Note. This article will be published in a forthcoming issue of the *International Journal of Sports Physiology and Performance*. The article appears here in its accepted, peer-reviewed form, as it was provided by the submitting author. It has not been copyedited, proofread, or formatted by the publisher.

Section: Invited Commentary

Article Title: Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner

Authors: Anna E. Saw[1], Michael Kellmann[2, 3], Luana C. Main[4], and Paul B. Gastin[1]

Affiliations: [1]Centre for Sport Research, Deakin University, Australia. [2]Faculty of Sport Science, Ruhr-University Bochum, Germany. [3]School of Human Movement and Nutrition Sciences, University of Queensland, Australia. [4]Institute for Physical Activity and Nutrition, Deakin University, Australia.

Journal: *International Journal of Sports Physiology and Performance*

Acceptance Date: October 12, 2016

DOI: http://dx.doi.org/10.1123/ijspp.2016-0395

“Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner” by Saw AE, Kellmann M, Main LC, Gastin PB]
International Journal of Sports Physiology and Performance

Athlete self-report measures in research and practice: Considerations for the discerning reader and fastidious practitioner

Commentary

A. E. Saw[1], M. Kellmann[2, 3], L. C. Main[4], P. B. Gastin[1]

1 Centre for Sport Research, Deakin University, Australia

2 Faculty of Sport Science, Ruhr-University Bochum, Germany

3 School of Human Movement and Nutrition Sciences, University of Queensland, Australia

4 Institute for Physical Activity and Nutrition, Deakin University, Australia

Corresponding author:

Anna E. Saw

Centre for Sport Research

School of Exercise and Nutrition Sciences, Deakin University

221 Burwood Highway, Burwood, Victoria, Australia, 3125

anna.saw@deakin.edu.au

Running head: Athlete self-report measures in research and practice

Abstract word count: 115

Text-only word count: 3853

Number of figures: 2

Number of tables: 1

“Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner” by Saw AE, Kellmann M, Main LC, Gastin PB]
International Journal of Sports Physiology and Performance

Abstract

Athlete self-report measures (ASRM) have the potential to provide valuable insight into the training response, however there exists a disconnect between research and practice which needs to be addressed. Namely, the measure or methods used in research are not always reflective of practice, or data primarily obtained from practice lacks empirical quality. This commentary reviews existing empirical measures, and the psychometric properties required to be considered acceptable for research and practice. This information will allow discerning readers to make a judgement on the quality of ASRM data being reported in research papers. Fastidious practitioners and researchers are also provided with explicit guidelines for selecting and implementing an ASRM, and reporting these details in research papers.

Introduction

Modern sport has adopted an increasingly scientific approach to athletic preparation by means of athlete monitoring[1], with high-performance sport programs investing substantial human and financial resources in the process[2]. The challenge is to carefully monitor and manage the load and recovery of an athlete to optimise their performance capacity, and avoid deleterious outcomes such as underperformance, injury, or illness. One particular method of athlete monitoring which has gained considerable popularity in recent years is athlete self-report measures (ASRM)[3]. Athlete self-report measures are paper-based or electronic records of an athlete's perceived physical, psychological, and/or social wellbeing, completed on a regular, often daily, basis.

The utility of ASRM for athlete monitoring is well-supported[4]. However, this support is limited to established measures in the literature. Applied practice tends to favour brief, custom measures over empirical measures[3], thus creating a disconnect between research and practice. Applied research is attempting to bridge this gap by utilising data from custom ASRM to answer practical problems. For instance, retrospective analysis of data which extends across one or more seasons may provide valuable insight into the training response and its evaluation. However, this approach requires careful consideration from researchers, practitioners, and readers of applied research.

Deviating from the principles of basic research faces the risk of dubious or flawed concepts and methods becoming repeatedly adopted and published[5]. This is particularly evident as the use of custom ASRM is often solely justified as 'similar to that used previously'. Therefore, to avoid the field progressing down an undesirable path, it is a critical time for applied research using ASRM to ensure scientific rigour is upheld.

The purpose of this commentary is not to curtail the current enthusiasm for ASRM research, but to sensitise the scientific community to potential shortcomings and to offer a potential solution to help advance the field. Specific guidance is provided to practitioners and researchers to improve ASRM selection, data quality and analysis, and the standard of reporting in research papers.

Review of empirical athlete self-report measures

Empirical ASRM assess one or more of the dimensions[*] of mood, stress, recovery, symptoms, and emotions (Table 1). Scales and items may relate to general wellbeing (e.g., I was fed up with everything[6]), or be sport-specific (e.g., I felt frustrated by my sport[6]). It may be argued that sport-specific items are more applicable to, and better received by, athletes. Alternatively, the argument that the training response also manifests as general wellbeing signs and symptoms suggest general items or a composite measure may be more appropriate for athlete monitoring.

Research using ASRM to monitor the training response gained momentum in the 1980s, with a series of studies[7] measuring mood disturbance using the Profile of Mood States (POMS)[8]. The POMS has consistently been shown to respond in a dose-response manner with training load[4], in addition to identifying overtrained athletes[9]. Shorter[10-14] and sport-oriented[9,15] derivatives of the POMS have also been developed and used to monitor athletes.

