Chapter 6

Reliability and validity of the Dutch recovery stress questionnaire for athletes

This chapter is submitted as:
Nederhof, E., Brink, M.S., Lemmink, K.A.P.M. Reliability and validity of the Dutch Recovery-Stress Questionnaire for Athletes. *International Journal of Sport Psychology.*

Acknowledgements:
The authors kindly thank Wytske Altenburg, Hillona Beugels, Ida Buist, Marieke de Boer and Inge Luttikhuizen for their contribution to the study.

The purpose of the present study was to investigate the cross-cultural validity of the Recovery Stress Questionnaire for Athletes (RESTQ-sport) by analysing reliability and validity of a Dutch translation. Two studies were performed to assess test-retest reliability, internal consistency and factor structure. Criterion validity was assessed in the first study only, with the Profile of Mood States as criterion measure. The test-retest reliability of the Dutch RESTQ-sport was acceptable, especially as the RESTQ-sport aims to measure stress and recovery states. Internal consistency was good for most scales. In both studies, internal consistency was higher at the second compared to the first measurement. Factor analyses confirmed the stress-recovery structure of the Dutch RESTQ-sport. Criterion validity was also supported. Overall, it was concluded that the Dutch RESTQ-sport has sufficient reliability and validity. This gives support to the cross-cultural usefulness of the scale.

Introduction

Monitoring stress and recovery has become increasingly important in sports. A disturbed stress-recovery balance can increase an athlete's vulnerability for injuries, overtraining or burn-out (Tenenbaum et al., 2003b; Williams & Andersen, 1998). The Recovery Stress Questionnaire for Athletes (RESTQ-sport) is a recently developed questionnaire that can be used for this purpose (Kellmann & Kallus, 2001). The RESTQ-sport is an extension of the general Recovery Stress Questionnaire (Kallus, 1995), which exists of 12 stress and recovery scales. Three sport-specific stress scales and four sport-specific recovery scales were added to the original questionnaire in the version for athletes. The questionnaire focuses on the frequency of stress and recovery related behaviours (e.g. *I had a good time with friends*) as well as on psycho-physical states (e.g. *I was in a bad mood*).

The RESTQ-sport has predominantly been used in endurance athletes such as rowers (e.g. Kellmann et al., 2001; Purge et al., 2006), cyclists (Bouget et al., 2006; Chapter 3 of this thesis), and triathletes (Coutts et al., 2006). The RESTQ-sport is currently being used in research projects with soccer players as well.

The RESTQ-sport was successfully used to study perceived stress and recovery states in monitoring studies with adolescent (e.g. Kellmann et al., 2001) and adult athletes (e.g. Purge et al., 2006). In other studies the RESTQ-sport has been used to measure the effect of high load training on perceived stress and recovery (e.g. Bouget et al., 2006; Coutts et al., 2006). The questionnaire has been used successfully in different countries (e.g., France: Bouget et al., 2006; Australia: Coutts et al., 2007; Germany: Kellmann et al., 2001; the Netherlands: Chapter 3 of this thesis; Estonia: Purge et al., 2006) which can be seen as indirect proof of its cross-cultural usefulness.

However, the RESTQ-sport is not all good news. Davis et al. (Davis et al., 2007) performed a validation study on the individual items of the questionnaire. Davis and colleagues could not confirm the 19 subscales of the RESTQ-sport. This type of validation had not been performed by the original authors (Kellmann & Kallus, 2001).

The purpose of the present study was to examine cross-cultural usefulness by analysing reliability and validity of the scales in the Dutch version. As the RESTQ-sport has shown to be a useful tool in research and practice the authors decided to validate the Dutch version following the procedures of the original psychometric evaluation.

Study 1

The purpose of the first study was to test initial reliability and validity of the Dutch RESTQ-sport. The questionnaire was translated from English into Dutch by a native Dutch speaker who was an expert in English as well as in the field of sports. The translation was checked by two other experts of whom one was a native English speaker. Consistency and test-retest reliability of the translated questionnaire were determined as well as construct and criterion validity. The RESTQ-sport was compared to the Profile of Mood States (POMS) for the determination of criterion validity.

