Injury, illness and physical demands of professional contemporary dancers: Health outcomes and methodological issues

by Annie Caroline Jeffries

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

Under the supervision of
Professor Franco Impellizzeri
Distinguished Professor Aaron Coutts
Dr Alan McCall
Dr Lee Wallace

University of Technology Sydney
Faculty of Health

February 2021

Certificate of Authorship and Originality of Thesis

I, Annie Jeffries declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the school of Sport, Exercise & Rehabilitation, Faculty of Health at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Annie Jeffries

23/02/2021

Date submitted

Acknowledgements

I would like to thank my principal supervisor Professor Franco Impellizzeri. Your scholarship and knowledge steered me in the right direction. Thank you for the superb guidance and advice you have provided throughout my time as a student. I have been extremely lucky to have a supervisor who is so enthusiastic and passionate about science and research, and who always responded to my questions so patiently and promptly (even at 4am). Thank you for your excellent mentorship and support.

To my co-supervisors, I wish to firstly thank Distinguished Professor Aaron Coutts for your optimism, support and encouragement throughout my studies. To Dr Alan McCall (the great LoTD), and Dr Lee Wallace thank you both for your guidance and advice.

Thank you to all my co-authors, particularly Professor Samuele Marcora. Our collaboration on the framework paper was an absolute highlight for me. I regret I couldn't speak Italian in our meetings, but I am working on it for the next time.

I wish to thank the Sydney Dance Company, the dancers and principle physiotherapist Ashlea Mary Cohen for their support and participation in my research.

To my late mentor Dr Ken Crichton, it was your passion for dance medicine that inspired me on the path of research and for that I thank you. Also thank you to Professor Elizabeth Sullivan, for your guidance, encouragement and belief in me.

Finally, to my long-suffering family and friends. To Adam, thank you for your never-ending support and patience without which I could not have completed my studies. To my dad, thank you for your support. I know it is not a Nobel Prize, but I will keep working on it. To my best friends Jason and Hels, thank you for your loyal friendship and steadfast support. Above all my eternal gratitude to my Mum. Your unwavering belief in me has always carried me through. This thesis is dedicated to your memory.

Preface

This thesis for the degree of Doctor of Philosophy is in the format of published or submitted manuscripts and abides by the 'Procedures for Presentation and Submission of Theses for Higher Degrees – University of Technology Sydney; Policies and Directions of the University'.

Based on the research design and data collected by the candidate two manuscripts have been published and two manuscripts are currently under review in peer reviewed journals. These papers are firstly brought together by an *Introduction*, which provides background information, an explanation of the research problem and the aims of the series of studies. A *Literature Review* then follows with an overview of health-related issues and measures in professional dance including injury, illness and training load, highlighting the gaps within the literature. The body of the research in this thesis is then presented in manuscript form, in a sequence following the development of research ideas in this investigation. As such, each manuscript outlines and discusses the individual methodology and the findings of each study separately. These chapters are formatted according to the specific journal requirements and therefore may slightly vary from each other. The *General Discussion* chapter provides an interpretation of the collective findings and practical applications from the series of investigations conducted. The *Summary* chapter incorporates the flow of research ideas and conclusions from each project and outlines directions for future research.

List of manuscripts submitted for publication

Jeffries, A.C., Wallace, L., Coutts, A.J., Cohen, A.M., McCall, A., Impellizzeri, F.M. (2020). Injury, Illness, and Training Load in a Professional Contemporary Dance Company: A Prospective Study. *Journal of Athletic Training*. 55(9), 967–976.

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*. 15 (9), 1203-1215.

Jeffries, A.C., Marcora, S., Coutts, A. J., Wallace, L., McCall, A., & Impellizzeri, F. M. Development of a revised conceptual framework of physical training for measurement validation and other applications. *Sports Medicine* [under revision, minor modifications requested].

Jeffries, A.C., Novak, A.R., Coutts, A.J., McCall, A., McLaren, S.J, Impellizzeri, F.M. Development and validation of single items for fatigue and recovery. *Journal of Sports Sciences.*[under review]

Conference proceedings

Jeffries, A.C., Wallace, L., Coutts, A.J., Cohen, A.M., McCall, A., Impellizzeri, F.M. Injury, Illness, and Training Load in a Professional Contemporary Dance Company: A Prospective Study. Paper presented at the International Association of Dance Medicine and Science conference Montreal, Canada, 2019.

Statement of candidate contribution

Table 1. 1: Percentage contribution of authors to peer-reviewed manuscripts of thesis

Author	Study One (Chapter Three)					
	Annie Jeffries	**Lee Wallace**	**Aaron Coutts**	**Ashlea Cohen**	**Alan McCall**	**Franco Impellizzeri**
Research design	60%					40%
Ethics application	90%		10%			
Subject recruitment	80%			20%		
Data collection	70%			30%		
Data cleaning	100%					
Statistical analysis	70%					30%
Manuscript preparation	100%					
Manuscript revision	40%	5%	10%		10%	35%

Author	Study Two (Chapter Four)					
	Annie Jeffries	**Lee Wallace**	**Aaron Coutts**	**Shaun McLaren**	**Alan McCall**	**Franco Impellizzeri**
Research design	60%					40%
Ethics application						
Subject recruitment						
Data collection	70%	20%				10%
Data cleaning	80%					20%
Statistical analysis	90%					10%
Manuscript preparation	100%					
Manuscript revision	40%	5%	10%	15%	10%	20%

Author	Study Three (Chapter Five)					
	Annie Jeffries	**Samuele Marcora**	**Aaron Coutts**	**Lee Wallace**	**Alan McCall**	**Franco Impellizzeri**
Research design	40%	20%				40%
Ethics application						
Subject recruitment						
Data collection	60%					40%
Data cleaning						
Statistical analysis						
Manuscript preparation	100%					
Manuscript revision	20%	25%	10%	5%	10%	30%

Author	Study Four (Chapter Six)					
	Annie Jeffries	**Andrew Novak**	**Aaron Coutts**	**Alan McCall**	**Shaun McLaren**	**Franco Impellizzeri**
Research design	60%					40%
Ethics application	90%					10%
Subject recruitment	90%					10%
Data collection	100%					
Data cleaning	100%					
Statistical analysis	50%	20%				30%
Manuscript preparation	100%					
Manuscript revision	30%	10%	10%	10%	10%	30%

Abstract

Introduction: Professional dance is a highly demanding physical activity, with both high injury rates and training load and no research examining illness occurrence. Additionally, few studies have investigated these relationships concurrently. In order to investigate relationships conceptual models and valid measurement tools are required. Despite the widespread use of these measurement tools few have been validated. Therefore, this thesis sought to investigate training load, injury and illness in dance and also develop an appropriate conceptual framework for monitoring and also validating measurement tools in dance. In *Study One*, medical attention and time loss injury, illness and training load data were recorded across one year in a cohort of professional contemporary dancers (n=16). The results show that professional dancers experienced high training loads relative to other athletes and concomitantly high injury and illness incidences and risks. In addition, dancers continued training and performing, even when affected by medical attention injury or illness. *Study Two* was a two-part systematic review. Part one identified the most commonly used athlete reported outcome measures (AROMs) in sport for monitoring training responses; part two assessed the risk of bias, measurement properties, and level of evidence, based on international clinimetric guidelines. Results demonstrated that whilst the measurement properties of multiple-item AROMs derived from psychometrics were acceptable (with the exclusion of content validity and measurement error) the single-item AROMs most frequently used in sport science have not been validated. Until proper validation studies are completed, all conclusions based on these AROMs are questionable. *Study Three* was the development of an updated conceptual framework providing an overarching model that may help understand and guide the development, validation, implementation, and interpretation of measures used for athlete monitoring. Using the conceptual framework (*Study Three*) as a foundation, *Study Four* examined the construct validity and reliability of the single items fatigue and recovery for measuring the training effects in dancers. Results provide preliminary evidence confirming the construct validity, reliability and agreement for the single item of fatigue. The recovery item, despite acceptable reliability and agreement, was only partially confirmed in terms of construct validity, when using the SRSS recovery items as reference. *Conclusion:* Collectively, this thesis provides novel information regarding factors affecting dancer's health and the development of a conceptual framework for monitoring and also validating measurement tools in dance.

Keywords

Rating of perceived exertion

Upper respiratory tract infection

Injury surveillance

Measurement properties

Training effects

Performance

Subjective measures

Athlete monitoring

Fatigue

Recovery

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Absolute	unipolar scale
ABQ	athlete burnout questionnaire
ACSM	American college of sports medicine
ARSS	acute recovery stress scale
AROM	athlete reported outcome measure
AU	arbitrary units
BRUMS	Brunel Mood Scale
CK	creatine kinase
CI	confidence interval
CMJ	countermovement jump
COSMIN	COnsensus-based Standards for the selection of health Measurement Instruments
COVID-19	corona virus disease of 2019
CR-10	category ratio
DALDA	daily analysis of life demands for athletes
DOMS	delayed-onset muscle soreness
ECSS	European college of sport science
Fatigue ABS	fatigue absolute
Fatigue REL	fatigue relative
GPS	global positioning system
HREC	human research ethics committee
H	hours
HRV	heart rate variability
IQR	inter quartile range
IR	impulse-response
LW	linearly weighted
Med-Inj	medical attention injury
Neg.emo.state	negative emotional state
Mental perf	mental performance
MTDS	multi-component training distress scale
NIH	national institute of health

No	number
OFSM	overtraining questionnaire of the societe francaise de medecine du sport
Overall recov	overall recovery
PerPot	Performance potential metamodel
Physical perf	physical performance
POMS	Profile of mood state
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PROMIS	Patient-Reported Outcomes Measurement Information System
PRSS	Perceived recovery stress scale
PSS	perceived stress scale
REDCap	research electronic data capture
Recov ABS	recovery absolute
Recov REL	recovery relative
Relative	bipolar scale
RESTQ-Sport	Recovery-stress questionnaire athletes
SI	single item
SRSS	Short recovery and stress scale
SRSS Recov	Short recovery and stress scale recovery
SD	standard deviation
sRPE	session rating of perceived exertion
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
Tim-Inj	time loss injury
TQR	Total quality recovery scale
TL	training load
URTI	upper respiratory tract infection
y	years

Chapter One

General Introduction

Background

Dance overview

There are many forms of dance including contemporary, classical ballet, tap, ballroom and jazz. Despite having different skills, music and style, each form can be categorized as a predominantly high intensity, intermittent exercise involving aerobic and anaerobic energetic demands (Guidetti, Gallotta, Emerenziani, & Baldari, 2007; Rafferty, 2010; Schantz & Astrand, 1984). As such, dance has some similarities to sports such as football and tennis, in which explosive bursts of exercise are followed by movements requiring precision and skill (Twitchett, Koutedakis, & Wyon, 2009). Attainment of optimal performance in dance requires a complex interaction of aesthetic, artistic, technical, and psychological components (Kuno, Fukunaga, Hirano, & Miyashita, 1996), often involving extreme physical demands and a high level of athletic ability (Gamboa, Roberts, Maring, & Fergus, 2008; Khan et al., 1995).

Dance has been a part of human culture since the very earliest civilisations, with recorded evidence of dancing being found dating back to 10,000 years ago (Dubey-Pathak, 2014). Over time different dance styles have merged and evolved into what we know today as the most well-known dance genres: contemporary, classical ballet, ballroom and jazz. One of the most popular and technical forms of dance is contemporary dance (Figure 1a). First developed during the mid-twentieth century contemporary dance draws on classical ballet, modern and jazz dance styles. Contemporary involves extreme range of motion, strong torso and legwork, with varying changes in speed and rhythm. Classical ballet was first developed during the Italian renaissance and is traditionally performed to classical music. It is highly technical, known for its aesthetics and rigorous technique such as pointe work (e.g., standing on toes), turn out (e.g., external rotation of the legs) and high leg extensions (Figure 1b). Additionally, classical ballet requires precise movement together with ethereal like qualities. Other forms of dance that are common are jazz and ballroom; ballroom originated at the end of the sixteenth century in France and is a type of partnered dance and also comprises several different sub styles (e.g., tango, waltz) (Figure 1c). Jazz originates from seventeenth century African traditions and is an umbrella term that encompasses several different sub styles (e.g., hip hop, Broadway). It consists dramatic body movements, improvisation and isolations of specific body parts (Figure 1d).

Figure 1. 1: Contemporary dance

Figure 1. 2: Classical ballet

Figure 1. 4: Ballroom dance

Figure 1. 3: Jazz dance

Dancers regularly undertake high training loads often with little recovery time (Jeffries, Wallace, & Coutts, 2017). However, the concept of monitoring training load in dance is still in its infancy with only a few studies validating internal training load in contemporary and classical ballet. Additionally, high injury rates are prevalent across all levels and genres of dance (Allen, Nevill, Brooks, Koutedakis, & Wyon, 2012; Bronner, Ojofeitimi, & Rose, 2003; Bronner & Wood, 2017). Whilst there has been some health research regarding eating disorders (Liu, Tseng, Chang, Fang, & Lee, 2016; Nascimento, Luna, & Fontenelle, 2012) and vitamin D deficiency in dancers (Beck, Mitchell, Foskett, Conlon, & von Hurst, 2015; Rowan et al., 2019), there is no research examining illness, specifically upper respiratory tract infections. Within sport, variation in training load, particularly high training loads, have been associated with the occurrence of upper respiratory tract infections (Gleeson et al., 2012). However, this relationship has not yet been investigated in dance. Additionally, no research exists examining training load, injury and illness concurrently.

Injury in dance

Traditionally, large training volumes form the foundation for the refinement and development of a dancer's technique and performance, and typically involve up to 10 hours (hrs) per day for professional dancers (Wyon, 2010). However, a consequence of large training volumes is increased susceptibility to musculoskeletal injury (Motta-Valencia, 2006), leading to high injury rates in both professional ballet and contemporary dancers (Jacobs et al., 2017). The scope of epidemiological literature suggests that the most predominant injuries in dance relate to overuse, particularly of the lower extremities (Allen et al., 2012; Bronner et al., 2003; Gamboa et al., 2008). Moreover, the prevalence of musculoskeletal injuries in dancers, regardless of the style and level of participation, indicates dance is a high-risk activity. For example, higher levels of injury rates (4.7/1000 dance hours) have been reported for professional ballet dancers (Luke, Kinney, & d'Hemecourt, 2002), compared to elite gymnasts (2.6/1000 h) (Kolt & Kirkby, 1999) and ice skaters (1.4/1000 h) (Kjaer & Larsson, 1992). Furthermore, based on the Safe dance report II, 89% of all professional dancers within Australia experience one or more injuries sufficient to adversely impact dance performance over the course of their career and 50% have persistent or recurrent injuries (Crookshanks, 1999).

Despite the high injury incidence rates reported within professional dance, there is a wide range from 0.16 to 4.7 per 1000 h of exposure (Allen et al., 2012; Bronner et al., 2003; Luke et al., 2002). This large variation may be due to a number of factors, including injury definition, style of dance (i.e., contemporary or ballet), repertoire, and tour schedule. For example, many injury classification systems are available, however researchers may select an injury definition without justification, which is a source of degrees of freedom potentially changing the results and interpretation (Hamilton, Meeuwisse, Emery, & Shrier, 2012). Furthermore, some studies use self-reported injury data as opposed to an injury diagnosis made by a health care practitioner (e.g., physician or physiotherapist). Again, this may lead to different conclusions based on the same population data.