The Recovery-Stress Questionnaire for Athletes (RESTQ-Sport)[6,16] has become widely popular in research over the previous decade. A dose-response relationship between stress and training load, and an inverse relationship between recovery and training load, has been

[*] Terminology related to the composition of ASRM may differ between measures, hence for the purposes of this commentary, dimensions are the overarching constructs being assessed. Within a measure, dimensions are assessed by scales, which may comprise multiple items.

observed in various athletes[4]. A similar response has also been observed for symptoms of stress measured by the Daily Analyses of Life Demands for Athletes[17].

The assessment of symptoms provides useful clinical insight into physical (e.g., headache, sore throat, digestive complaints) or behavioural (e.g., ability to work, eating habits) manifestations of the training response. However, in practice, it may be more advantageous to monitor mood disturbance and perceived stress and recovery in order to detect maladaptation prior to the manifestation of symptoms[18].

Findings from a recent systematic review demonstrated that measures of subjective mood disturbance, perceived stress and recovery, and symptoms of stress were similarly responsive to acute changes in training load[4]. Measures of perceived stress and recovery were also responsive to chronic training load. Therefore, there is no one dimension which may be recommended as the best option for athlete monitoring. This decision should be based upon the intended purpose for implementing an ASRM, and practicalities of the sport context. Additionally, empirical measures are not all equal in their ability to measure a certain dimension, hence consideration should also be given to the psychometric properties of a measure.

Psychometric properties

For an ASRM to be highly regarded and accepted in the literature, the development and psychometric properties of the measure must be openly documented (i.e., published paper or manual). This documentation may evolve as the measure is refined and additional validity, reliability, and reference values are obtained from different athlete populations and contexts. Key psychometric considerations are outlined below, and illustrated using the RESTQ-Sport[6,16] as an example.

Theoretical basis

An ASRM intends to evaluate a particular outcome of athletic preparation. The relevance of this outcome, and process by which it occurs, should be based upon sound theory. A single theory may be adopted, perhaps focused on a specific dimension, or multiple theories may be adopted to explain the complex interrelations of dimensions. For instance, the RESTQ-Sport has a basis in theories on physiological and psychological responses to stress and recovery, and their interaction, summarised by the 'scissor model'[19]. The sport-specific component of the questionnaire also draws upon theories of burnout and self-efficacy[6,16].

Instrument development

The selection and refinement of ASRM items must go through several phases. An initial pool of items and scales may be drawn from theories, related literature, other measures, and experts in the field. Exploratory Factor Analysis is used to refine and reduce the number of items by determining whether scales are related (load cleanly), or are distinct (do not load cleanly). The RESTQ-Sport initially consisted of 86 items, and proceeded through iterations of 85 and 80 items before the final RESTQ-Sport-76[6], RESTQ-Sport-52[6], and the new RESTQ-Sport-36[16] versions were published[6].

Confirmatory Factor Analysis is subsequently used to test the hypothesised relationship of scales and dimensions. This hypothesis should be grounded in the theoretical basis of the measure. For the RESTQ-Sport-76, analytical models support the theoretical separation of stress and recovery dimensions, although some correlations between scales provides a better model fit[16].

Reliability

Reliability refers to the consistency or precision of measurements. It is important to be aware that the degree of reliability impacts the precision for individual interpretation (confidence interval) and hence the quality of feedback. Ultimately, the more error contributed by various sources (i.e., the individual and their environment), the less reliable and sensitive a measure is for assessing the athlete state. The internal consistency of multiple items of a scale is typically reported as a Cronbach's alpha coefficient, with 0.7 or above considered acceptable[†]. Internal consistency tends to increase as athletes become more familiar with a measure, raising the importance of including or allowing for a period of familiarisation in a measurement protocol.

Test-retest reliability is of particular importance for detecting true change in the athlete state over time. A reliable state-oriented measure should possess a high degree of stability (Pearson's correlation (*r*) 0.7 or above[†]) over the short-term, however such stability is undesirable over the longer-term when the athlete state is anticipated to change. Repeated measures of the RESTQ-Sport have shown sufficient reliability in the short term (r = 0.79), with the strength of correlations decreasing over three or more days, reflecting a sensitivity to changes in the recovery-stress state over time[16].

Validity

Validity is not a property of an ASRM *per se*, but an assessment of whether the data obtained is appropriate, meaningful, and useful for a specific purpose. Construct validity is dependent upon several components of validity[20]. The phases of instrument development provide subjective assurance that items measure the specific dimension of interest. Validity may also be calculated from the relationship with another measure which is known or assumed to be valid, or inferred from more operational criteria. For instance, the RESTQ-Sport claims validity if the measure: is sensitive to changes in training and competition cycles; allows prediction of the intensity of current training conditions on the basis of past situations; is related

[†] This value is considered acceptable for research or analysis of group data, however higher values (i.e., $\alpha \geq 0.9$ or $r \geq 0.9$) may be more appropriate for individual diagnosis or decision making[20].

to performance, and; represents systematic changes in stress states[6]. However, it should be noted that assessment of such validity is complex, as each criterion also possesses a degree of error.