Method

Subjects

Participants in the first study were 116 athletes (59 male, 57 female) with a mean age of 23.1 (SD = 3.6) years. The mean amount of training hours in the week before the first data collection was 6.8 (SD = 3.2) hours. Participants trained for basketball (n = 11), korfball (n = 16), rowing (n = 11), speed skating (n = 31) and volleyball (n = 47).

Instruments

Recovery Stress Questionnaire for Athletes (RESTQ-sport).

The RESTQ-sport consists of 77 questions. Questions are answered on a seven point Likert type scale. Questions have a stem that specifies the period the athlete should look back. In the present study the questionnaire asks about activities in the past week. The first question is a practice question that does not go into the analysis. The 19 scales of the RESTQ-sport consist of the other 76 questions. Each scale consists of four questions. There are 12 general and 7 sport-specific scales of which seven scales measure general stress, five general recovery, three sport-specific stress and four scales measure sport-specific recovery (Kellmann & Kallus, 2001).

Test-retest reliability of the RESTQ-sport is higher for shorter measurement intervals. Test-retest reliability lies above .7 for most scales over three days, it declines to around .5 over 9 days. Longer measurement intervals give even lower test-retest reliabilities (Kellmann & Kallus, 2001). This finding is consistent with the construct the questionnaire measures, stress and recovery vary over time.

Internal consistency of the RESTQ-sport is good for most scales with Cronbach's alpha of .7 and higher. Although differences between samples and fluctuations over time are present in nearly all scales, in some samples insufficient reliability has been found for the scales *Conflicts/ Pressure*, *Lack of Energy*, *Success* and *Disturbed Breaks* (Kellmann & Kallus, 2001).

Construct validity of the RESTQ-sport has been shown to be good. Both the general and the sport specific scales loaded on either the factor stress or on the factor recovery. In some samples some scales loaded negatively on the other factor as well (Davis et al., 2007; Kellmann & Kallus, 2001).

Criterion validity with the POMS as criterion measure has also been shown to be good. Correlations were all in the expected direction. The stress scales correlated positively with negative mood states and the recovery scales correlated positively with the positive mood state and vice versa (Kellmann & Kallus, 2001).

Profile of Mood States (POMS).

The Dutch POMS is a shortened version of the original POMS (McNair et al., 1971) and consists of the five scales *Depression*, *Anger*, *Fatigue*, *Vigour* and *Tension*. It consists of 32 items that are answered on a five point Likert scale. The Dutch POMS has shown to have good reliability and validity (Wicherts & Vorst, 2004).

Procedure

Coaches and trainers were asked for permission to contact their athletes before or after a training session. If permission was given the procedure was explained to the athletes and they were asked if they agreed on participation. The athletes who agreed filled out the RESTQ-sport and the POMS before or after a regular training session. The second measurement occasion took place exactly one week after the first. This time only the RESTQ-sport was filled out. Procedures of the study were in accordance with ethical standards of the Helsinki declaration.

Statistical analysis

If more than one out of four values within a scale were missing, participants were excluded from further analyses. Otherwise, the missing value was replaced with the mean score of the corresponding scale.

Both absolute and relative test-retest reliability were calculated. Absolute reliability was calculated with the Bland and Altman (1999) method. The mean difference between the scores on both days was calculated. The 95% confidence interval to the mean difference was calculated as follows:

$$95\%CI = \bar{d} - t \times \left(s_d/\sqrt{n}\right)$$

With $\bar{d}$ the mean difference, t the critical *t* value for the number of degrees of freedom and s_d the standard deviation of $\bar{d}$. If zero lays within the 95% confidence interval, it was concluded that no bias existed. Intraclass correlation coefficient (ICC) was calculated for determination of relative test-retest reliability. Cronbach's alpha was calculated for internal consistency. A maximum likelihood analysis with oblique rotation and two fixed factors was done to examine the stress-recovery factorial structure. Factor loadings smaller than .30 were omitted from the tables for increased clarity. Finally, Pearson's correlations between RESTQ-sport scales and POMS scales were calculated for criterion validity. All analyses were performed using SPSS (version 11.01).