Further issues with epidemiology studies within dance is that they do not adequately detail the typology of injuries. For example, whether an injury is the direct result of training or performance (e.g., primary) or occurred outside of training (e.g., secondary). Additionally, an accurate description of the nature of the injury (e.g., acute, chronic, traumatic, non-traumatic), together with a detailed description of the tissue level and body region. Finally, all of this information should then be provided with time to event modelling in order to investigate the injury aetiology in a dance context. Currently, there is paucity of research in which injury typology is accurately described and no research exists examining survival (time to event) analysis in dance. As such, research on injury incidence and risk factors in dance is both limited and methodologically inconsistent compared with investigations of many other athletic activities.

Injury risk factors

Injury aetiology is multifactorial, involving both intrinsic and extrinsic risk factors (Meeuwisse, Tyreman, Hagel, & Emery, 2007), which can be further identified as modifiable and non-modifiable (Maffey & Emery, 2007). Modifiable risk factors are those that can be varied to reduce injury rates through the implementation of appropriate prevention strategies (Meeuwisse, 1991). For example, decreased levels of sport-specific training, endurance, strength and balance have been associated with increased risk of injury in football, ice hockey and the military (Cahill & Griffith, 1978; Dryden, Francescutti, Rowe, Spence, & Voaklander, 2000; Emery, 1999, 2003; Jones, Bovee, Harris, & Cowan, 1993; Pinto, Kuhn, Greenfield, & Hawkins, 1999; Tropp, Ekstrand, & Gillquist, 1984). Additionally, research suggests that training load is a major risk factor for injury (Drew & Finch, 2016). Moreover, inappropriate balance between loading and recovery may induce fatigue, maladaptations (Fry, Morton, & Keast, 1991; Kenttä & Hassmén, 1998; Meeusen et al., 2013) and increased risk of injury and illness (Soligard et al., 2016). However, currently there are no validated instruments for the measurement of fatigue or recovery in dance. Future research in dance should examine the relationship between modifiable risk factors such as training load and injury and illness.

Numerous factors have been as been proposed as injury mechanisms in dance including previous injury, psychological (Noh, Morris, & Andersen, 2011), anthropometric (Benson, Geiger, Eiserman, & Wardlaw, 1989; Twitchett, Angioi, Koutedakis, & Wyon, 2010); poor aerobic capacity (Twitchett et al., 2010), dance technique (Fietzer, Chang, & Kulig, 2012) and fatigue (Liederbach, Schanfein, & Kremenic, 2013). It has also been suggested that rehearsal and performance regimes that include high training loads and under estimation of recovery time also contribute to overuse injuries (Ekegren, Quested, & Brodrick, 2014). In a previous study on pre professional dancers, we reported dancers undertake very high training loads especially in comparison with other high-level athletes (Jeffries et al., 2017). However, to date, studies designed to identify risk factors for injury in dancers have several limitations including inconsistent injury definitions, limited sample size and non-optimal study designs (Kenny, Whittaker, & Emery, 2016). Whilst there is strong evidence that lower extremity overuse injuries are the most common form of injury in dancers, there is a lack of evidence identifying key risk factors for these injuries. Therefore, further studies are warranted that provided a detailed description of injury type in order to accurately identify risk factors within dance.

Training load as a risk factor for injury

Over recent years training load has been proposed as a risk factor for injury (Blanch & Gabbett, 2016; Gabbett, 2016; Hulin et al., 2014; Hulin, Gabbett, Lawson, Caputi, & Sampson, 2016). In order to induce a training outcome, training load is a necessary input variable, that together with other contextual and individual factors, may partially contribute to injury occurrence. For example, stress fractures or acute injury of bone are aligned with excessive stress being experienced by the bone (Melvin, 1993) and therefore potentially related to the training load. Additionally, it has been proposed that excessive training in relation to the ability to recover can favour injuries (Kibler & Chandler, 1998). Only one study has examined training load in relation to injury (overuse) in dance

(Boeding, Visser, Meuffels, & de Vos, 2019). The authors reported that the training load of dancers with self-reported symptoms of overuse injury was higher than that of dancers with no symptoms. However, caution is warranted when interpreting these findings due to methodologic concerns, including a short observation period (7 weeks), missed data (time-loss or medical-attention injuries) and a lack of detailed injury reporting or quantification of training load measures. Therefore, longitudinal studies that provide detailed injury description together with training load data are needed in order to hypothesise (and later verify) how training load may influence injuries and, eventually, what different types of injuries.

Illness in dance

Currently no research exists examining the incidence or severity of illness in dance, in particularly upper respiratory tract infections. Upper respiratory tract illness has been reported as the most common reason for non-injury related presentations to sports medicine clinics, accounting for 35% to 65% of illnesses (Gleeson & Pyne, 2016). Furthermore, high training load has been proposed as a risk factor for illness in athletes (Cunniffe et al., 2011; Hellard, Avalos, Guimaraes, Toussaint, & Pyne, 2015; Novas, Rowbottom, & Jenkins, 2003). Additionally, changes in training load have been associated with an increased risk of illness (Schwellnus et al., 2016). For example, increases in training volume have been related to an increased risk of illness in elite swimmers (Hellard, Avalos, Guimaraes, Toussaint, & Pyne, 2015) and elite junior tennis players (Novas et al., 2003). Currently, there is no research concurrently examining the incidence of upper respiratory tract infections together with training load data in dancers.

Dancer and athlete monitoring

Monitoring athletes is essential in understanding the biological adaptations that occur during training and competition loads. As training imposes stress on an athlete, their physical and psychological well-being move along a continuum that progresses from homeostasis through to stages of acute fatigue, overreaching, overtraining syndrome and ultimately time-loss injury (Coutts & Cormack, 2014; Fry et al., 1991; Soligard et al., 2016). Such maladaptations may arise through both poor load management and interaction with psychological non-sport stressors (Fry et al., 1991; Johnson & Ivarsson, 2011; Lehmann, Foster, & Keul, 1993).

Monitoring involves collecting information on both training stimulus, its effect and responses (Impellizzeri, Marcora, & Coutts, 2019). Options for athlete monitoring include objective measures such as performance, biochemical and physiological and subjective measures such as self-reported questionnaires (Coutts, Crowcroft, & Kempton, 2017; Saw, Main, & Gastin, 2016). In a systematic review, subjective measures (i.e., questionnaires) were found to correspond with acute and chronic training related changes in athlete “well-being”, particularly with respect measures of mood disturbance, perceived stress, recovery and symptoms of stress (Saw et al., 2016). However, the psychometric properties and quality of these subjective measures have not been assessed according to current guidelines. Additionally, in the context of athlete-reported outcome measures, a reference framework is required in order to evaluate the validity of an instrument.

Whilst there are reference frameworks and models available for physical training (Calvert, Banister, Savage, & Bach, 1976; Impellizzeri et al., 2019; Impellizzeri, Rampinini, & Marcora, 2005; Morton, Fitz-Clarke, & Banister, 1990; Perl, 2001, 2004), they do not include all components necessary to provide a suitable reference for validation studies particularly involving subjective measures. Within dance, a previous study has examined the validity of the session rating of perceived exertion method for quantifying training load (Jeffries et al, 2017), however, no subjective measures have been specifically validated for measuring training effects in dancers.

Statement of the Problem

Dance is a highly demanding activity, with high injury rates and training load and no research examining illness occurrence. Additionally, few studies have investigated these relationships concurrently. In order to investigate relationships conceptual frameworks and valid measurement tools are required. Despite the widespread use of these measurement tools few have been validated. Therefore, this thesis sought to investigate training load, injury and illness in dance and also develop an appropriate conceptual framework for monitoring and also validating measurement tools in dance.

Research Objectives

In order to cover the lack in the literature of the aforementioned key aspects, the aims of the series of investigations included in this thesis are:

1. Examine injury, illness, and training load in professional contemporary dancers.
2. Identify commonly used Athlete Reported Outcome Measures (AROMs) and assess the risk of bias and measurement property qualities.
3. Develop an updated conceptual framework intended to facilitate the validation and interpretation of physical training measures.
4. Establish construct validity and reliability of two single items for fatigue and recovery in dancers.

Purpose of Studies

Chapter Three- Injury, Illness, and Training Load in a Professional Contemporary Dance Company: A Prospective Study

- The purpose of this study was to provide a detailed description of injury, illness, and training load occurring in professional contemporary dancers.

Chapter Four- Athlete-Reported Outcome Measures for Monitoring Training Responses: A Systematic Review of Risk of Bias and Measurement Property Quality According to the COSMIN Guidelines

- The first purpose of this review was to identify the most commonly used AROMs in sport for monitoring training responses.
- The second purpose was to assess risk of bias, measurement properties and level of evidence, of the identified instruments from part 1, based on the COnsensus based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

Chapter Five- Development of a revised conceptual framework of physical training for measurement validation and other applications

- The primary purpose was to provide an updated conceptual framework intended to facilitate the validation and interpretation of physical training measures.
- The secondary purpose was to provide theoretical and practical applications of the framework such as validation and conceptualization of constructs (e.g., performance readiness), and understanding of higher order constructs, such as training tolerance when monitoring training to adapt it to individual responses and effects.

Chapter Six- Development and validation of single items for fatigue and recovery: preliminary results in dancers

- The primary purpose was to examine construct validity (convergence and divergence using a hypothesis testing approach) and reliability of two single items for fatigue and recovery, one taken from the literature and one purposely created for dancers.
- A secondary purpose was to explore the comprehension and the interpretation of these two items using a concept identification approach.

Research progress linking the manuscripts

The original aim of this thesis was to examine associations of injury and illness with training load and subjective measures (such as wellness) in professional dancers (*Study One*). When the initial project was designed this was a common approach in research and therefore exploration of these associations formed the initial goal of the first study. However, in order to prove associations, potential explanatory or predictive variables need to be valid. As such we began to examine the validity of the wellness items (*Study Two*). We identified commonly used subjective measures such as the 'wellness' items which did not appear to be valid or validated. Based on this finding we excluded the analysis of these variables from *Study One.* Additionally, analysis of training load was also excluded given there was no underlying rationale together with the possibility of increased researcher degrees of freedom leading to inflated rate of false-positive findings and over estimation of effect sizes. Indeed, we demonstrated this scenario in the results of *Study One*. The project then evolved to the validation of subjective measures used for monitoring the training effect, in particularly the wellness items (*Study Two*). Given the lack of studies showing the validity of the single items commonly used in research, *Study Four* was planned to select and examine the validity of items that can be useful to monitor training. However, the validation process necessitates a framework and for this reason *Study Three* was conducted. *Studies Two and Three* were developed encompassing athletes in general, not only dancers, given most of the instruments also used in dance arise from sport science. However, while the conceptual framework may be more general, the validation process is population specific. Therefore, *Study Four*, although based on the new conceptual framework, examined the construct validity for single item measures of fatigue and recovery (*Study Four*) in a relatively large cohort of dancers. Relevance of concepts included in the wellness items was also investigated from final user perspective (*Study 5, ongoing*).

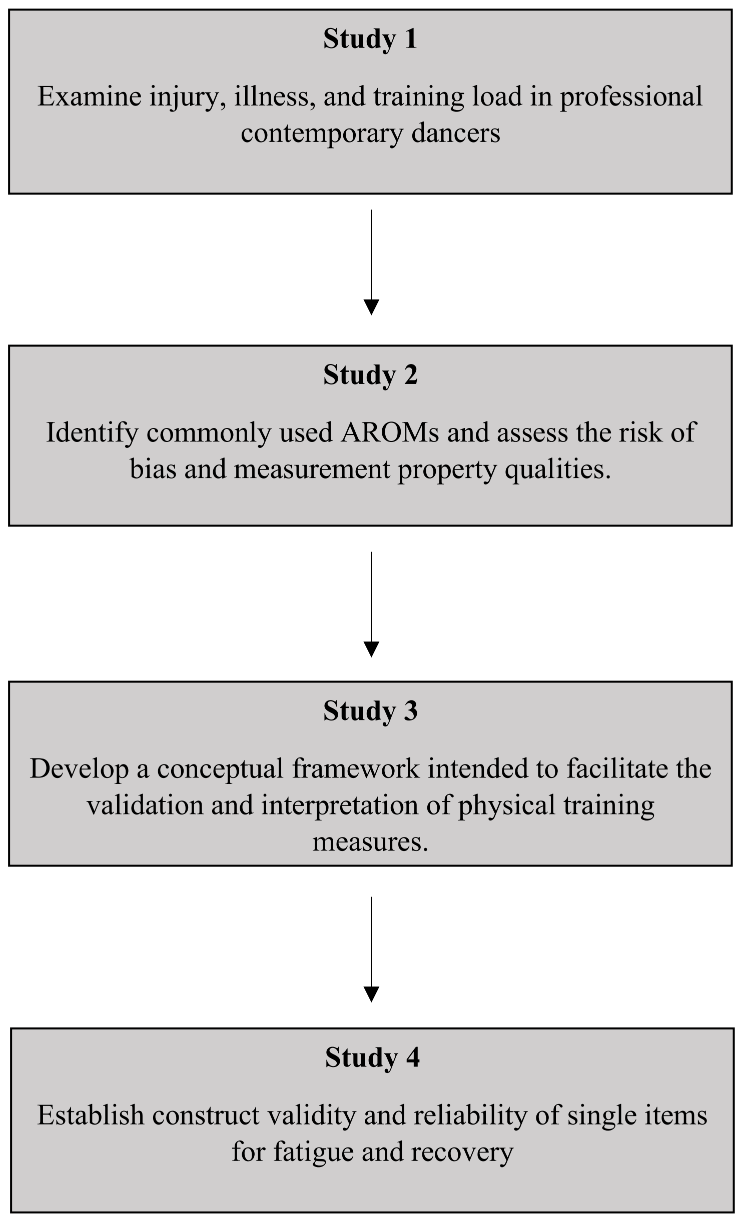


Figure 1. 5: General outline of the research progress linking the four studies undertaken in this thesis.

Chapter Two

Literature Review

This literature review contains five sections. First an overview of dance, the burden of injury and current practices in training load monitoring. Second, an examination of training load and injury research. Third, we review illness in athletes and the current research on relationship between training load and illness. We then examine subjective monitoring tools and the validity of these instruments. Finally, a brief review on available frameworks for the physical training process. As there is very limited research in dance regarding training load, illness and monitoring practices this literature review will combine information from both sport and dance science. Additionally, as there is limited research specifically on professional contemporary dancers, we will explore research regarding other dance styles and skill levels.

Introduction

Dance is a popular activity on a global scale, with a wide-ranging demographic profile (Emery, Meeuwisse, & McAllister, 2006; O'Neill, Pate, & Liese, 2011). Within adolescent sport, dance is ranked in the top three (22%) in terms of participation (Emery et al., 2006). Although dance has many positive health and well-being aspects (Bize, Johnson, & Plotnikoff, 2007), it is a highly demanding activity with an increased risk of injury and thus reduced future participation (Emery, 2010). Whilst much research has investigated injury rates in dance, there is relatively little information regarding illness or the incidence of injury and illness together with training load over a long period of time.

Injury in dance

Dance represents a unique combination of athleticism and artistry, in that speed and power are required to support graceful and intricate movement. Dancers routinely undertake high training volumes, leading to increased levels of fatigue, impaired performance, and high injury rates (Wyon, 2010). Furthermore, the risk of injury is further heightened by the intense physical demands and repetitive nature of dance movement patterns (Bronner, Ojofeitimi, & Spriggs, 2003). Collectively, this environment, may adversely impact career longevity and cause long term disability (Gamboa et al., 2008; Garrick & Lewis, 2001; Jacobs et al., 2017; Ramel & Moritz, 1998).