Reference values

Reference values are used to place data in context according to specific populations (norm-referenced) or outcomes (criterion-referenced). Therefore, it is preferable for an ASRM to have reference values for different athlete cohorts and contexts, and also for different training states. For instance, the RESTQ-Sport manual includes individual-specific profiles of different recovery-stress states[16]. These examples highlight the importance of considering the relative profile of two or more scales, or the pattern of responses over time, rather than a single critical value for an item or scale.

Recommendations

It is possible for an ASRM to be successfully implemented for both practice and applied research. This hinges on careful planning and strategies being put in place to support good quality data. Sports programs should therefore take the time to work through the following considerations to establish the needs and limitations of the sport context, select a measure, and to develop a culture which supports ASRM use. Recommendations are also provided to optimise data analysis and the reporting of ASRM data in research papers.

Establish needs and limitations of sport context

There are many questions which a practitioner or researcher should consider to determine whether or not it is appropriate to implement an ASRM in their particular sport context. A logical progression of these questions is presented in Figure 1.

Purpose and stakeholder engagement

The decision to implement an ASRM should be based upon an identified need, and the capacity for an ASRM to meet that need. This need and capacity, and ideally a working hypothesis, should be stablished through conversation with the coaching team. There should also be a clear intention for how the data obtained will be used (i.e., to inform practices on a day-to-day basis, and/or with a longer-term perspective of improving understanding of athletic preparation for the future). Commitment of the coaching team to the process is of utmost importance, whilst efforts should also be invested in developing the buy-in of athletes, support staff, and the sports organisation[21].

Feasibility

The feasibility of ASRM use relates to the time and effort required of athletes and staff, in addition to the associated financial costs. While there may be an initial willingness to invest considerable resources in the process, the ongoing success of an ASRM is more likely if costs are minimised, and are justified by the benefits provided to all stakeholders.

If the intended purpose of an ASRM does not align with available resources, it may be necessary for the sports organisation to acquire additional resources (e.g., staff, funding), or reallocate resources. Practitioners should have a portion of their contracted time allocated to overseeing data input, following up with athletes and other practitioners with specific expertise (e.g., sport psychologist), analysing and interpreting the data, and providing feedback to athletes and coaches. In addition to practitioner hours, financial investment may also include the purchase/development and upkeep of specific software if desired. Investment in software may help offset the cost of practitioner hours by automating much of the day-to-day processes such as athlete reminders, analysis and alerts, and basic feedback.

Select a measure

Three steps are required to select a measure which will yield quality and meaningful data: determine the dimensions to assess; choose a theoretical model, and; select a psychometrically sound measure[5]. Each step is complex, and each question outlined in Figure 2 should be answered with clarity before proceeding.

Dimensions

One or more distinct dimensions may be assessed, depending on the intended purpose of the ASRM. Dimensions should be responsive to changes in acute and chronic training load, and be related to athlete wellbeing and/or performance[4]. Selecting multiple dimensions allows for variable athlete responses, and avoids emphasising one aspect of the training response over others.

Practitioner involvement may also dictate what dimensions are assessed by an ASRM. For instance, dimensions related to a practitioner's expertise may assist them to do their job. Conversely, it should be considered whether it is appropriate to assess a dimension for which there is lack of expertise amongst those involved. For instance, if a sport psychologist is not part of the support team, some may question whether psychometric items can be effectively interpreted and acted upon[22].

Theoretical basis

The selection of appropriate dimensions may also form part of an iterative process with the selection of a theoretical model. It is important for the practitioner to understand the strengths and limitations of the chosen model, including the evidence for and against. Theoretical models are often the subject of academic debate spanning decades, hence the practitioner must be able to justify their decision.

Empirical or custom measure?

The selection of a measure should be based upon the chosen dimensions and theory, in light of previously determined feasibility. Choosing an existing empirical measures is recommended, however this option is not free of further work. It is the responsibility of the practitioner to ensure reliability and validity are upheld in their particular sport context[23].

Developing a new custom ASRM is no small task, and the decision to take this approach deserves due consideration and commitment. The development process should be guided by the chosen theoretical model, and be supported by a clear rationale (e.g., assessing a more applicable dimension, being shorter and easier to administer, a better predictor of outcomes, or more sensitive to change compared to existing measures). A measure which is more specific to a particular sport or sport context is not necessarily more relevant or sensitive to change[5]. Hence any purported benefit over existing measures must be supported by evidence.

Several phases of development and refinement are required to be completed in order to ensure acceptable psychometric properties, as outlined previously. The availability of large athlete cohorts (N = 50-500[24]) to partake in instrument development, and assessment of other measures (e.g., training load) to establish validity, must also be planned for.

To date, the composition of custom ASRM in the applied setting are typically informed by empirical measures (e.g., POMS, RESTQ-Sport), personal experience, and the recommendations of Hooper and Mackinnon[25]. If a custom measure is to be largely based upon an empirical measure, it is recommended that the author(s) of the measure be consulted to avoid a breach of copyright, and also for quality assurance.

