovery the correlation is .49. The inter-item correlation between ARSS_stress and the RPE of the past training is .19 (p = .08) and for SRSS_stress the correlation is .18 (p = .11).

Table 4

Correlation of the ARSS scales and SRSS items preceding training load and recovery.

			Training			Recovery	
			Duration last training	RPE	sRPE	TQR	Recovery rate[a]
Recovery dimension / Short Recovery Scale	PPC	ARSS	-.24*	-.07	-.14	.61**	.27*
		SRSS	-.14	.07	-.08	.41**	.19
	MPC	ARSS	-.24*	-.06	-.13	.48**	.28*
		SRSS	-.14	.03	-.10	.31**	.23*
	EB	ARSS	-.18	-.10	-.16	.36**	.15
		SRSS	-.14	.02	-.06	.21	.16
	OR	ARSS	-.29**	-.11	-.19	.79**	.46**
		SRSS	-.21	.01	-.14	.60**	.29**
Stress dimension / Short Stress Scale	MS	ARSS	.12	.08	.14	-.67**	-.39**
		SRSS	.17	.15	.19	-.63**	-.37**
	LA	ARSS	.20	.19	.21	-.37**	-.15
		SRSS	.07	.04	.01	-.27*	-.24*
	NES	ARSS	.06	-.02	.08	-.17	-.03
		SRSS	.25*	.06	.19	-.28**	-.18
	OS	ARSS	.10	.16	.22	-.66**	-.33**
		SRSS	.19	.14	.23*	-.47**	-.26*

Note. ARSS scales/SRSS items: 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; RPE = Rate of Perceived Exertion; sRPE = session Rate of Perceived Exertion (duration*RPE); TQR = Total Quality Recovery Scale. * Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed); [a] Recovery rate was calculated by multiplying the time since last training by the TQR.

Discriminant validity

To see if an athlete who scores low on recovery, experiences more stress on the stress dimensions and vice versa, a scatterplot of the total score on the recovery dimension of the ARSS vs. the total score on the stress dimension of the ARSS is made. The same procedure is also done for the SRSS. In Figure 2 the scatterplot between ARSS_recovery and ARSS_stress whereas in Figure 3 the scatterplot between SRSS_recovery and SRSS_stress is shown. For the ARSS, the correlation between recovery and stress dimensions is $r = -.71$ and for the SRSS this value is $r = -.73$. Both correlations are significant at the 0.01 level (2-tailed). Note that the scores within the recovery and stress dimensions are combined in a total score.

Figure 2

Association between the recovery and stress dimensions of the ARSS visually shown in a scatterplot with stress on the X-axis and recovery on the Y-axis