Results

The first questionnaire of one of the subjects was excluded from further analyses, because too many data were missing. Eighty-eight participants took part in both measurements

Absolute and relative test-retest reliability were sufficient for most scales (Table 6.1). The scale *Success* had a mean difference between the two test scores significantly different from zero, as the 95% confidence interval of the mean difference (Bland & Altman, 1999) of the scale *Success* ranged from 0.02 to 0.33. This means that a test-retest bias was present. The scales *Physical Complaints* and *Sleep Quality* showed poor relative test-retest reliability with ICC below .50.

Internal consistency was sufficient for most scales (Table 6.1). Cronbach's alpha for the first sample was above .60, except for the scales *Conflicts/Pressure* and *Self-regulation*. In the second sample only the scale *Conflicts/Pressure* was below .60. In general, internal

Table 6.1. Absolute and Relative Test-Retest Reliability and Internal Consistency for the Preliminary Version of the Dutch RESTQ-sport.

		Bland and Altman (*N*=88)				ICC (*N*=88)			Cronbach's Alpha	
	Scale	*d*	Sd	95% CI		ICC	95% CI		T1 (*N*=115)	T2 (*N*=89)
1	General Stress	-0.09	0.75	-0.24	0.07	.66	.52	.76	.84	.88
2	Emotional Stress	0.00	0.78	-0.17	0.16	.53	.36	.66	.81	.86
3	Social Stress	0.01	0.77	-0.16	0.17	.55	.38	.68	.74	.84
4	Conflicts/Pressure	0.01	0.71	-0.14	0.16	.56	.40	.69	.47	.55
5	Fatique	0.05	0.98	-0.16	0.26	.54	.37	.67	.75	.71
6	Lack of Energy	-0.12	0.74	-0.27	0.04	.51	.34	.65	.62	.70
7	Physical Complaints	0.07	0.89	-0.12	0.26	.37	.18	.54	.69	.75
8	Success	0.17	0.73	0.02	0.33	.54	.38	.68	.61	.67
9	Social Recovery	0.05	0.80	-0.12	0.22	.55	.39	.68	.78	.78
10	Physical Recovery	-0.07	0.77	-0.23	0.09	.54	.38	.68	.67	.76
11	General Well-Being	0.07	0.73	-0.09	0.22	.63	.48	.74	.88	.90
12	Sleep Quality	0.00	0.58	-0.12	0.13	.44	.26	.60	.80	.75
13	Disturbed Breaks	-0.03	0.75	-0.19	0.13	.54	.37	.67	.72	.83
14	Emotional Exhaustion	-0.01	0.80	-0.18	0.16	.62	.47	.73	.67	.76
15	Injury	0.14	0.92	-0.05	0.34	.64	.50	.75	.70	.77
16	Being in Shape	-0.08	0.89	-0.27	0.11	.67	.54	.77	.91	.90
17	Personal Accomplishment	0.03	0.85	-0.16	0.21	.54	.38	.67	.64	.74
18	Self-Efficacy	-0.14	0.82	-0.32	0.03	.67	.53	.77	.78	.87
19	Self-regulation	-0.02	0.94	-0.22	0.18	.60	.45	.72	.53	.76

Table 6.2. Maximum Likelihood Factor Analysis of the Preliminary Version of the Dutch RESTQ-sport.