Selectively combining scales or items from multiple empirical measures negates established psychometric properties. Validity and reliability are specific to the number and order of items as presented in the complete measure (e.g., priming effect of earlier items). Furthermore, attempting to evaluate a scale with a single 'top level' item such as 'fatigue' introduces ambiguity. For instance, to rate their fatigue, an athlete may reflect upon their physical and/or psychological fatigue, their fatigue before, during, or after training, or in general. It is recommended that athletes be involved in the development and refinement of a measure to ensure items are clearly understood and perceived as relevant[23]. Careful consideration should also be given to the elements of a measure (e.g., type of scale, response options) and other factors related to implementation (e.g., access to technology, data security) specific to the sport context[21,26].

Develop a supportive culture

The environment surrounding ASRM use is a critical factor in encouraging athlete compliance and honesty[21]. Efforts should be directed at creating a culture whereby there is a mutual understanding among all stakeholders, and transparency in regards to all aspects of intended and potential data uses. To achieve this, the following steps are recommended:

1. define the purpose of implementing an ASRM;
2. establish guidelines regarding data access and use;
3. define individual roles and responsibilities;
4. provide education, transparency, and feedback;
5. build confidence and integrate into the normal routine.

In sports programs where several practitioners are involved in athlete monitoring, steps 2 and 3 are critical to protect athlete confidentiality and ensure a consistent message is being fed back to the coaching team and athlete. It is recommended that there is one key practitioner who coordinates inputs and acts as an intermediary between the support staff and coaching team[27].

Formal or informal education for athletes and practitioners should include why an ASRM is to be used, the purpose of items, who looks at the data, when the athlete will get feedback, and how data is to be used to the athlete's benefit, and not used to their detriment (including selection)[21]. Athletes and practitioners should also be afforded the opportunity to ask questions to fill any gaps in understanding and quell any apprehensions.

Analysis and interpretation

The interpretation of data requires the assessment of whether or not a change is meaningful. Interpretation of a meaningful change in ASRM scores needs to take into account the individual's reporting habits. Some athletes habitually report within a very narrow range of values whilst others fluctuate considerably. Furthermore, the value an athlete considers as their normal may be the mid-point on the scale, or at the lower-end or upper-end of the scale. To account for this, a common method used to analyse data is to calculate the deviation in an athlete's score from their mean. Practitioners may then set a threshold for what deviation from the mean reflects a meaningful change, and therefore is 'red-flagged'. For example, a threshold of a 50% increase in mood disturbance (and less than a 10% increase when an increase was expected) from an off-season baseline has been used to successfully modulate training in elite canoeists[28]. In practice, thresholds of 5% or one standard deviation from the mean have been reported[3]. Practical experience has also led to the recommendation of 1.5 standard deviations from a baseline mean[2]. However, in each instance the stated thresholds and baselines (e.g., off-season, pre-season, previous similar training phase, rolling mean) remain arbitrary and non-specific to particular phases of training and competition. Furthermore, it should be noted that mean and standard deviation calculations are inappropriate for categorical variables. A preferable approach is to determine the typical error for each individual from repeated measures between which no change in state is anticipated to have occurred. A threshold of 1.5-2.0 times the typical error has been recommended[29], however again this threshold is arbitrary.

Day-to-day practice

Day-to-day practice has adapted to the use of potentially inaccurate arbitrary 'red-flag' thresholds by incorporating a step of contextualisation[30]. Flagged responses are considered in light of other available information (e.g., other measures, knowledge of the athlete and their current situation), in order to determine whether or not action is justified. This approach may be practical and even favoured in contexts where practitioners have a close working relationship with a small number of athletes. In fact, it may be preferable to increase sensitivity for 'red-flags' by lowering arbitrary thresholds, and accepting a higher rate of false positives. However, this additional burden on practitioners to follow-up each flagged response may not be sustainable, and is impractical in less intimate or well-resourced contexts. Therefore, it is necessary to direct efforts towards refining thresholds for red-flags.

Whilst attention is typically focussed on detecting a meaningful change in an individual's data, it may equally be telling to detect a *lack* of change when change is expected. Therefore, consideration should also be given to the acute training load and, if applicable, the response of other athletes completing similar training. For instance, an athlete who's subjective wellbeing remains stable rather than improving with a decrease in training load[4], should be flagged, rather than their peers all being flagged for improving as expected.

The complex interplay of factors influencing an athlete's training response and wellbeing suggest more sophisticated analytical techniques may be necessary. For instance, including training load and individual characteristics as cofactors in an analytical model will increase the sensitivity to changes which are not explained by these cofactors. This approach is becoming increasingly accessible with advances in analytical techniques and software, along with software which collates various athlete monitoring data sources into a single database.