The lifetime prevalence of musculoskeletal injury in professional dancers varies from 20% to 84%, with 95% of dancers experiencing musculoskeletal pain (Hincapie, Morton, & Cassidy, 2008; Jacobs, Hincapie, & Cassidy, 2012). It has been reported that 89% of all professional dancers within Australia experience one or more injuries sufficient to adversely impact dance performance over the course of their career and 50% have persistent or recurrent injuries. Moreover, 36% of dancers have reported commencement of chronic injuries before 18 years of age during pre-professional training (Crookshanks, 1999). However, a lack of consensus regarding the definition of injury in dance may have led to inconsistencies in the reporting of injury rate statistics (Hincapie et al., 2008; Jacobs et al., 2012). Indeed, until recently there was no universally accepted definition of musculoskeletal injury in dance leading to difficulties in comparison of prevalence estimates, incidence rates and risk factors. In addition, few studies are longitudinal or utilise multivariable analyses to determine independent risk factors with injury (Hincapie et al., 2008). Nonetheless, the consensus of research confirms that musculoskeletal injury is a significant health factor for dancers of all skill levels. Future research would benefit from well-designed, longitudinal and statistically sound studies that examine the multifactorial nature of injury in dance.

Injury definition

The definition of injury is critical to the process of injury surveillance and risk factor identification. The most common injury definitions utilised within dance include medical attention (an injury diagnosed by a medical professional) (Timpka et al, 2014), time loss (unable to train/perform one or more days after the injury event) (Bronner et al, 2017) or a self-reported injury. Variations in the definition of injury have been shown to lead to different estimates of injury (Liederbach, Hagins, Gamboa, & Welsh, 2012; Orchard & Hoskins, 2007). Accordingly, to ensure reliable and comparable results across studies, many consensus statements have been developed (Fuller et al., 2006; Hodgson, Gissane, Gabbett, & King, 2007; Liederbach et al., 2012; Orchard & Hoskins, 2007; Timpka et al., 2014). However, the influence of various injury definitions on both the process of injury surveillance and risk factor identification has not been well documented. Interestingly, despite the high estimates of injury reported in dance, it has been suggested that variations in the types of injury surveillance methods (i.e. third party) and the definition of injury employed may lead to under reporting of musculoskeletal injury (Kenny, Palacios-Derflingher, Whittaker, & Emery, 2018). For example, across 47 studies included in a systematic review on injury in dance, ~40% (n=18) did not include an injury definition, with the rest containing variations of time loss or modified dance participation. Therefore, given that injury definitions are the foundation of epidemiological based injury studies, it is imperative that surveillance systems utilise injury definitions that are sufficiently sensitive to capture all injuries.

Risk factors of injury in dance

Since the late 1980s, many authors have suggested that intrinsic characteristics contribute to dance injuries (Bronner et al., 2003; Campbell, Lehr, Livingston, McCurdy, & Ware, 2019; Garrick & Lewis, 2001; Schafle, 1990; Steinberg et al., 2012). For example, imbalances in strength and flexibility, insufficient or excessive joint range of motion, poor postural alignment and reduced body mass index (BMI) have been identified in the aetiology of injury (Benson et al., 1989; Bronner et al., 2003; Hamilton, Hamilton, Marshall, & Molnar, 1992; Reid, Burnham, Saboe, & Kushner, 1987). For example, dancers with a BMI <19.0 spent more days with a low-grade injury (mean 24 days) than dancers with a BMI >19.0 (mean 11.6 days) (Benson et al., 1989). Furthermore, pre-professional dancers who experienced lateral hip and knee pain had significantly reduced flexibility in hip adduction and internal rotation than the control group (Reid et al., 1987).

In contrast, many studies in both dance and the wider field of sports medicine have reported no relationship between intrinsic variables and the prediction of risk in injury (Bennell et al., 1996; Ostenberg & Roos, 2000; Wiesler, Hunter, Martin, Curl, & Hoen, 1996). For example, (Wiesler et al., 1996) did not find BMI to be a risk factor for lower extremity injuries among ballet and contemporary students across a one year study. Additionally, (Bennell et al., 1996) found no difference in body fat, height, weight and total lean mass among male and female track athletes who experienced stress fractures than those who did not. Moreover, in a cohort of female soccer players BMI was not found to be a risk factor for injury (Ostenberg & Roos, 2000). Collectively, therefore it would seem there is no consensus regarding intrinsic risk factors that contribute to injury in dance or sport.

Despite many studies investigating risk factors for injury, a systematic review examining 47 studies on risk factors for musculoskeletal injury in contemporary and classical ballet dancers highlighted poor quality and low levels of evidence, together with inconsistencies in injury definition (Kenny et al., 2016). For example, in terms of quality, many studies had inadequate description of participant characteristics (i.e., age, sex, dance level or style, censoring or follow up, training volume). Careful description of participant characteristics is critical in having a direct impact on appropriateness of study design, analysis and interpretation of results, together with generalisation of conclusions, comparison across studies and replicability (Lane, Wolery, Reichow, & Rogers, 2007). Another issue identified was the predominant study design utilised was cross sectional, which limits the ability to make causal or predictive inferences. Furthermore, many studies employed invalid or unreliable measurement tools, further contributing to the poor quality across studies. Additionally, many studies (Angioi, Metsios, Koutedakis, Twitchett, & Wyon, 2009; Boeding et al., 2019; Twitchett et al., 2010) lack adequate detail regarding the characteristics of injury which is critical for designing injury-prevention programs or for examining causal associations.

In summary, there is no consensus regarding the relationship between intrinsic factors and injury in dance. To date, studies aimed at identifying risk factors for musculoskeletal injury in dancers of all styles have utilised different injury definitions, inadequate study design (e.g., cross sectional), small sample sizes and methodologically inconsistent. In addition, many studies lack detailed descriptions of injury (e.g., the inciting event, description of the injury mechanism, location, injury typology) which is fundamental in understanding injury causation and for future injury prevention strategies. Therefore, further research is warranted in dance in which rigorous methods and detailed injury reporting are adopted.

Training load in dance

To maximize the adaptive response, coaches and scientists need to control the stress applied to the athlete at the individual level. To achieve this, precise control and manipulation of the training load are required (Impellizzeri et al., 2019). Currently, sports science principles are rarely used in the preparation of dancers, including the control of the stress applied to a dancer or measurement of the adaptive response. Dance teachers tend to use training time as a surrogate measure of training exposure (Liederbach, Gleim, & Nicholas, 1992), and together with their experience, intuition and perception assess the intensity and training load being experienced by the dancer. Indeed, subjective assessments made by dance teachers, together with the use of time as a measure of work volume, is inappropriate as it is not evidence based or controllable. More recently, internal training load in contemporary dance has been validated using the sRPE method (Jeffries et al., 2017) in sixteen pre-professional contemporary dancers during a seven-week period. sRPE was significantly correlated with internal training load measures determined from three different heart rate methods $\%HR_{peak}$ ($r = 0.69$), Edwards TRIMP ($r=0.72$), and Bannisters TRIMP ($r=0.77$). It was determined that sRPE was a valid indicator of internal training load in dancers and may be a practical tool for the development of periodisation strategies. Additionally, this study demonstrated that contemporary dancers undertake very high training loads, especially in comparison to many other high-level athletes (McLean, Coutts, Kelly, McGuigan, & Cormack, 2010; Rogalski, Dawson, Heasman, & Gabbett, 2013). For example, mean weekly training loads (4283 ±2442 AU) were much higher than those reported in professional Australian Football players (ranging from 2027 AU pre-season to 1651 AU in-season) (Rogalski et al., 2013) or elite soccer players (2,493 ± 821 AU) (Brito, Hertzog, & Nassis, 2016). Currently, only two studies have investigated training load in dance, both utilising the sRPE method in pre professional contemporary dancers (Jeffries et al., 2017) and amateur classical ballet dancers (da Silva et al., 2015). In summary, whilst there is limited research in this area, there are validated measures available for monitoring internal training load in contemporary dancers.

Training load and injury

The relationship between injury and training load dates back more than 30 years (Kibler, Chandler, & Stracener, 1992), however recently there has been a substantial increase in training load and injury research (Ahmun, McCaig, Tallent, Williams, & Gabbett, 2019; Blanch & Gabbett, 2016; Bowen, Gross, Gimpel, Bruce-Low, & Li, 2020; Bowen, Gross, Gimpel, & Li, 2017; Carey et al., 2017; Colby et al., 2017; Cummins et al., 2019; Hartwig et al., 2019; Hulin et al., 2014; Hulin et al., 2016; Jaspers et al., 2018; Johnston et al., 2019; Malone et al., 2017; Myers et al., 2020; Soligard et al., 2016; Stares et al., 2018;

Thornton, Delaney, Duthie, & Dascombe, 2017). Despite the complexity of injury aetiology in sport, involving both intrinsic and extrinsic factors (Meeuwisse, 1994; Meeuwisse et al., 2007), it has been proposed that training load management is a major risk factor for injury (Drew & Finch, 2016; Soligard et al., 2016). For example, high training loads have been related with increased risk of injury and illness in professional rugby league players (Rogalski et al., 2013). In contrast, low preseason cumulative workloads have been associated with increased in-season injury risk in Australian football (Colby et al., 2017). Furthermore, inappropriate balance between loading and recovery can lead to prolonged fatigue and abnormal training responses such as maladaptation (Fry et al., 1991; Kenttä & Hassmén, 1998; Kuipers & Keizer, 1988; Meeusen et al., 2013) and an increased risk of injury and illness (Drew & Finch, 2016; Soligard et al., 2016).

Monitoring and management of training loads is not common practice in dance. Indeed, there is a paucity of research examining the relationship between training load and injury and illness in dancers. Only one group (Boeding et al., 2019) has examined injury (overuse) symptoms and training load in dance: reporting dancers without musculoskeletal pain had lower training loads. For example, the training load of dancers with self-reported symptoms of overuse injury was higher than that of dancers with no symptoms. However, caution is warranted when interpreting these findings due to methodologic concerns, including a short observation period (7 weeks) and missed data (no time-loss or medical-attention injuries). Therefore, future investigations are required that are longitudinal in nature, have a longer risk period, and provide comprehensive data on injury incidence, symptoms and training load. These may assist our appraisal of the mechanisms of injury and prompt suggestions for preventive interventions.

Despite the recent rise in research in sport supporting the relationship between injury and training load (Ahmun et al., 2019; Blanch & Gabbett, 2016; Bowen et al., 2020; Gabbett, 2016; Hulin et al., 2014; Hulin et al., 2016; Myers et al., 2020) there is an absence of strong evidence confirming the quantitative use of training load data to manipulate training for the purpose of preventing injury (Impellizzeri, McCall, Ward, Bornn, & Coutts, 2020; Impellizzeri, Menaspà, Coutts, Kalkhoven, & Menaspà, 2020). The main issue is there is no reference model or framework on which to base the interpretation of results. Lack of a framework or model in which results are based may give researchers too many degrees of freedom (Smoliga & Zavorsky, 2017), potentially leading to the risk of false discoveries and confirmation bias (Impellizzeri et al., 2020; Impellizzeri et al., 2020). For example, in the absence of a framework it may be interpreted that the underlying mechanism of injury is indicted by any measure of training load (Impellizzeri et al., 2020). As such, a specific conceptual or theoretical framework is required to identify factors that may influence injury and whether there are validated methods to measure these factors. The role of theoretical and conceptual frameworks is further discussed in the Physical training frameworks or models section.

There are many other methodological issues of concern including inadequate quality and standards of reporting, suboptimal analysis, low sample size and generalizability, injury definitions, missing data, measures of exposure and categorisation method (Impellizzeri et al., 2020). For example, there are many measures of exposure (e.g., acute chronic ratio, ratios between low and high training load or week to week variations) that have been utilised in statistical models (Bowen et al., 2020; Bowen et al., 2017; Hulin et al., 2016). An issue with assessing associations from multiple associations of training load measures is it inflates the risk of a type I error. Other issues include both the lack of reporting and describing how missing data was handled. In epidemiology it is good practice to report the proportion of missing data, assess the randomness of missing values and details of how missing data was handled. Omitting observations with missing variables can severely bias parameter estimates if the values are not missing completely at random (Harel et al., 2018). Another concern regarding training load and injury studies is the categorisation of training load. For example, inclusion of participants into low, medium or high categories based on the training load completed by the whole group (Carey et al., 2017; Hulin, Gabbett, Caputi, Lawson, & Sampson, 2016). Most previous studies do not provide a rationale for using group training load as a reference. Given the individual nature of injuries, it would be reasonable to use individual data for categorization (e.g., placing individuals in the high, medium, or low category based on their training load range). Therefore, although training load and injury research is common, in the absence of strong evidence supporting the use of training load metrics, caution must be taken when interpreting results of previous studies.

In summary, injury prevention is a multifactorial and complex phenomenon. Whilst simple solutions for reduction of injury risk and training load modelling are attractive, careful consideration is required to the numerous factors that might contribute to injury. Consequently, future research is required to investigate essential areas such as injury typology and mechanism, such studies should be methodologically sound, of high quality and based on established epidemiologic methods.

Illness

In athletes, upper respiratory tract infections (URTI) account for 35–65% of all illness presentations to sports medicine physicians and are the most common reason for non-injury related presentations (Alonso et al., 2010; Engebretsen et al., 2013; Engebretsen et al., 2010; Mountjoy et al., 2010; Schwellnus, Derman, Lambert, Redhead, & Page, 2011; Soligard et al., 2015). The cause of upper respiratory tract infections remains uncertain; however, the majority of cases are typically related to allergic response to allergens, common respiratory viruses, virus latency (reactivation), and exercise related trauma to the integrity of the respiratory epithelial membranes (Gleeson & Pyne, 2016). Susceptibility to URTIs is influenced by a number of factors including environment, lifestyle (Hellard et al., 2015; König, Grathwohl, Weinstock, Northoff, & Berg, 2000; Walsh et al., 2011) and transient exercise-induced immune suppression (Walsh et al., 2011), with not all illness episodes having infective aetiology (Cox et al., 2008; Reid, Gleeson, Williams, & Clancy, 2004; Spence et al., 2007). Recurring or chronic respiratory illness can negatively impact not only the health but also the performance of an athlete (Schwellnus et al, 2016).

Illness and training load

Elite athletes often experience demanding training, competition and performance schedules, resulting in increased training and competition loads. In particular, inappropriate training load, combined with an overloaded or congested schedule challenges the ability of athletes to cope with these stressors. It has been suggested this can result in decreased performance, increased risk of injury and illness (Soligard et al., 2016). Indeed, high intensity and prolonged training and competition load is related with a heightened risk of both subclinical immunological changes that may increase the risk of illness and diagnosed illness (Gleeson & Pyne, 2016; Walsh et al., 2011).

High training loads can have either positive or negative influences on the risk of illness in athletes. High absolute training loads have been associated with an increased risk of illness in recreational and national level athletes (Gleeson et al., 2012; Gleeson, Bishop, Oliveira, & Tauler, 2013; Novas et al., 2003). Previous research has investigated the relationship between training loads and illness in a variety of sports including athletics, swimming, basketball, rugby league, and tennis (Foster, 1998; Heath et al., 1991; Hellard et al., 2015; Novas et al., 2003; Spence et al., 2007; Thornton et al., 2016). Collectively, these studies have demonstrated a relationship between the risk of illness and increases in the volume of training (Heath et al., 1991; Hellard et al., 2015; Moreira, Arsati, De Oliveira Lima-Arsati, Simões, & De Araújo, 2011; Novas et al., 2003). Additionally, in a prospective study in elite rugby union players, periods of increased intensity of training preceded the development of URTI symptoms (Cunniffe et al., 2011). Interestingly, it has been reported that the association between absolute training load and illness may not exist in the case of highest-level elite athletes. For example, several studies have shown that high absolute training loads in international-level (Hellard et al., 2015; Veugelers, Young, Fahrner, & Harvey, 2016) and medal winning athletes (Svendsen, Taylor, Tonnessen, Bahr, & Gleeson, 2016) are related to a lower risk of illness compared with sub-elite or national level athletes. Whilst the exact mechanisms for these findings are unclear, it has been suggested that it may relate to exceptional physical characteristics and superior immunological defences (Malm, 2006).