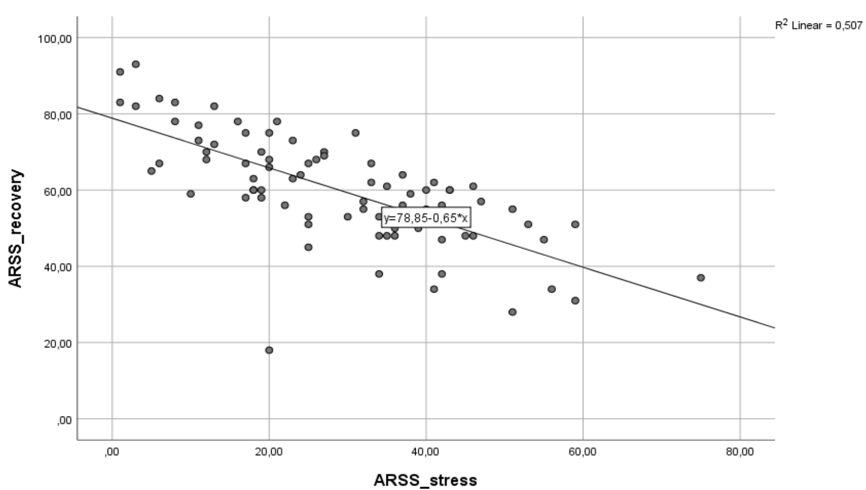
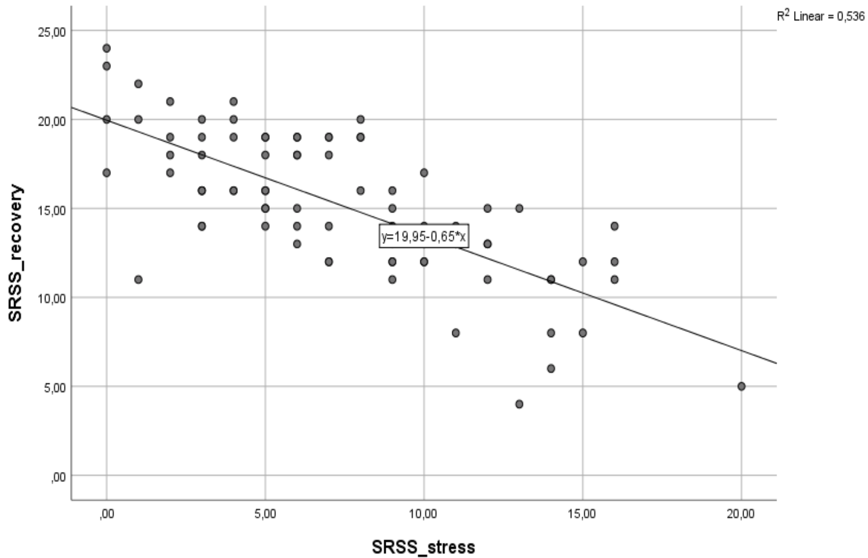


Figure 3

Association between the recovery and stress dimensions of the SRSS visually shown in a scatterplot with stress on the X-axis and recovery on the Y-axis



Concurrent validity

Table 5 shows the correlation coefficients between the corresponding scales/items of the two questionnaires. Correlations between the ARSS and SRSS were large to very large, with a Spearman's rho between .60 and .77. All correlations were significant ($p < 0.01$) which means that there was a significant, positive correlation between the ARSS and SRSS and that it seems that the SRSS measures the same as the ARSS.

Table 5

Correlations between ARSS scales and the corresponding SRSS dimensions

	Scale/item	Between ARSS & SRSS		Scale/item	Between ARSS & SRSS
Short Recovery Scale	PPC	.69**	Short Stress Scale	MS	.73**
	MPC	.60**		LA	.67**
	EB	.69**		NES	.73**
	OR	.77**		OS	.76**

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. ** Correlation is significant at the 0.01 level (2-tailed).

Discussion

The aim of this thesis was to examine the reliability and validity of the ARSS and the SRSS in Dutch. Both instruments intended to measure recovery and stress in a multidimensional, time-efficient, sport-specific, and practical way (Kellmann et al., 2018).

Descriptives

First of all, the characteristics of the included athletes and final statistical results of the ARSS and SRSS were in line with previous research (Hitzschke et al., 2016; Nässi et al., 2017; Kölling et al., 2020). The proportion between men and women is slightly similar, but athletes in the current study had a lower mean age than earlier studies (21.1 versus 23, 25.21 and 25 years) (Hitzschke et al., 2016; Nässi et al., 2017; Kölling et al., 2020). In addition, the training hours per week is higher (17.5 hours) than the other studies (14.9, 8.4 and 10.7 hours) (Hitzschke et al., 2016; Nässi et al., 2017; Kölling et al., 2020). Beside the included elite athletes, also not-elite athletes were included in this study. This can be an explanation for the dispersion, and this is also seen in the standard deviation of 7.8.