Research

A limitation of using ASRM data which has been used to inform day-to-day practice for research is the inability to control for intervening actions which may have taken place. Consider, for example, a research question investigating the efficacy of a certain training protocol within an applied context. Over the course of the study, it may have been necessary to modify the training protocol on an individual basis in the interest of athlete wellbeing and performance (e.g., injury, illness, unfavourable training response). Three suggestions to account for this research-intervention dilemma are:

i. If an intervention is made which results in deviation from the research protocol, only include data up to that intervention.
ii. Follow a standard, pre-determined intervention procedure (e.g., Berglund and Safstrom[28]).
iii. Follow the procedures of Action Research[31], whereby the decision making process, intervention, and outcomes are well-documented for subsequent reflection.

A further consideration is that intervening actions may also be taken by the athlete. Reported anecdotes suggest that completing an ASRM may have an educational effect upon athletes, leading them to recognise behaviours which may not align with their goals and to adjust their behaviours accordingly[30,32-34]. This is desirable from an athlete development perspective, however lacks the necessary control required for a research protocol. The occurrence of athlete-initiated interventions may be discerned qualitatively from athlete interviews or analysis of trends in reported behaviours.

Publishing

Minimum reporting guidelines

Moving forward, to improve the strength and transparency of ASRM data in future research, we recommend the following minimum reporting guidelines. Provide detail on:

1. dimension selection and theoretical basis;
2. rationale for instrument selection or development;
3. dimensions, scales, and items;
4. time frame;
5. answer mode and how items/scales/dimensions scored;
6. validity and reliability;
7. reference values;
8. what instruction and/or education was provided to athletes, and whether there was a period of familiarisation;
9. when, where, and how the measure was completed;
10. what methods of reinforcement were used, if any, such as rewards for completion, or punishment for non-completion;
11. compliance rate, and how determined;
12. assessment of meaningful change, and;
13. if and how the data was used to influence practice during the data collection period, and whether this was accounted for in analysis.

All points should be addressed for custom or modified empirical measures. Empirical measures used in a population or context not previously reported in the literature may omit points 3-5 and cite the original measure. Empirical measures used as previously reported in the literature may omit points 3-7 and cite the original measure and previous literature.

Conclusion

This commentary details considerations to evaluate and improve the quality of ASRM data for both research and practice. Greater scrutiny of ASRM in applied research should encourage researchers to acknowledge the limitations of ASRM data, and take steps to minimise sources of error in future research and practice. As custom ASRM are already widely used in practice, practitioners must weigh up whether to adopt an empirical measure instead, or to proceed to establish the psychometric properties of the existing measure. Taken as a whole, it is hoped that research and practice will become more closely aligned, allowing the training response to be better understood and managed.

References

1. Halson SL. Monitoring training load to understand fatigue in athletes. *Sports Medicine.* 2014;44(S2):139-147.

2. Coutts A, Cormack SJ. Monitoring the training response. In: Joyce D, Lewindon D, eds. *High-performance training for sports*. Champaign, IL: Human Kinetics Publishers; 2014:71-84.

3. Taylor K, Chapman D, Cronin J, Newton M, Gill N. Fatigue monitoring in high performance sport: A survey of current trends. *Journal of Australian Strength and Conditioning.* 2012;20(1):12-23.

4. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: a systematic review. *British Journal of Sports Medicine.* 2016;50(5):281-291.

5. Ekkekakis P. *The measurement of affect, mood, and emotion: A guide for health-behavioral research.* New York, NY: Cambridge University Press; 2013.

6. Kellmann M, Kallus K. *Recovery-Stress Questionnaire for Athletes: User manual.* Champaign, IL: Human Kinetics; 2001.

7. Morgan WP, Brown DR, Raglin JS, O'Connor PJ, Ellickson KA. Psychological monitoring of overtraining and staleness. *British Journal of Sports Medicine.* 1987;21(3):107-114.

8. McNair P, Lorr M, Droppleman L. *POMS manual.* 2nd ed. San Diego, CA: Education and Industrial Testing Service; 1981.

9. Raglin JS, Morgan WP. Development of a scale for use in monitoring training-induced distress in athletes. *International Journal of Sports Medicine.* 1994;15(2):84-88.

10. Shacham S. A shortened version of the Profile of Mood States. *Journal of Personality Assessment.* 1983;47(3):305-306.

11. Grove JR, Prapavessis H. Preliminary evidence for the reliability and validity of an abbreviated Profile of Mood States. *International Journal of Sport Psychology.* 1992;23(2):93-109.

12. Terry PC, Lane AM, Lane HJ, Keohane L. Development and validation of a mood measure for adolescents. *Journal of Sports Sciences.* 1999;17(11):861-872.

13. Terry P, Lane A, Fogarty G. Construct validity of the Profile of Mood States-Adolescents for use with adults. *Psychology of Sport and Exercise.* 2003;4:125-139.

14. Dean J, Whelan J, Meyers A. An incredibly quick way to assess mood states: The incredibly short POMS. Paper presented at: Annual Conference of the Association for the Advancement of Applied Sport Psychology1990.

15. Kenttä G, Hassmén P, Raglin JS. Mood state monitoring of training and recovery in elite kayakers. *European Journal of Sport Science.* 2006;6(4):245-253.

16. Kellmann M, Kallus KW. Recovery-Stress Questionnaire for Athletes. In: Kallus KW, Kellmann M, eds. *The Recovery-Stress Questionnaires: User manual*. Frankfurt am Main: Pearson Assessment & Information GmbH; 2016:86-131.