Illness in dance

In dance, there is a lack of research data concerning illness, with most research involving eating disorders (Liu et al., 2016; Nascimento et al., 2012; Ringham et al., 2006; Thomas, Keel, & Heatherton, 2005) and vitamin D deficiency (Beck et al., 2015; Ducher et al., 2011; Rowan et al., 2019; Wolman et al., 2013; Wyon, Koutedakis, Wolman, Nevill, & Allen, 2014). Indeed, no epidemiological studies have examined the incidence and severity of illness in this population. Given URTIs are the most common illness presented to sports medicine clinics and the potential negative impact on health and performance of athletes undertaking high levels of strenuous exercise, they are of particular interest (Gleeson & Pyne, 2016). Furthermore, no studies have examined whether there is an association between training load and risk of illness or explored the dose-response relationship of recovery time with the risk of illness in dancers. This may be relevant as previous research has demonstrated dancers undertake very high training loads (Jeffries et al., 2017), especially in comparison to many other athletes.

Monitoring athletes with subjective measures

Exercise is a stressor that induces various psychophysiological responses. It is this response that forms the basis for functional adaptations underpinning changes in training outcomes such as performance or injury (Impellizzeri et al., 2019). As such athlete monitoring provides coaching and medical staff useful information for the prescription and adjustment of training load in order to facilitate optimal adaptation and performance, together with minimisation of the risk of overtraining, illness and injury (Coutts & Cormack, 2014; Halson, 2014). Typically, implementation of athlete monitoring involves cost, time and equipment. However, self-report measures such as Athlete Reported Outcome Measures (AROMs) (e.g., questionnaires and surveys), are a relatively simple and inexpensive approach to athlete monitoring (Halson, 2014). Indeed, research suggests AROMs may be more sensitive and reliable than traditional physiological, performance and biochemical (Buchheit et al., 2013; Coutts, Slattery, & Wallace, 2007; Meeusen et al., 2013). Self-reported measures include psychological variables (such as mood, stress, recovery) and perceptions of well-being and have been associated with overreaching, overtraining (Hooper, MacKinnon, & Hanrahan, 1997; Meeusen et al., 2013; Urhausen, Gabriel, Weiler, & Kindermann, 1998) and increased risk of injury (Galambos, Terry, Moyle, Locke, & Lane, 2005; Johnson & Ivarsson, 2011) and illness (Anglem, Lucas, Rose, & Cotter, 2008; Zorrilla et al., 2001). A range of AROMs are currently used in athlete monitoring and have previously been cited as superior to objective measures in terms of reflecting acute and chronic training loads with superior sensitivity and consistency (Saw et al., 2016). Although widely used in sport, the psychometric properties and qualities of AROMs are yet to be established. Any measure used in research and practice should be valid in order to provide accurate, meaningful, useful, clear, and specific information. These properties are especially critical when using these measures to develop models and understand phenomena. One of the essential attributes is validity, that is, the extent to which a measure actually assesses what it is intended to measure. Therefore, to ensure AROMs measure the particular domains/constructs of interest, proper development and validation are imperative (Saw, Kellmann, Main, & Gastin, 2017).

Within dance, there are no validated AROMs or subjective measures for measuring training effects. Typically, instruments utilized in dance are sourced from sport science (Harrison et al., 2019). However, it essential that instruments are specifically developed for the target population and context of use (Mokkink et al., 2010). Content validity is considered to be the most important measurement property of an AROM measure because if it is unclear what the AROM is actually measuring, the assessment of the other measurement properties is not valuable. Therefore, aspects to consider before selecting an AROM are the target population (e.g., dancers) and the construct (i.e., outcome or domain) to be measured.

Based on the literature, two key constructs of interest for measuring training effects would be fatigue and recovery. For example, given that dancers undertake extremely high training loads with inadequate recovery periods (Jeffries et al., 2017) they may be at risk of increased fatigue levels. Additionally, 48% of dancers participating in the Safe Dance Report IV survey reported fatigue as a contributing factor to injury (Vassallo et al, 2017). Furthermore, as fundamental sport science principles are not commonly used within dance (e.g., periodisation) easy to use subjective measures may allow for teachers and researchers to monitor training effects more readily in dancers. Therefore, although there are validated tools for measuring internal training load in contemporary dancers, such as the sRPE method, currently no validated measures exist for measuring training effects.

Physical training frameworks or models

In order to investigate relationships conceptual models and valid measurement tools are required. While the use and development of reference theoretical and conceptual frameworks is common in disciplines such as psychometrics and clinimetrics (Rothman, Beltran, Cappelleri, Lipscomb, & Teschendorf, 2007; Wilson & Cleary, 1995), they are not frequently used when validating measurements used in sport science. Consequently, the absence of a reference framework presenting hypothesized relations between constructs (e.g., wellness) and other measures (e.g., training load), makes the examination of the validity and usefulness of these items introduced in the sport science literature potentially confusing and prone to bias. For example, some common subjective measures used in research and practice such as "wellness items" do not appear to have been validated following appropriate and established methods such as those presented by the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN). Potentially this may be due to the fact that available frameworks (Calvert et al., 1976; Impellizzeri et al., 2019; Impellizzeri et al., 2005; Morton et al., 1990; Perl, 2001, 2004) do not include all components necessary to provide a suitable reference for these kinds of validation studies. Therefore, a review of all available frameworks or models was conducted to identify areas that could be expanded and that are important in the conceptualization of the process and outcomes of physical training. From the examination of the literature, five relevant models specifically proposed for the training process were identified: the theoretical frameworks by Impellizzeri et al. (2019); (Impellizzeri et al., 2005), the Banister Impulse-Response (IR) (Calvert et al., 1976; Morton et al., 1990), the Performance Potential Metamodel (PerPot) (Perl, 2001, 2004) and the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM) joint consensus statement on the overtraining continuum (Meeusen et al., 2013).

The theoretical framework by Impellizzeri et al (Impellizzeri et al., 2019; Impellizzeri et al., 2005) was developed to define and conceptualize the important measurable components of the training process and its outcomes (Figure 2.1). This framework was further refined and extended in 2019 together with operational definitions (Figure 2.2).

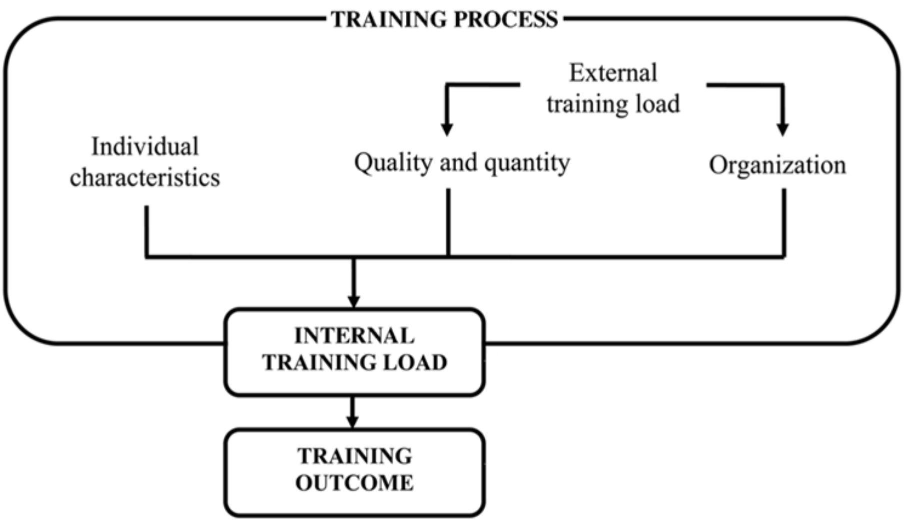


Figure 2. 1: The relationship between internal and external training load on training outcomes (Impellizzeri et al., 2005).

A fundamental concept of this physical training framework was that when delivered appropriately, and according to the training goal and plan, exercise will induce psychophysiological responses leading to adaptations. These responses *during* the exercise are the stimuli initiating the adaptations determining the training outcomes. The training load construct was differentiated into two components, external and internal. In this framework, the external training load, and its interaction with various individual and contextual factors, determines the internal training load, which ultimately will produce changes in the training outcome. In the context of training load, the term external load, refers to the external training load undertaken by an athlete. External load has been defined as the physical work prescribed in the training plan. Accordingly, the term external load accommodates quantification and prescription in a variety of manners, enabling the use of a diverse range of measures and metrics. Some common measures of external load include global positioning systems (GPS) derived units (speed, accelerations, etc.) and level of resistance. Similar to external load, the term internal load implicitly refers to internal training loads, where the term training is the qualifier of load. In the context of training load, internal load typically refers to the psycho-physiological stress experienced by an athlete during the exercise or the training session. Therefore, the concept of internal load incorporates all the psycho-physiological responses initiated to cope with the requirements elicited by the external load, i.e., the execution of the exercise (single or sequence) according to the modalities prescribed by the coach. Despite internal loads typically being psycho-physiological in nature (essentially due to relative ease of application and quantification), the stress and strain experienced by specific tissues in response to an applied force is also internal and should therefore also fall within this category. A possible area to extend this physical training framework is the inclusion of an additional construct 'training effects' with a negative and positive component, as in other models such as Banister Impulse-Response (IR) (Calvert et al., 1976; Morton et al., 1990) and PerPot (Perl, 2001, 2004).

The updated theoretical framework (Impellizzeri et al., 2019) included additional individual and contextual factors such as environment, training status, nutrition, health, psychological status, and genetics. These modifiable and non-modifiable factors can result in a different internal load experienced by an athlete, even when the same external load is applied. An area that may be further extended in this framework is the inclusion of these individual and contextual factors to other areas of the framework. For example, individual factors (e.g., psychological status) can affect the training outcome.

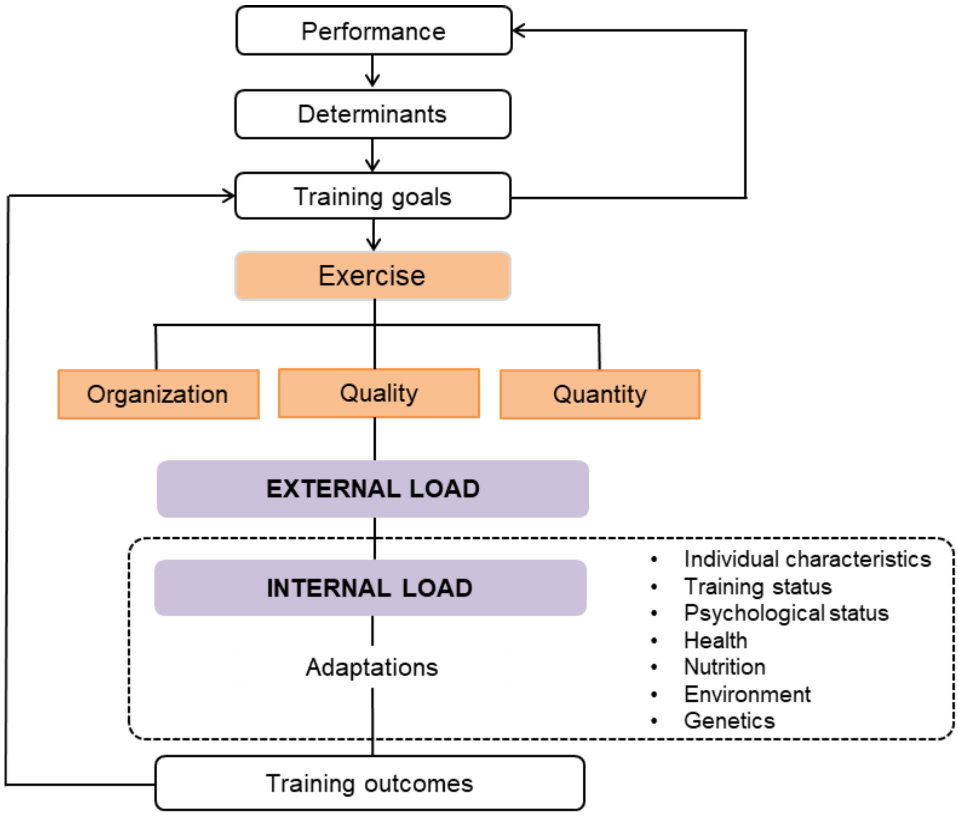


Figure 2. 2: Theoretical framework of the training process (Impellizzeri et al., 2019).

In the seventies Banister and colleagues introduced a system theory to describe and anaylse physical adaptation due to physical training. The model of Banister considered athletic performance to be measurable as the net outcome of two key training responses, called fitness and fatigue, from the application of a training impulse (Figure 2.3). The original works by Banister et al. (Calvert et al., 1976; Morton et al., 1990) applied a systems theory to assess the response to physical training using a mathematical function. In this mathematical model, performance is the output, with the athlete regarded as the system, and the training impulse as the input (Calvert et al., 1976; Morton et al., 1990).

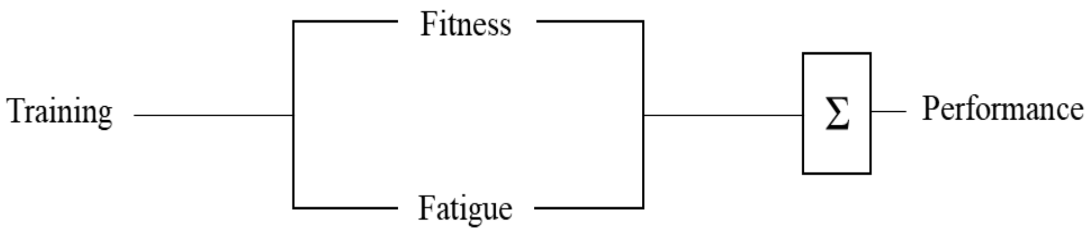


Figure 2. 3: Banister Impulse-Response (IR) model (Calvert et al., 1976; Morton et al., 1990). The training impulse results in two responses or after-effects, which either positively or negatively influence performance with the fitness after-effect having a positive physiological response, whereas the fatigue after-effect resulting in a negative physiological response. The resultant interaction between these two after-effects produces a change in performance with the magnitude and duration of the after-effects dependent on the stimulus. However, there is a conceptual problem related to the definitions, or lack thereof, for the terms 'fitness' and 'fatigue'. Originally, these two terms were arbitrarily used to define two mathematical components reflecting the positive and negative training responses. They were computationally and not conceptually defined since they just represented an assigned name for two "model components of performance ability" and not two theoretical constructs. However, the interpretation of these two terms may result in confusing overlapping with the more generic definitions reported in the literature, making it challenging to achieve a clear understanding of these constructs for conceptualising into a general framework.