Reliability and validity

The results showed adequate and promising initial reliability and validity for the sample of Dutch athletes. The values of Cronbach's alpha are equal to or higher than $\alpha = .78$ which means that the reliability can be considered as acceptable (DeVellis, 1991). Interestingly, the reliability was slightly higher than in the German versions (Hitzschke et al., 2015; Kellmann & Kölling, 2019). This difference can be explained through different circumstances under which the athletes filled out the questionnaire.

When considering the validity, the CFA revealed an acceptable model fit for the ARSS for almost all scales reaching the suggested cut-off values of at least three of the four fit indices (Bagozzi & Yi, 2012). However, further improvements are needed for the scales *Muscular Stress* and *Lack of Activation.* A possible reason for the poor model fit may be the result of the sample or of the individual items of the scales. In terms of the individual items, there can be a discrepancy between the meaning and interpretation of an item. For example, item 3 of the scale *Lack of Activation* is 'niet enthousiast'. It could be possible that athletes found it difficult to answer this item and relate it to their recovery and stress state. Also remarkable is the lower value of item 2 ('ontvankelijk') of the scale *Mental Performance Capability* compared to other values within this scale. This difference can be explained by the translation procedure because it was the most debated item during this process. The above-mentioned results almost support the findings of the original German version. In the German version, calculations of the CFA revealed a good model fit for all scales (Kellmann et al., 2018).

In addition, to assess the convergent validity, I examined the expected relationship between the recovery scales and the TQR and the stress scales and the RPE. I expected a high, positive correlation between the TQR and the recovery scales and between the RPE and the stress scales of the ARSS and SRSS. For ARSS_recovery, there was an expected high, positive, and significant ($p < 0.01$) correlation with the TQR ($r = .65$) and a negative, significant ($p < 0.01$) correlation with the stress scales as well ($r = -.55$). For ARSS_stress, there was a very low correlation ($r = .19$) with the RPE which means that the intensity of the training does not correlate directly with the stress an athlete experienced. This finding may be explained by the fact that the time since the last training plays a role. The same pattern is found for the SRSS. Earlier research from Meckel, Zach, Eliakim and Sindiani (2018) suggest that a physiological reaction may not specifically align with an athlete's perceived exertion when an athlete anticipate in a difficult part of a training program. This result in a feeling of calmness which stems from their accomplishments and the perception of effort is lower than calculated (Meckel et al., 2018). In the current study, it was not known what the circumstances were in the time between the training and filling out the questionnaires. Different results are expected when applying this in a more controlled setting like a training camp. For example, higher stress scores are expected when using the ARSS directly after a training, and higher recovery scores are expected after waking up in the morning.

Moreover, another research question was about the discriminant validity and to see if athletes who score low on the recovery dimensions, experience more stress on the stress dimensions and vice versa. From this study, I can conclude that when an athlete experiences high stress, the recovery level is low and when an athlete experiences low stress, the recovery level is high. This result can be underlined with the study of Hitzschke et al. (2015). They found that a decreased level of recovery is associated with an increased level of stress. For this study, the same pattern is found for both the ARSS and SRSS with a slightly higher correlation for the SRSS.

Lastly, when looking at the recovery and stress scales, a negative correlation was expected between both scales. Within the ARSS, a negative correlation was expected between the recovery scales and the stress scales. The results from the analysis were in line with the expected relation and this was also the case for the SRSS and previous research for the German and English versions (Hitzschke et al., 2016; Nässi et al., 2017; Kölling et al., 2020). This result revealed that there is a high necessity to consider the negative correlations between recovery and stress scales as separate. Earlier research suggested the same that related dimensions showed positive correlations and opposite dimensions are showing negative correlations (Nässi et al., 2017; Kellmann et al., 2018).

To further expand on this, the same correlation pattern was found in the German and English versions (Hitzschke et al., 2015; Nässi et al., 2017) with a higher overall correlation for the Dutch version. Interestingly, the English versions showed a higher correlation pattern than the original German versions. The ARSS and SRSS are measuring the same concepts, therefore a large correlation was expected between the corresponding scales. A possible explanation why the correlation is not one may be that the ARSS consist of four items per subscale and the SRSS of one item per subscale which can cause a discrepancy in answering. In conclusion, the results support the expectation because there were only large correlations between both measures.