17. Rushall BS. A tool for measuring stress tolerance in elite athletes. *Journal of Applied Sport Psychology.* 1990;2(1):51-66.

18. Meeusen R, Duclos M, Foster C, et al. 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). *European Journal of Sport Science.* 2013;13(1):1-24.

19. Kallus KW, Kellmann M. Burnout in athletes and coaches. In: Hanin YL, ed. *Emotions in Sport*. Champaign, IL: Human Kinetics; 2000:209-230.

20. Vaughn BK, Daniel SR. Conceptualizing validity. In: Tenenbaum G, Eklund RC, Kamata A, eds. *Measurement in sport and exercise psychology*. Champaign, Il: Human Kinetics; 2012:33-39.

21. Saw AE, Main LC, Gastin PB. Monitoring athletes through self-report: Factors influencing implementation. *Journal of Sports Science and Medicine.* 2015;14:137-146.

22. Collins D, Moore P, Mitchell D, Alpress F. Role conflict and confidentiality in multidisciplinary athlete support programmes. *British Journal of Sports Medicine.* 1999;33(3):208-211.

23. Beckmann J, Kellmann M. Procedures and principles of sport psychological assessment. *Sport Psychologist.* 2003;17(3):338-350.

24. Petscher Y, Schatschneider C. Validating scores from new assessments: A comparison of Classical Test Theory and Item Response Theory. In: Tenenbaum G, Eklund RC, Kamata A, eds. *Measurement in sport and exercise psychology*. Champaign, Il: Human Kinetics; 2012:41-52.

25. Hooper SL, Mackinnon LT. Monitoring overtraining in athletes. *Sports Medicine.* 1995;20(5):321-327.

26. Saw AE, Main LC, Gastin PB. Impact of sport context and support on the use of a self-report measure for athlete monitoring. *Journal of Sports Science and Medicine.* 2015;14:732-739.

27. Norris SR, Smith DJ. Planning, periodization, and sequencing of training and competition: The rationale for a competently planned, optimally executed training and competition program, supported by a multidisciplinary team. In: Kellmann M, ed. *Enhancing recovery: Preventing underperformance in athletes*. Champaign, IL: Human Kinetics; 2002:121-141.

28. Berglund B, Safstrom H. Psychological monitoring and modulation of training load of world-class canoeists. *Medicine and Science in Sports and Exercise.* 1994;26(8):1036-1040.

29. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Medicine.* 2000;30(1):1-15.

30. Saw AE, Main LC, Gastin PB. Role of a self-report measure in athlete preparation. *Journal of Strength and Conditioning Research.* 2015;29(3):685-691.

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

32. Botterill C, Wilson C. Overtraining: Emotional and interdisciplinary dimensions. In: Kellmann M, ed. *Enhancing recovery: Preventing underperformance in athletes*. Champaign, IL: Human Kinetics; 2002:143-159.

33. Kellmann M, Patrick T, Botterill C, Wilson C. The Recovery-Cue and its use in applied settings: Practical suggestions regarding assessment and monitoring of recovery. In: Kellmann M, ed. *Enhancing recovery: Preventing underperformance in athletes*. Champaign, IL: Human Kinetics; 2002:219-229.

34. Kenttä G, Hassmén P. Underrecovery and overtraining: A conceptual model. In: Kellmann M, ed. *Enhancing recovery: Preventing underperformance in athletes*. Champaign, IL: Human Kinetics; 2002:57-77.

35. Main L, Grove JR. A multi-component assessment model for monitoring training distress among athletes. *European Journal of Sport Science.* 2009;9(4):195-202.

36. Grove JR, Main LC, Partridge K, et al. Training distress and performance readiness: Laboratory and field validation of a brief self-report measure. *Scandinavian Journal of Medicine and Science in Sports.* 2014;24(6):e483-490.

37. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior.* 1983;24(4):385-396.

38. Hitzschke B, Holst T, Ferrauti A, Meyer T, Pfeiffer M, Kellmann M. Entwicklung des Akutmaßes zur Erfassung von Erholung und Beanspruchung im Sport [Development of the Acute Recovery and Stress Scale]. *Diagnostica.* 2016.