Another model has been proposed to investigate the adaptive physiologic processes by means of antagonistic dynamics is the PerPot metamodel (Figure 2.4) (Perl, 2001, 2004). This model simulates the interaction between load and performance in adaptive physiological processes in sport, by means of antagonist dynamics (response flow and strain flow). The basis of the PerPot model is that adaptation to physical training is primarily dominated by individual conditions, is an extremely complex process and characterised by a diversity of parameters and their interrelation (Mester & Perl, 2000). In the metamodel the output potential (performance potential) is influenced by input load (training) and dynamically controlled by two internal potentials- strain and response. Together these potentials are influenced by each training impulses in equal measure and affect performance in an antagonistic way. The components of the PerPot (Perl, 2001, 2004) have some similar theoretical characteristics to Banister et al, (Calvert et al., 1976; Morton et al., 1990) such as the concept of antagonist role of strain and response potential, whilst using a different computational approach.

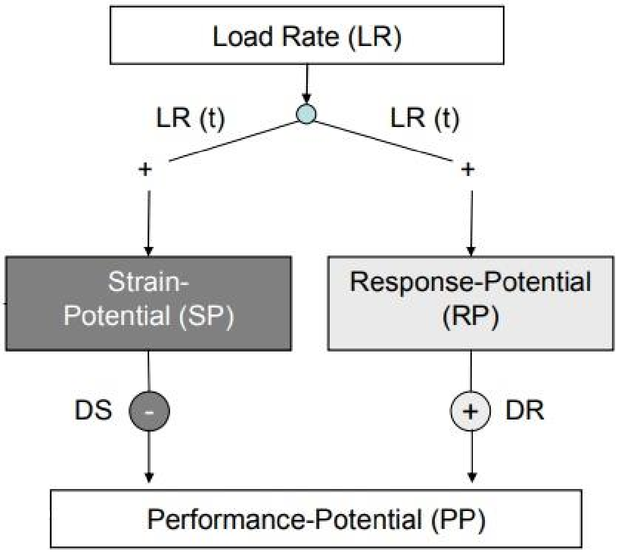


Figure 2. 4: Basic antagonistic structure of the Performance Potential metamodel. DS: delay in strain flow; DR: delay in response flow (Perl, 2001).

Although not presented formally as a framework, the joint consensus statement of the ECSS and the ACSM (Meeusen et al., 2013) can be seen as a theoretical framework presenting the continuum of the effects of training and individual/contextual factors on sport performance. According to the ECSS/ACSM consensus on overtraining there is a possible range of outcomes ranging from improvement, no change, and then a progression from functional to non-functional overreaching and finally over training (Meeusen et al., 2013) (Figure 2.5). Functional overreaching is used to improve performance and often applied during normal training cycles. It involves intensified training, which generally results in reduced performance, however after adequate rest an athlete's performance will improve relative to baseline levels. Conversely, if intensified training continues, without adequate rest, this may result in non-functional overreaching causing decrease in performance or stagnation potentially lasting for several weeks or more (Meeusen et al., 2013). Despite this, with appropriate rest an athlete should fully recover. Towards the end of the continuum is overtraining in which performance decrement may last for months. Therefore, from a practical perspective, it is essential to monitor acute and negative training effects in order to identify where an athlete is placed with this overtraining continuum. However, currently no physical training frameworks exist incorporating this overtraining continuum.

Process	TRAINING (overload)	INTENSIFED → TRAINING		
Outcome	Acute Fatigue	Functional OR (short term OR)	Non-functional Overreaching (extreme OR)	Overtraining syndrome (OTS)
Recovery	Days(s)	Days-weeks	Weeks-months	Months -….
Performance	Increase	Temporary performance decrement (e.g., training camp)	Stagnation decrease	Decrease

Figure 2. 5: European College of Sport Science and the American College of Sports Medicine joint consensus statement. Possible presentation of the different stages of training, overtraining and the overtraining syndrome (Meeusen et al., 2013).

Therefore, whilst there are physical training frameworks and models available, we have identified several areas that may be clarified and elaborated. This includes the lack of a positive and negative training effects in the previous framework (Impellizzeri et al., 2019; Impellizzeri et al., 2005), lack of clarity surrounding fitness and fatigue definitions (Calvert et al., 1976; Morton et al., 1990), further extension of the influence individual and contextual factors and absence of an overtraining continuum in all identified frameworks.

Summary

The preceding literature review highlighted a number of important findings regarding high injury rates, inadequate and poor quality of previous research and a complete lack of research regarding illness in dance. Secondly, we underlined the conceptual and methodological issues with training load and injury research in sport and the paucity of research in dance exploring these relationships. Third, we highlighted that whilst many AROMs are available in sport, the quality and level of evidence is yet to be investigated. Additionally, within dance, there are no validated tools for measuring training effects. Finally, we identified several concepts important in the conceptualization of the process and outcomes of physical training, with particular reference to subjective measures.

1. High injury rates exist in dance, however there is a lack of definitional consistency in the literature together with concerns regarding detail, methodology and quality in reporting.
2. Previous research in dance has identified various risk factors for injury, unsupported however by a consensus.
3. Some evidence suggests that training load management is a risk factor for injury in sport, although there are major issues with the methodology used in these studies. Currently, no epidemiologic data exist in dance regarding the incidence of injury together with training load over a long period of time, with detailed injury descriptions.
4. High intensity and prolonged training and competition load has been related with a heightened risk of illness in sport. In contrast, no research exists examining

illness in dance. Additionally, no studies have concurrently examined training load with illness in dance.

5. Although, a range of self-reports measures are currently used in athlete monitoring their psychometric properties and qualities are yet to be examined.
6. Studies examining the relationship between variables (e.g., training load or subjective measures such as ‘wellness’) with outcomes such as injury or illness are not based on a reference framework.
7. There are validated tools such as sRPE for quantification of internal training load. However, there are no validated instruments for measuring training effects in dance.

Chapter Three

A prospective study on injury, illness and training load in a professional contemporary dance company

Jeffries, A. C., Wallace, L., Coutts, A. J., Cohen, A. M., McCall, A., & Impellizzeri, F. M. (2020). Injury, Illness, and Training Load in a Professional Contemporary Dance Company: A Prospective Study. *Journal of Athletic Training*, 55(9), 967–976. https://doi.org/10.4085/1062-6050-477-19

Published online 20th August 2020.

Preface

Professional contemporary dance is a physically demanding and high-risk activity. Within sport, the cumulative effects of repeated and intense physiological and psychological training have been shown to disrupt immune system function, in particularly the increased risk of upper respiratory tract infections. Currently, no research exists in professional dance regarding the occurrence of upper respiratory tract infections. Therefore, the aim of *Study One* was to provide a detailed description of the nature and incidence injury, preliminary information of the occurrence of illness together with training load in professional contemporary dancers.

Abstract

Context: Professional dance is a demanding physical activity with high injury rates. Currently, no epidemiological data exists regarding the incidence of injury and illness together with training load over a long period of time.

Objective: To provide a detailed description of injury, illness and training load occurring in professional contemporary dancers.

Design: Descriptive epidemiology study.

Setting: A single professional contemporary dance company during a one-year period.

Patients or Other Participants: Sixteen male and female professional contemporary dancers.

Main Outcome Measures: Injury data included medical attention (Med-Inj) and time loss injury (Time-Inj). Illness was measured using the Wisconsin Upper Respiratory Tract Infection Survey. Training load was collected for each dance session using the session rating of perceived exertion. Training load was classified into three groups based on individual and group percentiles: low, medium and high.

Results: Seventy-nine injuries were reported (86.1% new, 6.3% re-injury and 7.6% exacerbation). The Med-Inj incidence rate 4.6 per 1000 hours (95% CI, 3.8-5.8) and Time-Inj rate 1.4 per 1000 hours (95% CI, 0.8-2.1). The median time till injury for Med-Inj and Time-Inj injury was 3 months. The number of days dancers experienced illness symptoms was 39.9 ± 27.9 [1-96], with an incidence rate of 9.1 per 1000 hours (95% CI, 7.7-10.7). Mean weekly training load was 6685 ± 1605 [4641-10391] (arbitrary units).

Inconsistent results were found for the incidence of injury and illness based on individual and group categorisation of training load.

Conclusion: Professional dancing is associated with both high injury and illness rates. This is worrying from a health perspective and underlines the need for further studies to understand how to decrease the risk. Training load is higher than other sport disciplines but whether the high incidence of injuries and illnesses is related to high training demands needs further investigations, possibly conducted as international, multicentre collaborative studies.

Key words: Dance, injury, illness, upper respiratory tract infection, training load.

Introduction

The injury incidence rates reported in professional dancers range from 0.16-4.44 per 1000 hours (h) exposure (Allen et al.; Bronner, McBride, & Gill, 2018; Bronner et al., 2003). This large variation may be due to a number of factors including injury definition, style of dance (i.e., contemporary or ballet), repertoire and tour schedule. However, research on injury incidence and risk factors in dance is both limited and methodologically inconsistent compared to many other athletic activities.

Previous research has demonstrated that the cumulative effects of repeated, intense physiological, and psychological physical training can disrupt immune system function (Hellard et al., 2015). In particular, the risk of upper respiratory tract infection (URTI) has also been shown to increase during periods of intensive training in elite athletes competing in a variety of sporting contexts (e.g., swimming, tennis, and team sports) (Cunniffe et al., 2011; Hellard et al., 2015; Novas et al., 2003). Furthermore, URTI is usually preceded by a prodromal period involving pathophysiological changes associated with fatigue, myalgia and headache. These early warning symptoms may be indicators of acute illness as well as symptomatic of overreaching and overtraining (Meeusen et al., 2013). Accordingly, monitoring athletes is essential in understanding the relationship between physical training and immunological changes. However, despite high psychophysiological demands, no research has investigated the incidence of URTI or the impact associated with illness in dance.

Inappropriate balance between training load and recovery has also been suggested to induce fatigue, abnormal training responses (Meeusen et al., 2013) and increased risk of injury and illness (Schwellnus et al., 2016). Despite substantial research examining training load and injury results, are inconsistent and appear to vary with the type and time frame of load measured (Eckard, Padua, Hearn, Pexa, & Frank, 2018). Only one study has examined injury (overuse) and training load in dance with results showing dancers without musculoskeletal pain had lower training loads (Boeding et al.). For example, the training load of dancers with self-reported symptoms of overuse-injury, was higher compared to dancers with no symptoms. However, caution should be taken when interpreting these findings due to methodological issues, including a short observation period (7 weeks) and missed data (time loss or medical attention injuries). Furthermore, data from single small cohorts cannot be generalised, however the combination of results from multiple cohorts may provide more insight and external validity. Therefore, further investigation is required that is longitudinal in nature, longer risk period, and comprehensively describes injury incidence and training load. This may assist our appraisal of the mechanisms of injury and the suggestions of preventive interventions for future studies.

To date, no studies in dance have examined injury, illness and training load in the same cohort utilising a prospective approach. Given the potential risks to the health of dancers (e.g., injury and illness) and their multifactorial nature, it is important to gain more data regarding the characteristics of injury, incidence of illness and training load volume completed over a season. This would provide information useful for the identification of potential factors for prognostic research. Therefore, the purpose of this study was to provide a detailed description of injury, illness and training load occurring in professional contemporary dancers.

Methods

Population

Active professional contemporary dancers from the same dance company were eligible to participate in the study, on a voluntary basis. Sixteen participants, 7 male and 9 females (participant baseline characteristics Table 3.1) agreed to participate in this study. Ethical approval was granted by the University of Technology Sydney, Human Research Ethics Committee before commencement of the study (HREC ETH17-1082). Each participant provided written informed consent and were familiarised with all procedures before the commencement of the study.

Study Design

This study used a prospective cohort research design in which injury, illness and training load data were collected during a one-year period 2018-2019, in Sydney Australia. The study location was at the dance company's training base or domestic and international performing venues. All injuries that were sustained during the surveillance period were recorded by the company physiotherapist, with no participants excluded from the analysis. Injury reporting and methodology was based on guidelines by the International Association of Dance Medicine and Science together with international consensus statements (Timpka et al., 2014). The reporting of this study follows the STROBE statement (Strengthening the Reporting of Observational Studies in Epidemiology) (von Elm et al., 2007).

Injury Definition, Classification, Severity, and Injury Rate

Injuries were defined as (i) Medical attention injury (Med-Inj): assessment of a participant's medical condition by a qualified physiotherapist (Timpka et al.). (ii) Time loss injury (Time-Inj); in which a participant was unable to dance for one or more days following the event (Bronner & Wood, 2017). Information was collected regarding injury status (new, no previous history; re-injury, repeat episode of a fully recovered injury; exacerbation, worsening in the state of a non-recovered injury) (Timpka et al.), mechanism of injury, location, injury occurrence, type of onset (Timpka et al.), report of cause, contact or non-contact, detailed injury assessment and exposure status.

Injuries were classified into 3 categories of severity according to the length of absence from full training with the day on which the injury occurred being counted as day zero: minor (1-7 days), moderate (8-28 days) and severe (>28 days). Dance exposure was defined as the time at which participants were at risk of injury (hours spent participating in dance class, rehearsal or performance). Injury incidence rate was calculated as all new Med-Inj per 1000 h and all new Time-Inj per 1000 h. The Med-Inj and Time-Inj incidence rate was also calculated for each training mode. Additionally, the Med-Inj and Time-Inj incidence rate was also calculated for group and individual daily and weekly training load percentiles (low, medium and high).

Illness

The Wisconsin Upper Respiratory Tract Infection Survey includes 10 items assessing symptoms, 9 items assessing functional impairments and 1 item assessing global severity and global change. For an URTI to be recorded, participants must have had upper respiratory signs and symptoms for ≥ 48 h. For calculation of illness incidence rate, participants were considered at risk of a new illness during dance training exposure minus the duration of each illness episode and minus the 7 days after each episode. Additionally, illness incidence rates were calculated based on group and individual daily and weekly training load percentiles (low, medium and high).

Training Load

Sessions were classified into the following training modes: ballet and contemporary class, rehearsals and performance. The internal training load for each session was calculated using the session rating of perceived exertion method (sRPE) for each dancer during the study period (Foster et al.). Session RPE was calculated as the product of exercise intensity and exercise duration. Exercise intensity was determined by sRPE according to the Borg CR-10 scale (Borg, Ljunggren, & Ceci) and was collected within 10 minutes after each class. The sRPE method has been previously validated as a measure of internal training load in pre-professional dancers (Jeffries et al.). In addition, the sum of the daily and weekly training load was calculated for individual and group percentiles (low, medium and high) and each type of session (ballet, contemporary, rehearsal and performance).

Data Collection

All injury, illness and training load data was prospectively collected and recorded in a centralised database (Edge 10, London, UK), using a custom designed App. Each injury was assigned a 4-character injury diagnosis code (Orchard Sports Injury Classification System 10.1). Injuries were time recorded according to the date of injury and all data linked by a unique identification code (Finch & Fortington, 2018). This study was part of a larger two-year project (surveillance system).

Analyses

Descriptive analyses were performed on injuries and illness sustained during the 2018 surveillance period. Injury and illness incidence rates were calculated with the following formula: number of new injuries/exposure in hours (individual hours collected from duration time calculated in sRPE) X 1000 and 95% confidence intervals (CI). Kaplan-Meier injury survival analysis was used to calculate the proportion of injury-free participants at monthly intervals for both Med-Inj and Time-Inj. All analyses were performed in Excel.

Results

A STROBE flow-chart of participants is presented in (Figure 3.1), indicating the sequential recruitment of participants.