Strengths and limitations

Besides answering the current research question, this research has some promising strengths. A first strength of this study was the translation procedure followed by the use of different translators which are from different sport practices (team and endurance sports), and different nationalities (the Netherlands and Belgium). This resulted in less potential variation in terminology which is useful for the validation procedure. There was little disagreement between the translators about the translation which means that the current version of the translation has been confirmed by different translators and experts from the field.

Moreover, we included participants from a variety of sports which were both individual and team sports, ranging from regional to Olympic level. The selected athletes were abled and physically disabled and were from the Netherlands or from Belgium. Because of this variation, generalizability is enhanced.

Next to the strong generalizability, a strong point of this study is considering the relationship with often used measures for determining current recovery and preceding load. This is not done by previous studies on the ARSS and SRSS, because they only focused on the construct validity and convergent validity. The advantage of implementing this point in this study was the possibility to see if there might be a connection between the results on the ARSS and SRSS and the intensity, volume, and quality of the training. This information could be used for the validity to see if the instrument can discriminate between recovery and stress in combination with other aspects and to see if it can predict future scores on recovery and stress dimensions.

A first limitation of this study is that the participants filled out the questionnaires under heterogeneous conditions. The online platform made it possible to fill out the questionnaires from every place, so the conditions were not controlled. On the other side, this can be a strength because there is no peer-pressure and there can be a generalization to daily practice. Future research should take place under more controlled conditions (e.g., a training camp) to get more insight in the sensitivity of the measures in practice. This type of research can also obtain repeated measures data to investigate the test-retest reliability of the instrument. In line with this, there is a possibility that the circumstances were different due to COVID-19. The different circumstances could impact the recovery and stress levels of the athletes because the training program can be different, and even the circumstances in daily life. In this way, it may be difficult to compare the results with previous done studies (Wahl et al., 2020). At the same time, the different circumstances can lead to an increased generalizability.

A second limitation is the small sample size. Regarding the norms of the COTAN, the current sample size of this study is too small for a reliable estimation of the reliability and validity of the ARSS and SRSS. The COTAN is a committee of the Dutch institute of Psychologists (NIP) and assesses different tests and questionnaires for use in practice (Drenth & Sijtsma, 2006). This means that no conclusive decisions can be made based on current measures of the ARSS and SRSS. For a reliable estimation of the reliability and validity, a sample size of at least 200 athletes is needed (Drenth & Sijtsma, 2006). On the other side, all athletes in the sample are representative for the population of athletes, which suggests that inferences can be made. Another possible advantage is the ability to study patterns, effects, or differences of individual athletes. This is linked to the ergodicity problem which states that statistical relationships that are showing up during data comparison between large groups of individuals, are not directly applicable on the statistical relationships that exist within individuals (Molenaar & Campbell, 2009). Ergodicity is a problem and thus a limitation because the average behavior of recovery and stress over time can be linked to the average of all behavior related to recovery and stress of individuals (Molenaar & Campbell, 2009).

Another limitation of the study is a possible response bias. Athletes can see the evaluation areas which means that it might be easier to give answers that coaches want to see. Important to note is that the scores on the questionnaires are not shared with the coach, and this is communicated to the athletes. Therefore, the chance of having a response bias is possibly lower. For a more detailed and complete overview and to lower the possibility of response bias, the SRSS must be used together with the ARSS. Also, using the SRSS only will result in a loss of information through item reduction (i.e., emotional area). Hence, the Dutch ARSS and SRSS are showing good evidence, but improvement with the use of repeated measures is needed for the content validity (Jeffries et al., 2020).