	Scale	T1 (N = 115)		T2 (N = 89)	
		factor 1	factor 2	factor 1	factor 2
1	General Stress	.71	-.30	.75	
2	Emotional Stress	.75		.88	
3	Social Stress	.71		.90	
4	Conflicts/Pressure	.61		.68	
5	Fatigue	.69		.69	
6	Lack of Energy	.69		.59	
7	Physical Complaints	.72		.78	
8	Success		.51		.70
9	Social Recovery		.57		.60
10	Physical Recovery	-.42	.65	-.42	.62
11	General Well-Being	-.32	.67	-.35	.63
12	Sleep Quality		.32		.33
13	Disturbed Breaks	.71		.42	
14	Emotional Exhaustion	.67		.61	
15	Injury	.39		.51	
16	Being in Shape	-.33	.57		.69
17	Personal Accomplishment		.50		.69
18	Self-Efficacy		.72		.68
19	Self-regulation	.30	.60	.44	.60
	Eigenvalue	5.55	3.80	6.07	4.41

consistency was better for the second measurement.
Factor structure of the Dutch RESTQ-sport was comparable to the original version. The stress recovery factor structure showed to be the strongest structure (Table 6.2). All general and sport-specific stress scales loaded on the first and all general and sport-specific recovery scales loaded on the second factor. The factor structure of the second measurement was generally stronger than the factor structure of the first measurement. The scales *General Stress*, *Physical Recovery*, *General Well-Being* and *Being in Shape* gave additional negative loadings to the other factor. The scale *Self-Regulation* loaded positively on both scales.
Criterion validity was good for most scales (Table 6.3). Stress scales of the RESTQ-sport correlated positively with the negative mood scales of the POMS and negatively with the positive mood scale. Recovery scales of the RESTQ-sport correlated positively with the positive mood scale and negatively with the negative mood scales of the POMS.

Study 2

All items of the scales with insufficient reliability or validity in study 1 were evaluated and modified if appropriate. Additional adjustments were made according to our experience with the preliminary Dutch RESTQ-sport. In total, 13 out of 76 items were modified. The purpose of the second study was to evaluate reliability and validity of the improved version of the Dutch RESTQ-sport.

Method

Subjects, Instrument and Procedure

Participants were 123 athletes (66 male, 57 female) who participated in basketball (n=14), gymnastics (n=11), handball (n=13), rowing (n=53) and soccer (n=32). Their mean age was 21.9 (±2.5) years and they practiced on average 8.6 (±4.6) hours a week. The subjects filled out the improved version of the Dutch RESTQ-sport. The same procedures as in study 1 were followed.

Statistical Analysis

The same analyses were performed as in study 1. Absolute test-retest reliability was calculated with the Bland and Altman (1999) method. ICC's were calculated for relative test-retest reliability. Two factor structure was studied using a confirmatory maximum likelihood factor analysis with oblique rotation.

Results

The second questionnaire of one of the subjects was excluded from analysis, because too many data were missing. 90 Subjects participated in both measurements.
Test-retest reliability can be found in Table 6.4. Absolute test-retest reliability was insufficient for the scales *Conflicts/Pressure*, *Lack of Energy*, *Success*, *Physical Recovery* and *Emotional Exhaustion*. Relative test-retest reliability was insufficient for the scales *General Stress*, *Emotional Stress*, *Social Stress*, *Physical Complaints*, *Success*, *General*

Table 6.3. Pearson's Correlations Between the POMS and the Preliminary Version of the Dutch RESTQ-sport.