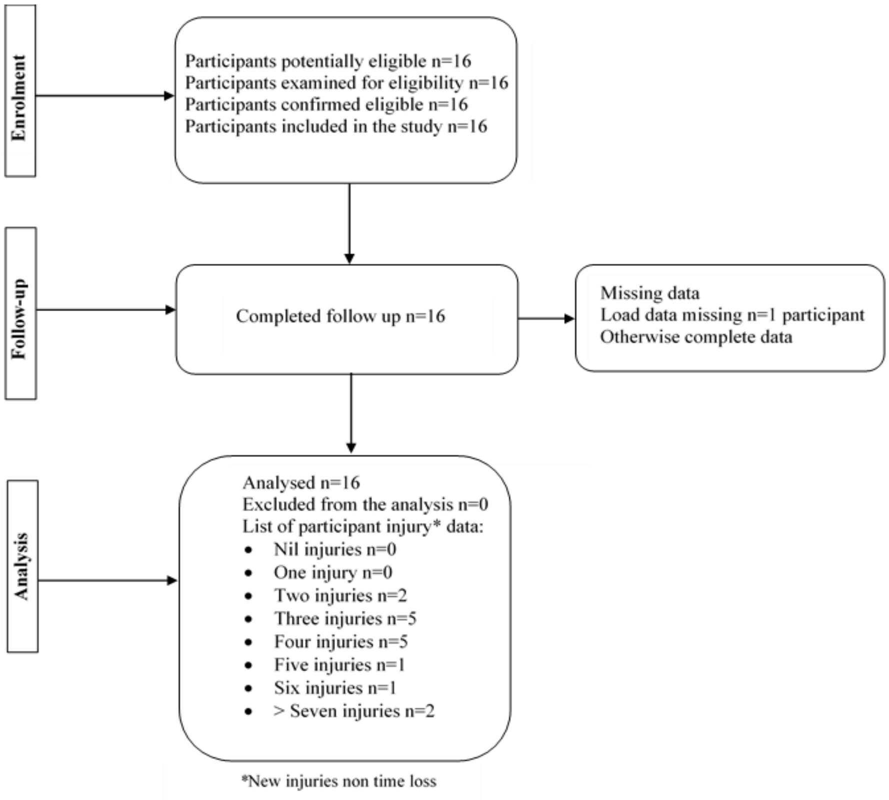


Figure 3. 1: Flow diagram of study design.* Injury: Medical attention injury (Med-Inj).

Participation and Injury Occurance

Sixteen professional contemporary dancers (9 female and 7 male) were under surveillance during a one year period over which 14,689 participation hours were completed. A total of 79 injuries were reported across the one year period (new: 86.1%, n=68; re-injury: 6.3%, n= 5; and exacerbation: 7.6%, n= 6) by the by 16 particpants (100%) with a Med-Inj incidence rate of 4.6 per 1000 h (95% CI, 3.8-5.8) and a Time-Inj incidence rate of 1.4 per 1000 h (95% CI, 0.8-2.1). A detailed summary of the injury distribution characteristics and incidence rates are presented in Table 3.2. The majority of reported injuries (70.6%, n= 59) resulted in no time loss from dance training or performance. The Med-Inj incidence rate for each session was: Ballet 4.4 per 1000 h (95% CI, 1.9-9.1), Contemporary 4.7 per 1000 h (95% CI, 2.3-9.2), Rehearsal 4.9 per 1000 h (95% CI, 3.6-

6.5) and Performance 3.1 per 1000 h (95% CI, 1.1-8.1). The greater risk difference was found between performance and contemporary: 2.1 (-2.6 to 6.7), p=0.58. The Time-Inj incidence rate for each session was: Ballet 0.0 per 1000 h, Contemporary 1.2 per 1000 h (95% CI, 0.3-4.6), Rehearsal 1.4 per 1000 h (95% CI, 0.8-2.2) and Performance 2.1 per 1000 h (95% CI, 0.7-6.5).

Table 3. 1: Participant baseline characteristics surveillance period 2018-2019

Characteristic	Males n = 7		Females n = 9		All n=16	
	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range
Age (y)	27.6 ± 5.5	19-34	24.9 ± 4.7	18-32	26.1 ± 5.1	18-34
Height (cm)	177.1 ± 7.9	170.9 - 191.2	161.7 ± 3.5	156.4 - 165.5	168.4 ± 9.6	156.4 - 191.2
Weight (kg)	72.1 ± 6.9	64.4 - 81.6	51.8 ± 4.0	47.2 - 60.2	60.7 ± 11.6	47.2 - 81.6
Professional Dance experience (y)	14.1 ± 4.8	6 – 20	21.2 ± 3.3	17 - 26	18.1 ± 5.3	6 - 26.0
Previous injury n (%)	6 (85.7)		9 (100)		15 (93.7)	

**Previous injury includes new injury, re injury or exacerbations.*

Table 3. 2: Injury distribution characteristics, participants injured, and incidence rates across 1-year (N = 16) [a]

	2018-2019		
	Male	**Female**	**Total**
Total participants	n=7	n=9	n=16
Injured	7	9	16
Non injured	0	0	0
Total injuries sustained			
Non time loss	24	44	68
Time loss	11	9	20
Injuries			
Median (range)	3 (2-8)	4 (3-11)	4 (2-11)
Incident rate (x 1000 hours) [b]			
Med-Inj (95% CI)	3.4 (2.2-5.0)	5.1 (3.8-6.8)	4.6 (3.8-5.8)
Time-Inj (95% CI)	1.6 (0.86-2.8)	1.0 (0.5-2.0)	1.4 (0.8-2.1)
No. of injuries			
0	0 (0.0)	0 (0.0)	0 (0.0)
1	0 (0.0)	0 (0.0)	0 (0.0)
2	2 (28.6)	0 (0.0)	2 (12.5)
3	2 (28.6)	3 (33.3)	5 (31.3)
4	2 (28.6)	3 (33.3)	5 (31.3)
5	0 (0.0)	1 (11.1)	1 (6.3)
6	1 (14.3)	0 (0.0)	1 (6.3)
7+	0 (0.0)	2 (22.2)	2 (12.5)
No. of time loss injuries			
0	1 (14.3)	5 (55.6)	6 (37.5)
1	2 (28.6)	2 (22.2)	4 (25.0)
2	3 (42.9)	0 (0.0)	3 (18.8)
3	1 (13.3)	1 (11.1)	2 (12.5)
4	0 (0.0)	1 (11.1)	1 (6.3)

[a] Values are presented as n (%) participants. [b] Number of new injuries per 1000 exposure hours. n, number.

Type of Injuries Sustained

The majority of Med-Inj sustained were to the knee (16.5%, n= 13), upper leg (15.2%, n=12) and torso (15.2%, n=12) which resulted in a median time loss period of 13 days (range 4-90), 3.5 days (range 1-10) and 6.5 days (range 2-14) respectively. The most common type of injury (69.6%, n= 55) was ligament/joint and muscle/tendon (26.6%, n=21) with a median time loss period of 4 days and 6 days (Table 3.3). Acute injuries were the most common type of injury accounting for 74.7% (n=59) with gradual injuries only 25.3% (n=20).

The majority of Time-Inj sustained were to the ankle (25%, n=5), upper leg (20%, n=4) and torso (20%, n=4) which resulted in a median time loss period of 21 days (range 5-28), 5 days (range 1-10) and 5 days (range 2-14) respectively. The most common type of injury was ligament and joint (65%, n=13). Acute injuries were most common accounting for 90% (n=18) with gradual only accounting for 10% (n=2).

Table 3. 3: Body region and tissue classification of new injuries

Tissue	Diagnosis	Frequency
Body region		Body region (%)
Bone Fracture/Stress reaction		**1 (1.3 %)**
Foot/toe	2nd MT stress reaction/fracture	1 (100 %)
Muscle/Tendon/bursa		**21 (26.6 %)**
Foot	Peroneal tendinopathy	1 (4.8 %)
	Webspace bursitis	1 (4.8 %)
Ankle	Achilles tendinopathy	1 (4.8 %)
	FHL Tenosynovitis	1 (4.8 %)
Lower Leg	Achilles tendinopathy	2 (9.5 %)
Knee	Patella tendinopathy	1 (4.8 %)
Upper Leg	Hamstring insertional tendinopathy	5 (23.8 %)
	Adductor & Quadricep fascial tear	1 (4.8 %)
	Quadricep fascial irritation	1 (4.8 %)
	Rectus femoris strain	1 (4.8 %)
	Biceps Femoris strain	1 (4.8 %)
	MTJ Hamstring strain	1 (4.8 %)
	Adductor longus strain	1 (4.8 %)
	Iliotibial band friction syndrome	1 (4.8 %)
Torso	External oblique strain	1 (4.8 %)
	Rectus abdonimis strain	1 (4.8 %)
Joint/Ligament		**55 (69.6 %)**
Toe	MTP joint sprain	3 (5.5 %)
Foot	Cuboid dysfunction	3 (5.5 %)
	Midfoot joint sprain	2 (3.6 %)
Ankle	Lateral ankle sprain	5 (9.1 %)
	Anterior ankle impingement	1 (1.8 %)
Knee	Patellofemoral pain syndrome	4 (7.3 %)
	Meniscus tear/irritation	3 (5.5 %)
	Meniscus irritation	1 (1.8 %)
	Fat pad impingement	2 (3.6 %)
	Tibiofemoral joint cartilage irritation	2 (3.6 %)
Hip	Hip impingement	8 (14.5 %)
	Hip instability	1 (1.8 %)
Pelvis	Sacro iliac joint dysfunction	2 (3.6 %)
Torso	Facet joint irritation	6 (10.9 %)
	Rib dysfunction	3 (5.5 %)
Shoulder	Shoulder impingement	3 (5.5 %)
	Shoulder dislocation	1 (1.8 %)
Elbow	Ligament sprain	1 (1.8 %)
Wrist	Scapholunate sprain	1 (1.8 %)
	TFCC sprain	2 (3.6 %)
Hand	Radiocarpal joint synovitis	1 (1.8 %)
CNS/PNS/Nerve		**2 (2.5%)**
Neck	Disc irritation	1 (50 %)
Torso	Neural tension	1 (50 %)

MT: Metatarsal; FHL: Flexor hallucis longus; MTJ: Musculotendinous junction; MTP: Metatarsophalangeal joint; TFCC: Triangular fibrocartilage complex

Injury Severity

Time loss injuries accounted for 29.4% (n=20) of all injuries resulting in 261 full days off dance. Sixty percent (n=12) of all-time loss injuries required the participant to take ≤ 7 days off dance, with 35% taking 8-28 days off and 5% > 28 days (Figure 3.2). Injuries requiring the greatest time off dance included: meniscal injury, ankle sprains and muscle strains.

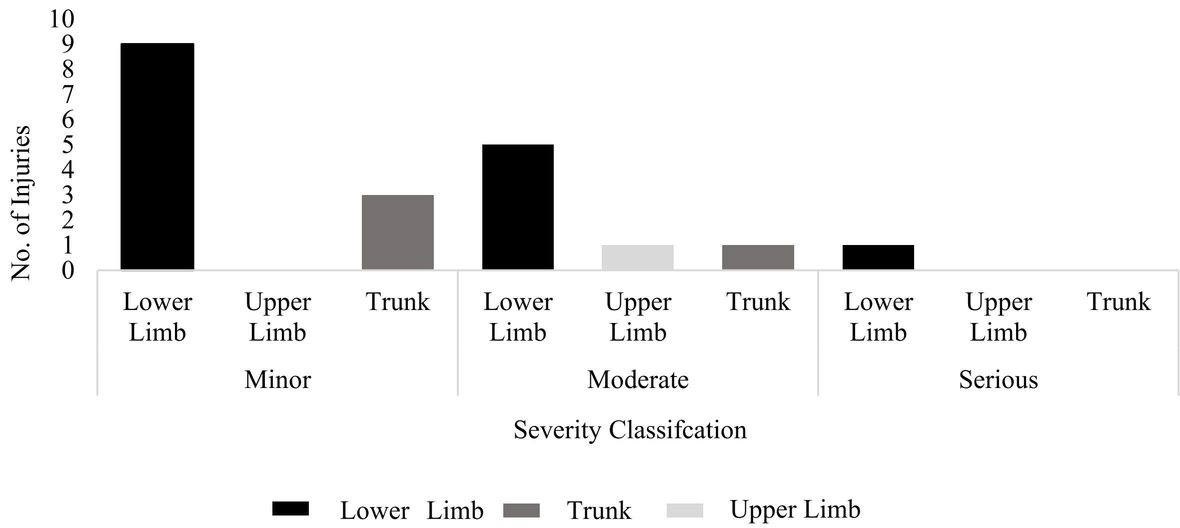


Figure 3. 2: The distribution of injury severity by injury location.

Time Till Injury

The starting point for the survival analysis was the commencement of the study. The participants were then followed until the first new injury occurred for both Med-Inj and Time-Inj injuries. Kaplan-Meier curves were used to estimate the survival function for the time to the first injury in all injured participants (Figure 3.3). All particpants were inury free at the commencement of the study.

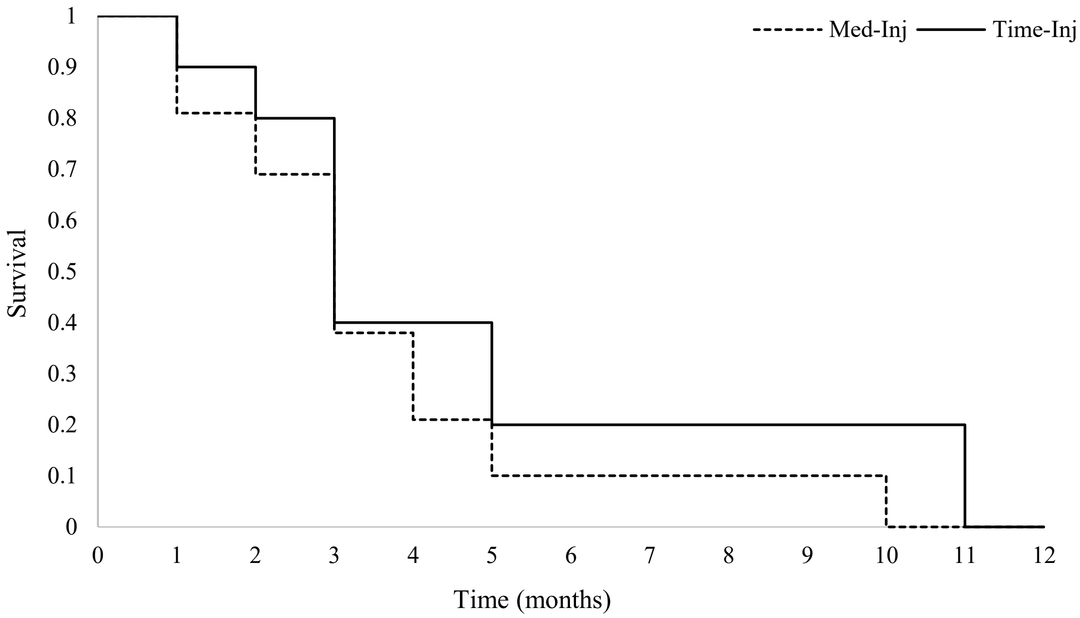


Figure 3. 3: Kaplan-Meier curve of the time till first injury for Medical attention injury (Med-Inj) as indicated by the dotted line (n=16) and Time loss injury (Time-Inj) as indicated by the continous line (n=10) in professional contemporary dancers during 2018. Survival: cumulative incidence probability. Vertical drop down lines signify when an injury occurred.

Illness

A total of 134 illness episodes were reported across the one-year surveillance period with no missing data. The illness incidence rate was 9.1 per 1000 h (95% CI, 7.7-10.7). The number of days participants experienced URTI symptoms were (mean ± SD, 39.9 ± 26.9; range 1-96).