Future research can expand on the current study by implementing repeated measures to see if there are correlations between the measures and how the results score on the longer term. Another point of interest is implementing subgroup analysis with larger samples. Due to COVID-19, the sample was not that high which can influence the generalizability. The last point of interest for future research is the responsiveness of the measures. In this study, there was only one measurement moment, and it might be interesting to see if the same results will appear in a second measurement. In addition, to further deepen into the model fit, I suggest analyzing these scales with the use of an explanatory factor analysis to see why they have a poor fit with the overall model and how this could be improved.

Conclusions and practical implications

In conclusion, the current results confirm the practical use of the ARSS and SRSS for measuring the current recovery and stress state of Dutch athletes. The study showed the factor level model structure of the ARSS and all reliability coefficients of the SRSS. I found that both the ARSS and the SRSS may be useful when monitoring the recovery and stress state of athletes, because they meet the requirements as being accurate, economical and obtain information about the recovery-stress state of an athlete.

The first practical implication is the validation of a shorter instrument. According to Heidari et al. (2018) using a short instrument for regular monitoring can function as a direct point of reference of the recovery status of an athlete. In addition, the ARSS and SRSS seem to be adequate tools for the monitoring of athletes because they provide economical and practical methods to gain information about the different aspects/dimensions of the recovery and stress state of an athlete (Nässi et al., 2017). The SRSS is a very economical and time-efficient tool to make daily use possible because it consists of only eight items. In this way, the threshold to test the recovery-stress state is lower. This instrument can be used in applied sport scientific settings (Horvath & Röthlin, 2018), but future longitudinal studies are needed to determine it in the Dutch population of athletes.

Another point that needs to be mentioned is that it is not known what the relationship is between the athlete and the corresponding coach. According to Heidari et al. (2018), a trustful and communicative relationship between an athlete and coach can increase the likelihood of obtaining significant, and important data that can be used to improve the quality of the training load and recovery. An athlete feels trust and respect in the relationship with the coach, so therefore the athlete will give more honest answers. Also, to get more information and insight in the recovery-stress balance of an athlete, looking at biological and social factors might be useful. Both the ARSS and SRSS are not doing this. Each specific domain gives valuable information and relying on only one approach can offer a limited view on recovery and stress processes

(Heidari et al., 2018). On the other hand, the aim of this study is to validate a potential new instrument instead of indicating something about biological and social factors.

For practical use, the ARSS and SRSS can be implemented in the daily life of Dutch athletes. It depends on the research question, but the ARSS gives a more cleared image of the current recovery-stress state by responding to single adjectives. The SRSS can be used when there is little time and gives the athlete quick feedback about the current recovery-stress state, when baseline measurements are assumed. However, it is recommended to also fill out the ARSS to get familiar with the items and the survey mode. When looking at the SRSS framework, the adjectives are solely used as explanators. Both the ARSS and the SRSS are reliable and showed satisfactory results and can be used for long-term athlete monitoring.

To conclude, with the development of the ARSS and SRSS, a gap in current research is fulfilled (Kellmann et al., 2016; Nässi et al., 2017). The Dutch versions of the ARSS and SRSS are showing internal consistency, factorial validity, and construct validity, and thereby promising implementation opportunities in sports.

Acknowledgements

After finishing this study, I would like to thank NOC*NSF for taking part in the process of data collection and the athletes for completing the questionnaires. Also, I would like to thank the different scientists who helped translating the ARSS and SRSS questionnaires. Finally, I would like to thank dr. J.R. den Hartigh (University of Groningen), dr. M.S. Brink (UMCG / University of Groningen), J.J. Brauers, PhD (UMCG/University of Groningen) and B.A Prins, BSc (University of Groningen) for their feedback.

References

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, *40*, 8–34. doi:10.1007/s11747-011-0278-x

Bompa, T. O., Haff, G. G. (2009). Periodization: *Theory and Methodology of Training*. 5th ed. Champaign, IL: Human Kinetics.

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scale*. Champaign, IL: Human Kinetics.