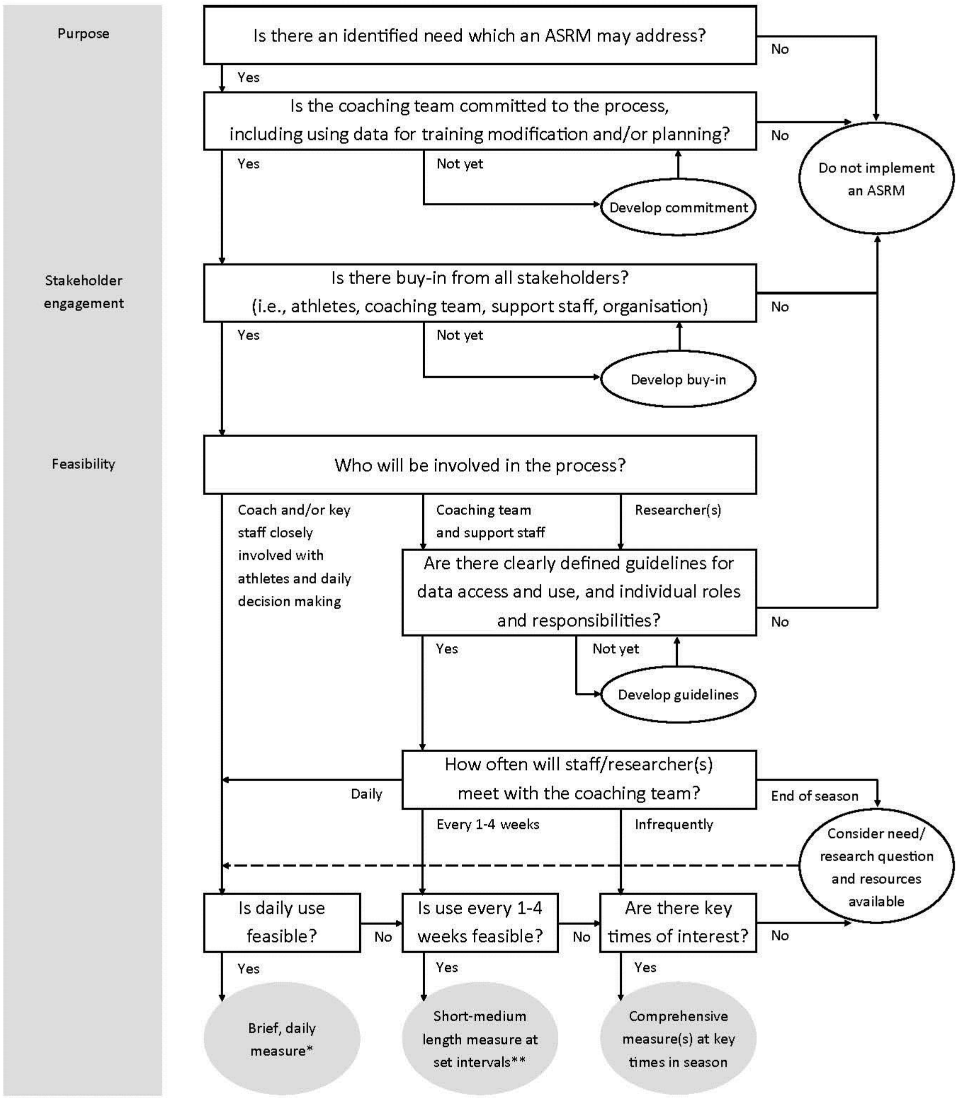
39. Kellmann M, Kölling S, Hitzschke B. *Das Akutmaß und die Kurzskala zur Erfassung von Erholung und Beanspruchung im Sport: Manual [The Acute Measure and the Short Scale of Recovery and Stress for Sports: Manual]*. Köln: Sportverlag Strauß; 2016.

40. Hitzschke B, Kölling S, Ferrauti A, Meyer T, Pfeiffer M, Kellmann M. Entwicklung der Kurzskala zur Erfassung von Erholung und Beanspruchung im Sport (KEB) [Development of the Short Recovery and Stress Scale for Sports (SSRS)]. *Zeitschrift für Sportpsychologie.* 2015;22:146-161.

41. Laurent CM, Green JM, Bishop PA, et al. A practical approach to monitoring recovery: development of a perceived recovery status scale. *Journal of Strength and Conditioning Research.* 2011;25(3):620-628.

42. Lundqvist C, Kenttä G. Positive emotions are not simply the absence of the negative ones: Development and validation of the Emotional Recovery Questionnaire (EmRecQ). *Sport Psychologist.* 2010;24:468-488.

"Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner" by Saw AE, Kellmann M, Main LC, Gastin PB]
International Journal of Sports Physiology and Performance



* Consider supplementing with a more comprehensive measure at set intervals or at key times in season.
** Suggested intervals include weekly, fortnightly, or once per microcycle. Consider supplementing with a brief daily measure during intense training phases, or a more comprehensive measure at key times in season.

Figure 1 Steps to establish the purpose, stakeholder engagement, and feasibility of implementing an athlete self-report measure (ASRM) in a sport context.

“Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner” by Saw AE, Kellmann M, Main LC, Gastin PB]
International Journal of Sports Physiology and Performance