Training Load

The average weekly training load (mean ± SD) was 6685 ± 1605 [Min-Max: 4641-10391] (AU). The average training load for individual sessions were; Ballet: 407 ± 120 [70-1200] (AU), Contemporary: 387 ± 153 [30-1620] (AU), Rehearsal: 578 ± 303 [20-1800] (AU) and Performance: 691 ± 350 [40-1800] (AU). The group average sRPE was highest in performance (RPE: 8.0 ± 1.4) and lowest for contemporary (5.0 ± 1.4) (Appendix A).

For the analysis of the effect of training load on injury there was complete data collected for injury and less than 10% missing data for training load. The missing data for training load was all from one participant (participant identification no. S3, injury n=4). The main analysis used single imputation (mean training load imputed for missing training load data) and the result are shown in Table 5. Sensitivity analysis were carried out as follows; (1) non imputed data (n=16), (2) imputation of maximum training load observed on that day or week respectively from the remaining participants entered as the missing value for dancer S3, (3) imputation of minimum training load observed on that day or week respectively from the remaining participants entered as the missing value for participant no. S3, and (4) non imputed data excluding participant no. S3 (n=15). The results of the sensitivity analysis were similar and did not change the interpretation of the data (data not shown). This suggests that the distribution of missing data is consistent with the assumption of the data being missing at random.

Injury and Training Load

Based on group training load percentiles the Med-Inj incidence in the high training load group was 5.7 per 1000 h, medium 3.6 per 1000 h (p=0.08) and low 3.2 per 1000 h (p=0.17). Based on participants' individual training load percentiles the greater daily Med-Inj incidence rate was low training load 6.6 per 1000 h (95% CI, 4.4-9.7) and the greater weekly injury incidence rate was medium training load 5.9 per 1000 h (95% CI, 4.2-8.3) (Table 3.4). Additionally, injury incidence rate was calculated for Med-Inj and Time-Inj injuries occurring during each type of session based on participants' individual training load percentiles. For Med-Inj ballet in the low training load percentile group had the greater injury rate 2.7 per 1000 h (95% CI, 1.0-7.31) and the smaller injury incidence rate 0.0 per 1000 h in the medium training load group. For Time-Inj performance had the greater incidence rate for all three categories of training load 0.7 per 1000 h (95% CI, 0.09-4.9) (Figure 3.4).

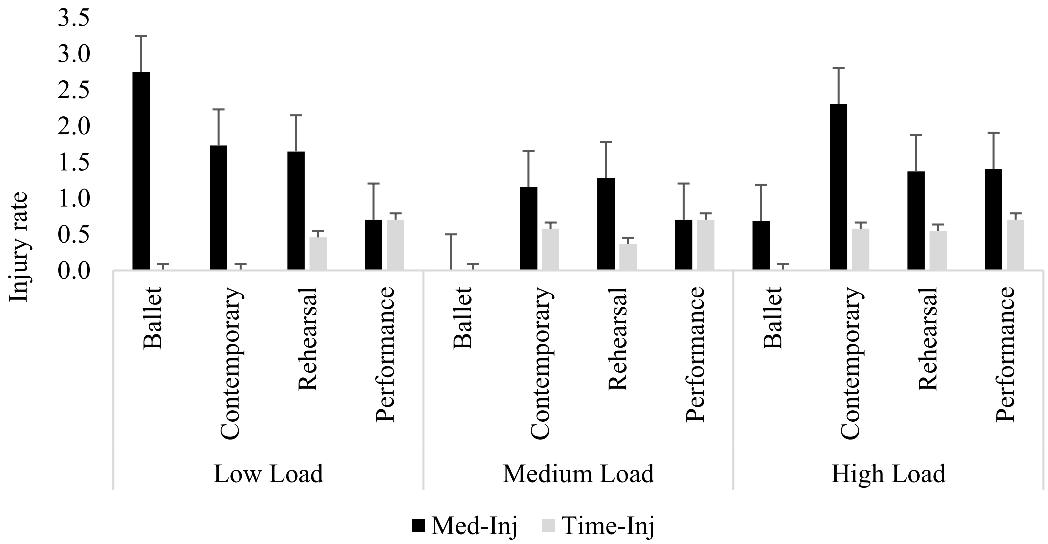


Figure 3. 4: Injury incidence rate per 1000 exposure hours (Medical attention injury: Med-Inj) and (Time loss injury: Time-Inj) for individual training load percentiles (low, medium and high) for each dance session: Ballet, contemporary, rehearsal and performance.

Table 3. 4: Individual Daily and Weekly injury and illness rate for load percentiles (mean imputation).

	Low		Medium		High	
	Daily	Weekly	Daily	Weekly	Daily	Weekly
Med-Inj	6.6	4.0	4.5	5.9	2.8	3.3
(95% CI)	(4.4-9.7)	(2.1-7.4)	(3.0-6.6)	(4.2-8.3)	(1.8-4.5)	(2.2-4.8)
Time-Inj	1.3	0.8	2.0	1.3	0.6	1.4
(95% CI)	(0.5-3.2)	(0.2-3.1)	(1.1-3.5)	(0.6-2.6)	(0.2-1.6)	(0.8-2.6)
Illness rate	11.1	9.6	6.8	9.1	8.5	7.8
(95% CI)	(8.2-15.0)	(6.4-14.2)	(4.9-9.4)	(6.9-11.9)	(6.5-11.0)	(6.0-10.0)

Illness and Training Load

The number of days participants experienced URTI symptoms in the high training load group (mean ± SD, 52.2 ± 26.6; range 19-96), medium training load group (mean ± SD, 47.8 ± 19.7; range 18-69) and low training load group (mean ± SD, 19.6 ± 21.3; range 1-60). The illness rate for the high training load group was 3.9 per 1000 h (95% CI, 2.9-5.0), medium training load group 3.4 per 1000 h (95% CI, 2.5-4.4) and low training load group 1.8 per 1000 h (95% CI, 1.2-2.6).

Discussion

The purpose of this study was to provide a detailed description of injury, illness and training load in professional contemporary dancers. This is the first study to comprehensively examine the combination of injury, illness and training load, in the same population, over an extended continuous surveillance period. The key findings of this study were (i) high injury rates, particularly Time-Inj (ii) high illness rates and duration of URTI symptoms (iii) median time to injury was 3 months (iv) injury and illness rate in relation to training load depends on the categorisation and reference method. Whilst a number of previous studies have examined injury rates in professional dancers (Bronner et al., 2018; Bronner et al., 2003; Nilsson, Leanderson, Wykman, & Strender, 2001), only one study has concurrently examined injury and training load. However this study was 7 weeks in duration and no time loss or medical attention injuries were reported (Boeding et al., 2019). Additionally, no studies have examined illness in dance.

Injury Occurrence

The present results reveal the entire cohort of participants (100%) acquired at least two Med-Inj injuries during the surveillance period, with several participants experiencing more than seven new injuries. The Med-Inj incidence rate (4.6 per 1000 h) is similar to previously reported prospective epidemiological studies in professional ballet (4.44 per 1000 h) (Allen et al., 2012). However, the Time-Inj incidence rate was higher (1.4 per 1000 h) compared to similar contemporary dance studies (0.16-0.22 per 1000 h) (Bronner et al., 2018; Bronner & Wood, 2017). Several factors may have contributed to these differences including repertoire, demographics and study duration. For example, previous research spanned a 15-year period compared to the one-year duration of this study (Bronner et al., 2018). In addition, in the current study the injury incidence rates for individual sessions ranged from 4.9 per 1000 h for rehearsal to 3.1 per 1000 h for performance. In contrast, a previous study in professional ballet dancers, reported lower injury incidence rates for rehearsal (2.43-2.99 per 1000 h) and higher in performance (4.45-5.19 per 1000 h) (Allen et al., 2012). This difference might be due to dance style and the intensity of rehearsal sessions. In our study 261 days of time lost injury were reported together with 447 days of training (in a modifed form) for dancers experiencing a medical attention injury. Furthermore, 4 dancers missed performances due to new injury and one due to re-injury over the surveillance period. This shows that dancers persist in training and performance despite the presence of injury or medical complaint.

Types of Injuries Sustained

In the current study, the most common type of injury was ligament and joint, predominantly affecting lower extremities, with hip impingement accounting for the majority of these injuries. In contrast, others reported that ligament and joint injuries only accounted for 35% of all injury types in a study of a similar cohort (Bronner & Wood, 2017). Additionally, despite overuse injuries being reported as the most common type of injury in dance (Bronner et al., 2003), the current study found overuse injuries only accounted for 25.3%. This variance may be due in part to differences in definitions of overuse injuries. For example, there is no current consensus on the definition of overuse injury, with some studies considering overuse as a mechanism of injury (Meeuwisse et al., 2007), whilst others use a diagnosis-based definition (Nilsson et al., 2001). Overall, our detailed report identified a wide range of injury types which should be considered when developing prevention programs or for examining causal associations in future aetiology studies.

Time till Injury

This is the first study to our knowledge to examine the time to event analysis of injury in professional contemporary dancers. The Kaplan Meier analysis involves calculation of the probability of injury at the time of occurence. In the current study, the median survival time without injury for both Med-Inj and Time-Inj was 3 months. Within the first five months of the surveillance period all but one participant experienced their first Med-Inj injury. Additionally, the injuries with the greatest severity (meniscus tear requiring surgical repair and grade three lateral ankle sprain) occurred in the first quarter of the year. Based on limited numbers, these results suggest that dance is a very high-risk discipline underlying the need for further research to support the implementation of prevention programs to protect the health of dancers.

Illness

Currently there are no studies regarding illness incidence in professional dancers, in particular URTI. Upper respiratory tract illness has been reported as the most common reason for non-injury-related presentations to sports medicine clinics, accounting for 35-65% of illness (Gleeson & Pyne, 2016). A major finding of this study was the high incidence of URTI with an illness incidence rate (9.1 per 1000 h) and 134 illness episodes. Additionally, the average duration participants experienced URTI symptoms was 39.9 ± 26.9 ranging from 1 day to approximately three months. Potential causes for high illness rates include training load management (Cunniffe et al., 2011; Hellard et al., 2015; Meeusen et al., 2013), international travel (Schwellnus et al., 2012) and lifestyle factors (Walsh et al., 2011). However, these measures were beyond the scope of this study with further research required to understand if these factors may influence the reduction of pathogen exposure and potential immunosuppression. It should be noted that no modifications to training or time loss were reported whilst dancers experienced an URTI. As such, persistent illness can have a negative impact on health and performance in dancers, particularly when undertaking high levels of strenuous exercise (Gleeson & Pyne, 2016).

Training load

The current study showed, as expected, that professional contemporary dancers undertake high training loads, especially in comparison with many other high-level athletes (Delecroix, McCall, Dawson, Berthoin, & Dupont, 2018; Veugelers et al., 2016). For example, weekly accumulated training loads in elite soccer 2994 AU (Delecroix et al., 2018) and average training loads in Australian football 4261 AU (Veugelers et al., 2016).

Additionally, the average weekly accumulated training loads (6685 ± 1605 AU) were much higher than those reported in pre professional contemporary dancers (4283 ± 2442 AU) (Jeffries et al., 2017) and adolescent ballet dancers (596 ± 153 AU) (da Silva et al., 2015). Collectively, this variation may be due to the increased training duration professional dancers undertake compared to pre professional dancers and football players. Although the average training load for individual sessions (contemporary, ballet and rehearsal) was similar to previous reports in pre-professionals (Jeffries et al., 2017), the dancers in this study undertook more sessions per day.

Injury and Training Load

In recent years many studies have examined the association between training load and injury in various sport disciplines (Eckard et al., 2018; Ehrmann, Duncan, Sindhusake, Franzsen, & Greene, 2016; Timoteo et al., 2018). However, most of these studies are underpowered since the number of injuries required to ascertain an association with training load is much higher than commonly used (van Smeden et al., 2016). Although the number of injury events in the current study is similar to previous publications, it is still inadequate for determining, in a reliable way, associations and hence we did not perform any predictive or associative analysis (Ehrmann et al., 2016; Timoteo et al., 2018). Rather, we accurately described the training load in order to help develop a conceptual framework and a causal structure that can serve for future epidemiological studies that may require international, and multicentre initiatives to establish an appropriate sample size.

An interesting finding of this study was participants in the high training load group experienced the highest injury incidence rate (5.7 per 1000 h) compared with medium and low (3.6 per 1000 h and 3.2 per 1000 h, respectively). This result is consistent with the findings of previous research in a variety of sports (Moller et al., 2017), but in contrast with other studies that have shown the opposite (lower injury risk corresponding to higher training load) (Brooks, Fuller, Kemp, & Reddin, 2008; Veugelers et al., 2016). These conflicting results may be due to a variety of methodological issues including low sample size, different injury definition and statistical approach. However, in the current study we also showed that the injury incidence rate in relation to the training load is dependent on the calculation method used. Most of the previous studies have used group data to create training load categories. For example, a study participant is placed in the high, medium or low category based on the training load completed by the whole group. However, there is no rationale for using the group training load as a reference and given the individual nature of injuries it would be reasonable to use individual data for categorisation, i.e., placing the individuals in the high, medium or low category based on the training load range they carried out. To our knowledge, only one study in cricket has categorised training load using individual data (Ahmun et al., 2019). Interestingly, in the current study when using individual training load, the Med-Inj injury rate was opposite in direction to the group training load injury rate, with dancers in the highest training load category experiencing lower injury rates. Therefore, a clear association within the training load level and injury rate cannot be shown and this is consistent with previous studies showing both higher and lower injury risk at high training load (Delecroix et al., 2018). Difference in categorisation methods, low sample size and sparse data bias may explain the inconsistencies of the current and previous studies (Eckard et al., 2018) and support the need for larger and hypothesis driven studies to avoid results simply reflecting “statistical noise” and potentially false discoveries.

Illness and Training Load

High training load has been proposed as a risk factor for illness in sport (Cunniffe et al., 2011; Hellard et al., 2015; Novas et al., 2003). Changes in internal and external training load have been associated with an increased risk of illness (Schwellnus et al., 2016). For example, increases in training volume have been related with increased risk of illness in elite swimmers (Hellard et al., 2015) and elite junior tennis players (Novas et al., 2003). Additionally, in team sports increased intensity of training preceded development of URTI symptoms (Cunniffe et al., 2011). In the current study, participants in the high and medium training load group had the highest illness incidence rate (3.9 and 3.4 per 1000 h, respectively) compared with the low training load groups (1.8 per 1000 h). The high training load group also experienced the longest duration of URTI symptoms compared with the medium and low groups. However, it is difficult to make comparisons between studies as there may be variations in methodology, characteristics of participants and travel schedules. In addition, there are many other factors that affect the immune system other than exercise. For example, nutrition, smoking, alcohol consumption, mental stress and lack of sleep have all been associated with the impaired immune function and increased risk of infection (Putlur et al., 2004). Regardless of the underlying cause of the URTI, a major concern, other than the health impact, is the accompanying fatigue that can limit or prevent training (Reid et al., 2004). It is recommended that future studies adopt multifactorial preventive strategies that may include clinical, training or lifestyle modifications (Walsh et al., 2011).

Similar to our injury results, when using individual training load, the illness rate was opposite in direction to the group training load illness rate, with dancers in the highest training load category experiencing lower illness rates. Therefore, depending on the method of categorisation, results in relation to training load are different. Without a strong rationale or conceptual framework for deciding the best way to categorise the training load, no conclusion can be drawn from the current study. Therefore, future research is warranted in the development of a conceptual framework which may explain particular causal phenomena and provide a basis for the interpretation of results.