	Scale	Depression	Anger	Fatigue	Vigour	Tension
1	General Stress	0.78**	0.44**	0.64**	-0.56**	0.47**
2	Emotional Stress	0.65**	0.68**	0.56**	-0.52**	0.39**
3	Social Stress	0.52**	0.68**	0.39**	-0.20*	0.30**
4	Conflicts/Pressure	0.45**	0.35**	0.44**	-0.21*	0.47**
5	Fatique	0.35**	0.37**	0.55**	-0.45**	0.31**
6	Lack of Energy	0.55**	0.42**	0.50**	-0.45**	0.41**
7	Physical Complaints	0.53**	0.40**	0.64**	-0.51**	0.29**
8	Success	-0.23*	0.01	0.01	0.31**	0.06
9	Social Recovery	-0.33**	-0.07	-0.15	0.44**	-0.21*
10	Physical Recovery	-0.60**	-0.35**	-0.47**	0.69**	-0.33**
11	General Well-Being	-0.62**	-0.32**	-0.40**	0.69**	-0.30**
12	Sleep Quality	-0.30**	-0.30**	-0.17	0.26**	-0.30**
13	Disturbed Breaks	0.32**	0.32**	0.45**	-0.22*	0.25**
14	Emotional Exhaustion	0.52**	0.44**	0.53**	-0.42**	0.27**
15	Injury	0.25**	0.20*	0.38**	-0.13	0.11
16	Being in Shape	-0.49**	-0.22*	-0.49**	0.60**	-0.23*
17	Personal Accomplishment	-0.32**	-0.23*	-0.06	0.16	-0.24*
18	Self-Efficacy	-0.44**	-0.15	-0.30**	0.57**	-0.21*
19	Self-regulation	0.00	0.17	0.12	0.21*	0.03

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

TABLE 6.4. ABSOLUTE AND RELATIVE TEST-RETEST RELIABILITY AND INTERNAL CONSISTENCY OF THE DUTCH RESTQ-SPORT.

		BLAND AND ALTMAN (N = 90)				ICC (N = 90)			CRONBACH'S ALPHA	
	SCALE	$\bar{d}$	SD	95% CI	$\bar{d}$	ICC	95% CI		T1 (N=123)	T2 (N=90)
1	GENERAL STRESS	0.02	0.86	-0.17	0.20	.35	.16	.52	.84	.84
2	EMOTIONAL STRESS	0.15	0.79	-0.02	0.31	.43	.24	.58	.77	.77
3	SOCIAL STRESS	0.13	0.68	-0.02	0.27	.48	.30	.62	.75	.72
4	CONFLICTS/PRESSURE	0.23	0.81	0.05	0.40	.60	.45	.72	.62	.73
5	FATIQUE	0.01	0.90	-0.18	0.20	.53	.37	.67	.78	.79
6	LACK OF ENERGY	0.21	0.73	0.05	0.36	.50	.33	.64	.60	.70
7	PHYSICAL COMPLAINTS	0.08	0.66	-0.06	0.21	.44	.26	.59	.58	.53
8	SUCCESS	0.25	0.74	0.09	0.41	.40	.21	.56	.56	.59
9	SOCIAL RECOVERY	0.03	0.66	-0.11	0.17	.70	.58	.79	.74	.84
10	PHYSICAL RECOVERY	-0.17	0.71	-0.32	-0.02	.51	.35	.65	.64	.70
11	GENERAL WELL-BEING	0.04	0.86	-0.14	0.22	.49	.32	.63	.87	.89
12	SLEEP QUALITY	-0.09	1.04	-0.31	0.13	.40	.21	.56	.83	.80
13	DISTURBED BREAKS	0.03	0.66	-0.12	0.17	.60	.44	.71	.73	.83
14	EMOTIONAL EXHAUSTION	-0.19	0.78	-0.36	-0.03	.52	.35	.65	.67	.72
15	INJURY	0.14	0.80	-0.03	0.31	.68	.56	.78	.73	.74
16	BEING IN SHAPE	-0.07	0.80	-0.24	0.10	.68	.55	.77	.85	.84
17	PERSONAL ACCOMPLISHMENT	0.14	0.88	-0.05	0.33	.50	.33	.64	.63	.69
18	SELF-EFFICACY	-0.04	0.75	-0.20	0.12	.69	.56	.78	.76	.80
19	SELF-REGULATION	-0.10	0.83	-0.27	0.08	.61	.46	.73	.56	.71

Table 6.5. Maximum Likelihood Factor Analysis of the Dutch RESTQ-sport.