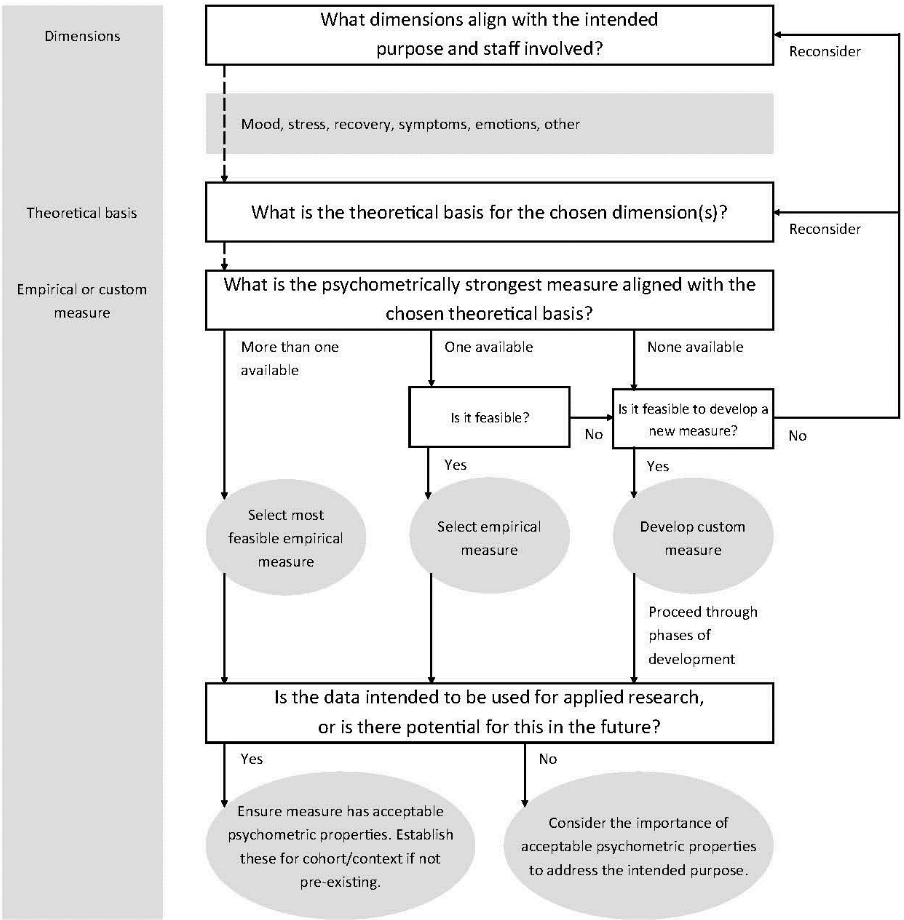


Figure 2 Steps to select an athlete self-report measure (ASRM).

"Athlete Self-Report Measures in Research and Practice: Considerations for the Discerning Reader and Fastidious Practitioner" by Saw AE, Kellmann M, Main LC, Gastin PB]
International Journal of Sports Physiology and Performance

Table 1 Characteristics of empirical athlete self-report measures.

Measure	Primary dimension(s)	General or sport-specific	Scales (n)	Items (n)	Time frame*	Answer mode	Psychometric properties
Profile of Mood States (POMS)[8]	Mood	G	6	65	1 week	5-point	T,D,V,R,N,C
POMS-Short Form (POMS-SF)[10]	Mood	G	6	37	1 week	5-point	T,D,V,R,N,C
POMS-Abbreviated (POMS-A)[11]	Mood	G	7	40	1 week	5-point	T,D,V,R,N,C
POMS-Adolescents/Brunel Mood Scale (BRUMS)[12,13]	Mood	G	6	24	1 week	5-point	T,D,V,N,C
Brief Assessment of Mood (BAM)[14]	Mood	G	6	6	1 week	5-point	T,D,V,R,N,C
POMS energy index[15]	Mood	G	2	15	Now	5-point	T,D,N,C
POMS training distress scale[9]	Mood	G	2	7	1 week	5-point	T,D,V,N,C
Multi-component Training Distress Scale (MTDS)[35]	Mood, stress, symptoms	G,S	6	22	1 day	5-point	T,D,V,N,C
Daily Analyses of Life Demands for Athletes (DALDA)[17]	Stress, symptoms	G,S	2	34	Now	a,b,c	T,D,V,R,N,C
Training Distress Scale (TDS)[36]	Symptoms	G,S	1	19	2 days	5-point	T,D,V,N,C
Perceived Stress Scale (PSS)[37]	Stress	G	1	14/10	1 month	5-point	T,D,V,R,N,C
Recovery-Stress Questionnaire for Athletes (RESTQ-Sport)[6,16]	Stress, recovery	G,S	19/19/12	76/52/36	3 days/nights	7-point	T,D,V,R,N,C
Acute Recovery and Stress Scale (ARSS)[38,39]	Stress, recovery	S	8	32	Now	7-point	T,D,V,R,N,C
Short Recovery and Stress Scale (SRSS)[39,40]	Stress, recovery	S	8	8	Now	7-point	T,D,V,R,N,C
Recovery-Cue[33]	Stress, recovery	G,S	7	7	1 week	7-point	T
Perceived Recovery Status Scale (PRS)[41]	Recovery	S	1	1	Now	11-point	T,D,V
Emotional Recovery Questionnaire (EmRecQ)[42]	Emotions	G	5	22	Now	5-point	T,D,V,R,N

*Time frame reflects response set of originally developed measure, however this may be modified to meet requirements of context if permitted by the manual. G = general wellbeing, S = sport-specific wellbeing, T = theoretically derived, D = documented instrument development, V = validity, R = reliability, N = norm-referenced values in athlete cohorts, C = criterion-referenced values.