Bresciani, G., Cuevas, M. J., Molinero, O., Almar, M., Suay, F., Salvador, A., de Paz, J. A., Marquez, S., & Gonzáles-Gallego, J. (2011). Signs of overload after an intensified training. *International Journal of Sports Medicine*, *32*, 338–343.

Brink, M. S., Visscher, C., Coutts, A. J., & Lemmink, K. A. P. M. (2012). Changes in perceived stress and recovery in overreached young elite soccer players. *Scandinavian Journal of Medicine & Science in Sports*, *22*(2), 285–292. https://doi-org.proxy-ub.rug.nl/10.1111/j.16000838.2010.01237.

Codonhato, R., Rubio, V., Oliveira, P. M. P., Resende, C. F., Rosa, B. A. M., Pujals, C., & Fiorese, L. (2018). Resilience, stress and injuries in the context of the Brazilian elite rhythmic gymnastics. *Plods ONE*, *13*(12). https://doi-org.proxy-ub.rug.nl/10.1371/journal.pone.0210174

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Sage Publications, Inc.

Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen* (4e herziene druk). Houten: Bohn Stafleu van Loghum.

Fletcher, D., & Sarkar, M. (2012). A grounded theory of psychological resilience in Olympic champions. *Psychology of Sport and Exercise, 13*(5), 669–678. https://doi-org.proxy-ub.rug.nl/10.1016/j.psychsport.2012.04.007

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *The Journal of Strength & Conditioning Research*, *15*(1), 109–115.

García Secades, X., Molinero, O., Ruíz Barquín, R., Salguero, A., De La Vega, R., & Márquez, S. (2017). Resilience and recovery-stress in competitive athletes. *Cuadernos de Psicología Del Deporte*, *17*(2), 73–80.

Hammes, D., Skorski, S., Schwindling, S., Ferrauti, A., Pfeiffer, M., Kellmann, M., & Meyer, T. (2016). Can the Lamberts and Lambert Submaximal Cycle Test (LSCT) indicate fatigue and recovery in trained cyclists? *International Journal of Sports Physiology and Performance, 11*, 328–336. doi:10.1123/ijspp.2015-0119

Hanson, S. J., McCullagh, P., & Tonymon, P. (1992). The relationship of personality characteristics, life stress, and coping resources to athletic injury. *Journal of Sport & Exercise Psychology*, *14*(3), 262–272.

Hausswirth, C., & Mujika, I. (Eds.). (2013). Recovery for performance in sport. Champaign, IL: Human Kinetics.

Heidari, J., Kölling, S., Pelka, M., & Kellmann, M. (2018). Monitoring the recovery-stress state in athletes. In M. Kellmann & J. Beckmann (Eds.), *Sport, recovery, and performance: Interdisciplinary insights*. (pp. 3–18). Routledge/Taylor & Francis Group.

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 (SRS). *Zeitschrift Für Sportpsychologie*, *22*(4), 146–161. https://doi-org.proxy-ub.rug.nl/10.1026/1612-5010/a000150

Hopkins, W. G. (2002). A new view of statistics. *Sportscience*. Retrieved from: http://www.sportsci.org/resource/stats/.

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(2), 241–248.

Hu, L., & Bertler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. doi:10.1080/10705519909540118

Jeffries, A. C., Wallace, L., Coutts, A. J., McLaren, S. J., McCall, A., & Impellizzeri, F. M. (2020). Athlete-Reported Outcome Measures for Monitoring Training Responses: A Systematic Review of Risk of Bias and Measurement Property Quality According to the COSMIN Guidelines. *International journal of sports physiology and performance*, 1–13. Advance online publication. https://doi.org/10.1123/ijspp.2020-0386

Kellmann, M., Bertollo, M., Bosquet, L., Brink, M., Coutts, A. J., Duffield, R., Erlacher, D., Halson, S. L., Hecksteden, A., Heidari, J., Kallus, K. W., Meeusen, R., Mujika, I., Robazza, C., Skorski, S., Venter, R., & Beckmann, J. (2018). Recovery and Performance in Sport: Consensus Statement. *International journal of sports physiology and performance, 13*(2), 240–245. https://doi.org/10.1123/ijspp.2017-0759

Kellmann, M., & Kallus, K. W. (2001). *Recovery-Stress Questionnaire for Athletes: User manual.* Champaign, IL: Human Kinetics.