	Scale	T1 (n=123)		T2 (n=90)	
		Factor 1	Factor 1	Factor 1	Factor 2
1	General Stress	.86	.73	.73	
2	Emotional Stress	.93	.86	.86	
3	Social Stress	.89	.78	.78	
4	Conflicts/Pressure	.66	.77	.77	
5	Fatique	.52	.47	.47	
6	Lack of Energy	.68	.56	.56	
7	Physical Complaints	.50	.62	.62	
8	Success				.58
9	Social Recovery				.60
10	Physical Recovery				.74
11	General Well-Being	-.47	-.38	-.38	.60
12	Sleep Quality	-.55	-.58	-.58	.29
13	Disturbed Breaks	.27	.59	.59	
14	Emotional Exhaustion	.51	.68	.68	
15	Injury	.18	.22	.22	
16	Being in Shape				.78
17	Personal Accomplishment				.68
18	Self-Efficacy				.79
19	Self-regulation		.39	.39	.56
	Eigenvalue	5.73	5.89	5.89	4.77

Well-Being and *Sleep Quality* with ICC's below .50.
Internal consistency was good for most scales, except *Physical Complaints* and *Success* in both samples and *Self-Regulation* in the first sample.
The factor structure was good for both samples (Table 6.5). All general and sport-specific stress scales of the Dutch RESTQ-sport loaded on the first factor. All general and sport-specific recovery factors loaded on the second factor. The scales *General Well-Being* and *Sleep Quality* had additional negative factor loadings on the other factor. *Self-Regulation* loaded positively on both scales in the second sample

Discussion

The purpose of the study was to examine cross-cultural usefulness of the RESTQ-sport by analysing reliability and validity of the Dutch translation. Both absolute and relative test-retest reliability were acceptable in both studies. It is difficult to draw conclusions from these reliability analyses, as the RESTQ-sport is a state oriented questionnaire (Kellmann & Kallus, 2001). Indeed, test-retest reliabilities of the original RESTQ-sport decreased with larger test-retest intervals. ICC's in the present study were comparable to those of the original version. It is noteworthy that test-retest reliability was lower in the second study compared to the first study of the present paper. Another notable finding is that absolute test-retest bias was found in some scales whereas relative test-retest reliability scores were low for other scales. This shows the importance of performing both absolute and relative test-retest analyses.
The other aspect of reliability that was tested in this study, internal consistency, was also comparable to the original version. In general, internal consistency was better in the second compared to the first measurement. This probably reflects the higher level of familiarity with the instrument. Scales with Cronbach's alphas below .60 in the Dutch RESTQ-sport also scored low in the original version. The only exception is the scale for Lack of Energy which scored below .60 in our second study. Thus, cross-cultural reliability of the questionnaire is good.
Criterion validity of the Dutch RESTQ-sport was good. Stress scales correlated positively with negative mood scales of the POMS, while recovery scales correlated negatively with the negative mood scales. The opposite was true for the positive mood scale of the POMS. Moderate correlations between the RESTQ-sport and the POMS show that similar but not identical constructs are measured. The correlations were all in the expected direction, except for the scale Self-Regulation.
The construct validity of the Dutch RESTQ-sport was also good. The factor structure of the Dutch RESTQ-sport clearly distinguished the stress scales from the recovery scales. Indeed, the stress recovery structuring was the strongest factor structure in the original version as well (Davis et al., 2007; Kellmann & Kallus, 2001). This gives support for the cross-cultural validity of the questionnaire.
As was the case with internal consistency, the factor structure was stronger for the second measurement. Apparently, it is important for athletes to become familiar with the RESTQ-sport for better reliability and validity. Like for the original version, it is our advise to always familiarise athletes with the RESTQ-sport before using it in research or in

practice.
The overall conclusion of the present paper is that the Dutch RESTQ-sport has sufficient reliability and validity for use in sports practice and research. The RESTQ-sport seems to be a tool that has cross-cultural validity.