Kellmann, M., & Kallus, K. W. (2016). Recovery-Stress Questionnaire for Athletes. In K. W. Kallus & M. Kellmann (Eds.), *The Recovery-Stress Questionnaires. User manual* (pp.86–134). Frankfurt am Main: Pearson Assessment & Information GmbH.

Kellmann, M., & Kölling, S. (2019). *Recovery and stress in sport: A manual for testing and assessment.* Routledge/Taylor & Francis Group. https://doi-org.proxy-ub.rug.nl/10.4324/9780429423857

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

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation.* Thousand Oaks, CA: Sage.

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 Sports Science & Coaching*, *10*, 529–542.

Kölling, S., Schaffran, P., Bibbey, A., Drew, M., Raysmith, B., Nässi, A., & Kellmann, M. (2020). Validation of the Acute Recovery and Stress Scale (ARSS) and the Short Recovery and Stress Scale (SRSS) in three English-speaking regions. *Journal of Sports Sciences*, *38*(2), 130–139.

Lambert, M. I. (2012). Quantification of endurance training and competition loads. In: Mujika I, ed. *Endurance Training: Science and Practice*. Vitoria-Gasteiz, Basque Country: Inigo Mujika S.L.U.;211-228.

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.

Meckel, Y., Zach, S., Eliakim, A., & Sindiani, M. (2018). The interval-training paradox: Physiological responses vs subjective rate of perceived exertion. Physiology & Behavior, 196, 144–149. https://doi-org.proxy-ub.rug.nl/10.1016/j.physbeh.2018.08.013

Meeusen, R., Duclos, M., Foster, C., Fry, A., Glesson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J., & 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.

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Nässi, A., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017). Development of two short measures for recovery and stress in sport. *European Journal of Sport Science, 17*(7), 894–903. https://doi-org.proxy-ub.rug.nl/10.1080/17461391.2017.1318180

Ohnhaus, E. E., & Adler, R. (1975). Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. *Pain, 1*, 379–384.

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

Raeder, C., Wiewelhove, T., de Paula Simola, R. A., Kellmann, M., Meyer, T., Pfeiffer, M., & Ferrauti, A. (2016). Assessment of fatigue and recovery in male and female athletes following six days of intensified strength training. *Journal of Strength and Conditioning Research, 30,* 3412–3427. doi:10.1519/JSC.0000000000001427

Raglin, J. S., Morgan, W. P., & O'Connor, P. J. (1991). Changes in mood states during training in female and male college swimmers. *International Journal of Sports Medicine, 12,* 585–589.

Saw, A. E., Kellmann, M., Main, L. C., & Gastin, P. B. (2017). Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner, *International Journal of Sports Physiology and Performance, 12*(s2), S2-127-S2-135. Retrieved Jan 20, 2021, from https://journals.humankinetics.com/view/journals/ijspp/12/s2/article-pS2-127.xml

Sharma, A. P., & Mujika, I. (2018). Quantification of training and competition loads in endurance sports: A key to recovery-stress balance and performance. In M. Kellmann & J. Beckmann (Eds.), *Sport, recovery, and performance: Interdisciplinary insights.* (pp. 132–147). Routledge/Taylor & Francis Group.

Vallerand, R. J. (1989). Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française [Toward a methodology for the transcultural validation of psychological questionnaires: Implications for research in the French language]. *Canadian Psychology*, *30*, 662–680. doi:10.1037/h0079856 van der Does, H. T., Brink, M. S., Otter, R. T., Visscher, C., & Lemmink, K. A. (2017). Injury risk is increased by changes in perceived recovery of team sport players. *Clinical Journal of Sport Medicine*, *27*(1), 46–51.

Wahl, C. A., Gnacinski, S. L., Nai, M. M., & Meyer, B. B. (2020). Psychological predictors of perceived stress and recovery in sport. *Sport, Exercise, and Performance Psychology*, *9*(3), 292–307. https://doi-org.proxy-ub.rug.nl/10.1037/spy0000175

Wiewelhove, T., Raeder, C., Meyer, T., Kellmann, M., Pfeiffer, M., & Ferrauti, A. (2016). Effect of repeated active recovery during a high-intensity interval training shock microcycle on markers of fatigue. *International Journal of Sports Physiology and Performance, 11*, 1060–1066. doi:10.1123/ijspp.2015-0494

Appendix A

Translation procedure

The process of the English translation of the German ARSS/SRSS

(As described by Kellmann et al. 2019)

Step 1. Parallel back-to-back translation (Vallerand et al. 1989) by experts in the field of sport psychology + three coaches. They all translated the items individually and met afterwards to discuss the translation.

Step 2. Several bilingual experts with different cultural backgrounds (Australian, British, North American) offered multiple back translations. There was 87% agreement for the items.

Step 3. Final call was made by experts from Australia.

Step 4. Based on the back-to-back translations, additional items were added for three scales.

Step 5. Final item wording was based on values of Cronbach's alpha and Confirmatory Factor Analysis (CFA).

Step 6. The total of 36 items was send to two experts in the field of sport science. Two items were modified based on feedback to ensure better understanding.

Step 7. The 36 items of the ARSS were reduced to 32 for the final version. The original order was used, as statistical measures to avoid sequence effects had been applied.

Step 8. Four additional items were listed at the end of the ARSS and as descriptors of the corresponding SRSS items.

Step 9. The survey links were distributed by the authors via sport institutions or clubs and sport associations.

Step 10. Each participant received info about design and content. Voluntary nature of participation was highlighted and informed consent obtained.

Step 11. Participants answered ARSS, SRSS and REST-Q or POMS and provided sociodemographic information.

Possible Dutch translation procedure

(Steps based on Vallerand et al. 1989 and Kellman et al. 2019)

Step 1 - 4. Back-to-back translation: Two bilingual (one German, one English or two of one language, depending on the recommendation of Michael Kellman) sport scientists translate the German and English version to Dutch. Two other bilingual sport scientists translate this first version back to German and English. Form a committee of four to six people including the translators, the researchers, a linguist, and author of the instrument.

- Each item and the meaning of the items of the new Dutch versions is compared to the original by the committee. When all translations are identical, the Dutch version satisfies the criteria.
- Technical words and expressions used in the Dutch language translation were scrutinized in order to verify the accuracy of their content.

Step 5. Base the final wording on Cronbach's alpha and Confirmatory Factor analysis.

Step 6. Check with experts (NOC*NSF, KNVB) if the experimental version is usable.

Step 7. Pre-test the experimental version:

Option 1: Draft target Dutch version to a group of Dutch speakers and ask them to explain why they responded as they did to individual items.

Option 2: Ask six athletes -three men and three women- to circle any expression or word that was equivocal. Discuss these expressions and words in the committee until consensus is reached. *(preferred option)*

Step 8. Analyze the cross-correlations which represent the temporal interlanguage relations, thus offering a quality index for translation. The two correlations should be statistically equal. An inequality could indicate an interaction between the time delay and the original language of the instrument.

Step 9. Repeat the steps 9-11 that were described by Kellmann et al. 2019.

For the Dutch version: include athletes from all levels via NOC*NSF, KNVB and sport clubs in Groningen. Provide information about the study, obtain informed consent, and provide the participants with the ARSS, SRSS, REST-Q or POMS.