There are some of limitations with this study. The sample size was restricted to the number of participants in the dance company. Accordingly, the goal of the current study was descriptive and exploration of associations between illness and injury with training load was presented to show the dependence of the results on the categorisation criteria. Although previous research has used similar sample sizes and number of events (Ehrmann et al., 2016; Timoteo et al., 2018), we believe this would provide unreliable results and would not satisfy the requirements for such a study (Eckard et al., 2018), that should also take into account the multifactorial nature of injuries and illnesses. However, these results increase the body of knowledge in which to create causal structures that can be verified by future studies.

The main aim of this study was to contribute to the accumulation of the body of knowledge used to identify potential mechanisms and prognostic factors. However, from a practical point of view our results suggest that a possible area of intervention is training load distribution by applying training principles. Indeed, few recovery periods occurred within and between macrocycles (weeks) as shown in Appendixes B and C (mean daily and weekly training loads). Considering professional dancers may need to undertake high training loads in order to achieve optimal performance, modifying the training load distribution and including recovery days and short periods of reduced load are potential immediate interventions. Furthermore, understanding training load distribution may require monitoring strategies (Impellizzeri et al., 2019), as well as injury and illness surveillance systems. Given dancers stoic nature and strong work ethic (e.g., continue to work when ill) may mean practitioners need to carefully monitor illness incidence and actively modify training loads accordingly.

Recommendations for future research

The goal of this study was to provide details on the nature and incidence of injury and preliminary information on illness occurrence. These results together with the training load experienced by the dancers can be used to develop and hypothesise an etiological framework and hence a causal structure (e.g. Directed Acyclic Graph) (Joffe, Gambhir, Chadeau-Hyam, & Vineis, 2012). This is necessary if the goal of future observational studies, as is common in sport and clinical settings, is to manipulate a variable to alter the likelihood of an event (e.g., injury and illness). Although causal inference from observational studies inevitably does not ensure causation (experimental studies are needed), it would provide some support to a causal link and, if strong enough, may support interventions. The variety of injury typology and nature found on this cohort also suggests that while the final mechanical causes are likely similar (stress and strain superior to the strength of the structure) the antecedent causes may be different. Combining all the injuries to determine associations with training load is not a reasonable approach as it would assume the same causal path for all the type of injuries experienced by the dancers. Although no relation between training load and injury and illnesses could be demonstrated, this association is worth exploring in future studies using appropriate statistical methods and sample size. Therefore, other than a conceptual framework to develop a causal structure, multicentre studies are necessary. Furthermore, prediction models for risk stratification (that do not require causation) are also warranted.

Conclusion

In this study, professional dancers experienced high training loads relative to other sports, and concomitantly high injury and illness incidence and risk. Interestingly, dancers in this study missed very few performances, despite injury and illness. Furthermore, dancers continued with training, albeit modified, even when affected by medical attention injuries and illness. Collectively, this suggests that dancers persist with training and performance despite the presence of injury and illness and may reflect their high level of commitment. However, the health of dancers should be a priority and the development of preventative interventions and education initiatives encouraged.

Funding

No funding was received for this research project.

Chapter Four

Athlete Reported Outcome Measures for monitoring training responses: a systematic review of risk of bias and measurement property quality according to the COSMIN guidelines

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*, *15*(9), 1203-1215. https://doi.org/10.1123/ijspp.2020-0386

Published online 25th September 2020.

Preface

The original aim of this thesis was to examine associations of injury and illness incidence with training load and subjective measures in professional contemporary dancers. When the initial project was designed the use of subjective measures (such as wellness) was a common approach in research and therefore exploration of the associations between injury and illness with subjective measures and training load formed the initial goal of *Study One*. Given the duration of *Study One* (two years) together with the timeframe for thesis completion (3 years), a preliminary examination of the literature revealed that training load and subjective measures were common predictive variables used within research. As such, given the above time constraints, subjective measures formed part of the initial data collection of *Study One*. However, upon a more detailed examination of the literature we identified some subjective measures, such as the wellness items, did not appear to be validated. Consequently, the project evolved to an examination of the validity of commonly used subjective measures used in the monitoring of training effects in athletes and dancers.

Abstract

Background: Athlete Reported Outcome Measures (AROMs) are frequently used in research and practice but no studies have examined their psychometric properties.

Objectives: Part 1) Identify the most commonly used AROMs in sport for monitoring training responses; Part 2) assess risk of bias, measurement properties and level of evidence, based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

Study appraisal and synthesis methods: Methodological quality of the studies, quality of measurement properties and level of evidence was determined using the COSMIN checklist and criteria.

Results: Part 1) From 9446 articles screened for title and abstract, 310 out of 334 full texts were included. 53.9% of the AROMs contained multiple items, while 46.1% contained single items. Part 2) From 1895 articles screened for title and abstract, 71 were selected. Most measurement properties of multiple item AROMs were adequate, but content validity and measurement error were inadequate. With the exclusion of two studies examining reliability and responsiveness, no validity studies were found for single items.

Conclusions: The measurement properties of multiple item AROMs derived from psychometrics were acceptable (with the exclusion of content validity and measurement error). The single item AROMs most frequently used in sport science have not been validated. Additionally, non-validated modified versions of the originally non-validated items are common. Until proper validation studies are completed, all conclusions based on these AROMs are questionable. Established reference methods such as those of clinimetrics should be used to develop and assess the validity of AROMs.

Keywords: Self-reports, questionnaire, measurement properties, risk of bias, quality.

Introduction

Traditionally, interest regarding athlete's subjective responses to exercise has been related to overtraining in sport. The most common application has been in the detection of early symptoms, or as a diagnostic tool, to subsequently inform decision-making and athlete management. In recent times the use of subjective responses within athlete monitoring has increased exponentially (Cormack & Coutts, 2016). A recent systematic review provided a critical appraisal of athlete monitoring methods including both subjective and objective measures (Saw et al., 2016). The authors suggested that subjective measures had superior sensitivity and consistency than objective measures and provide a more accurate reflection of changes to acute and chronic training loads. Therefore, given the training process is athlete-centred, and the ultimate goal is understanding how an athlete is tolerating training demands, subjective measures seems to provide the ability to measure constructs and dimensions that are not objectively measurable.

Patient Reported Outcome Measures (PROMs) are commonly used in clinical research and belong to the research field termed clinimetrics. Accordingly, international guidelines and initiatives have been established for the development and validation for standardising instruments used in clinical settings (e.g., NIH-funded initiative PROMIS). Clinimetrics was originally founded by Alvan R. Feinstein in the 1980's and is based on psychometric principles and methods. It refers to "the domain concerned with indexes, rating scales, and other expressions that are used to describe or measure symptoms, physical signs, and other distinctly clinical phenomena in clinical medicine" (Feinstein, 1987). These definitions reflect how subjective measures are also used in sport science to monitor the training process and therefore, clinimetrics can represent the strongest reference scientific area for sport science. With > 40-year history, the field of clinimetrics has well-established methods and guidelines. In our review, we refer to the subjective measurement of the training response as Athlete Reported Outcome Measures (AROMs) which is an adaption of the term PROM to athletes. With AROM we indicate self-report measures or questionnaires used in the assessment of the responses to training and competition (e.g. symptoms and physical signs).

Any measure used in research and practice should be valid in order to provide accurate, meaningful, useful, clear and specific information. These properties are especially critical when using these measures to develop models and understand phenomena (i.e., to avoid garbage in – garbage out). There are several AROM measurement properties that need to be assessed to understand whether their quality support their use. One of the essential attributes is validity, i.e., the extent a measure actually assesses what it is intended to measure. Therefore, to ensure AROMs measure the particular domains/constructs of interest, proper development and validation is imperative (Saw et al., 2017). Within sports science there are minimal guidelines for the development of AROMs. As such, we refer to COnsensus-based Standards for the Selection of health Measurement Instruments (COSMIN: http://www.cosmin.nl), a clinimetric-based initiative. COSMIN was developed by an international multidisciplinary team with backgrounds in epidemiology, psychometrics, medicine, qualitative research and health care, who have expertise in the development and evaluation of outcome measurement instruments. The COSMIN guidelines also provide instruments for assessing the risk of bias of the studies, quality of measurement properties and level of evidence.

Whilst a previous review (Saw et al., 2016) provided initial examination of athlete self-report measures, we extended and complemented this research with the first comprehensive assessment of AROM measurement properties. In part one, we systematically searched and retrieved the most commonly used AROMs (in research) for monitoring training responses in sport settings. In part two, as no previous systematic reviews have assessed the quality of instruments as common in other fields (Stephan, Mainzer, Kummel, & Impellizzeri, 2019), we examined the risk of bias, quality of the measurement properties and level of evidence of identified AROMs according to the COSMIN guidelines.

Methods

Study design and registration

Our manuscript was reported according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) (Shamseer et al., 2015). A protocol was prepared for this systematic review and pre-registered with the Open Science Framework registration DOI (10.17605/OSF.IO /J5G32). The review was conducted according to the COSMIN methodology (Mokkink et al., 2010).

Part One

Search strategy

To identify all available literature, a systematic review was conducted using the following electronic databases: Pubmed, Medline, EMBASE, PsycINFO, CINAHL, and SPORTSDiscus from inception up until (and including) December 2018. A further update was made in January 2020. We developed the search strategy after consultation with a university librarian. The search was conducted using medical subject heading (MESH) terms and text words (or synonyms) for the population of interest (athlete* OR player*), AND exposure (training OR overtrain* OR overreach*), AND outcome of interest (questionnaire* OR self-report* OR percept* OR self-evaluation* OR subjective rating).

In addition, references of included articles were searched manually for any further records.

Inclusion and exclusion criteria

Our criteria for article inclusion was as follows: AROMs monitoring responses to physical training in athletes of any gender, even in combination with other interventions (nutritional, psychological, etc.), all research designs, AROMs including self-reports, questionnaires, perceptual scales, no language restrictions, amateurs or professionals, and any sport. Articles in languages other than the review teams mastered (English, Italian, French, Spanish and Portuguese) were translated in the authors' network or google translation. The exclusion criteria were as follows: athletes with disease symptoms or diagnosis (e.g., cancer, diabetes, etc.); elderly (> 65 years) and children (< 16 years), reviews (any kind) or meta-analysis.

Data extraction

Titles and abstracts, and eligible full texts were screened by two independent raters (AJ and LW). Disagreements regarding inclusion and exclusion were discussed and resolved through consensus. A third reviewer (FI) was consulted if disagreements remained unresolved. Reasons for exclusion were noted and a flow chart of study article selection was prepared according to PRISMA (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). We divided instruments into multiple item (e.g., multiple item or combined score, used in traditional psychometric instruments) and single item (e.g., single items or self-developed multiple item measures with or without a combined score).

Part Two

Search strategy

To identify all available literature on the psychometric (measurement) properties, a second systematic review was performed in Pubmed, PsycINFO and Web of Science from inception up until (and including) January 2020. The search included MESH terms and text words (or synonyms) for the construct of interest (any identified instruments from part one), the population of interest (athlete* OR player*) and instrument of interest (questionnaire* OR self-report* OR percept* OR self-evaluation* OR subjective rating). For measurement properties, a validated search filter for studies on measurement properties (Terwee, Jansma, Riphagen, & de Vet, 2009) was applied in PubMed. In PsycINFO and Web of Science non-validated search filters were used. As there was such a low yield of results from the COSMIN validated search strategy, a second modified search was performed together with a manual search in Google Scholar.

Inclusion and exclusion criteria

Studies were included that reported on original data regarding at least one measurement property as defined in the COSMIN taxonomy (Mokkink et al., 2010), AROMs that monitor or examine athlete responses to physical training and all languages. Exclusion criteria were studies only available as abstracts, theses, studies involving cancer, diabetes, elderly or children (classified as < 16 years, however an exception was made if an original development study was conducted on children) and reviews (any kind).

Data extraction

First, descriptive data was extracted from the included papers: AROMs used, study setting, study population and sample size. Second, data was extracted on the measurement properties of the AROMs, which included: content validity (from qualitative research and or any theoretical framework), structural validity (determined using factor analysis), internal consistency of scales, construct validity, reliability, and responsiveness. One review author (AJ) extracted information and a second author (FI) was available if there was doubt on eligibility criteria. Reasons for exclusion were noted and a flow chart of study article selection was prepared according to the PRISMA statement (Moher et al., 2009).

Assessment of methodological quality

The methodological quality was assessed using the COSMIN checklist (Mokkink et al., 2012; Mokkink et al., 2010; Terwee et al., 2012). Based on the COSMIN taxonomy, measurement properties are divided into three domains: reliability, validity and responsiveness (Mokkink et al., 2010). Additionally, these properties are further expanded to a nine-box checklist (1) internal consistency, (2) reliability (test–retest), (3) measurement error, (4) content validity, (5) structural validity, (6) hypotheses testing, (7) cross-cultural validity, (8) criterion validity and (9) responsiveness. Each box contains numerous items with standards for the study design and statistical analysis for the measurement property. Items can be scored on a four-point scale as poor, fair, good or excellent. An overall score of each study was then determined by the 'worst score counts' method. Due to the absence of an established gold standard measure, criterion validity was not evaluated but instead analysed as construct validity. Predefined spreadsheets from COSMIN were used to systematically collect data on study characteristics, measurement properties and methodology.

Data analyses

Data analysis was conducted in three steps following the COSMIN methodology.(Prinsen et al., 2018) First, the methodological quality of eligible studies was performed using the COSMIN checklist (Mokkink et al., 2010; Terwee et al., 2012) for assessment of methodological quality of studies on measurement properties. Methodological aspects concerning statistical methods and design requirements, specific to the measurement properties were rated on a four-point scale: "very good", "adequate", "doubtful", and "inadequate". Based on the COSMIN recommendations, any methodological aspect assessed, the lowest rating per measurement property was obtained.

Second, the criteria for good measurement properties was applied to the results of the included studies following the COSMIN guidelines for systematic reviews of PROMs (Prinsen et al., 2018; Terwee et al., 2007). The measurement properties of each individual study was evaluated according to criteria based on (Terwee et al., 2012) and rated as "sufficient" (+), "insufficient" (-), "inconsistent" (±) or indeterminate (?) depending on the design, methodology and outcomes (Terwee et al., 2007). These results were then collectively summarized to obtain an overall score across all included studies "sufficient" (+), "insufficient" (-), "inconsistent" (±) or indeterminate (?).

In the final step, the overall rating of evidence was accompanied by a level of quality of the evidence with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach from the COSMIN methodology, to grade the confidence in the total body of evidence available for the measurement properties (Prinsen et al., 2018). The quality of the evidence was graded as high, moderate, low, or very low and based on (i) risk of bias, (ii) indirectness, (iii) inconsistency of results, and (iv) imprecision of studies. One reviewer (AJ) assessed all studies with ratings confirmed by a second reviewer (FI). Detailed information on the COSMIN standards can be found elsewhere (Terwee et al., 2012).

Results

Part one

The search resulted in 21,003 publications of which 9446 were screened for title and abstract after removal of duplicates. Based on the inclusion and exclusion criteria a further 9160 studies were excluded leaving a total of 334 studies for full text screening. Twenty-four studies were excluded due to either no full text availability or age group < 16 years. In total 310 full text studies were included (Figure 4.1) of which 116 independent AROMs were reported 471 times. Most studies identified using AROMs have been conducted over the past 50 years, with an increase in studies (49%, n=151) during the last 5 years.

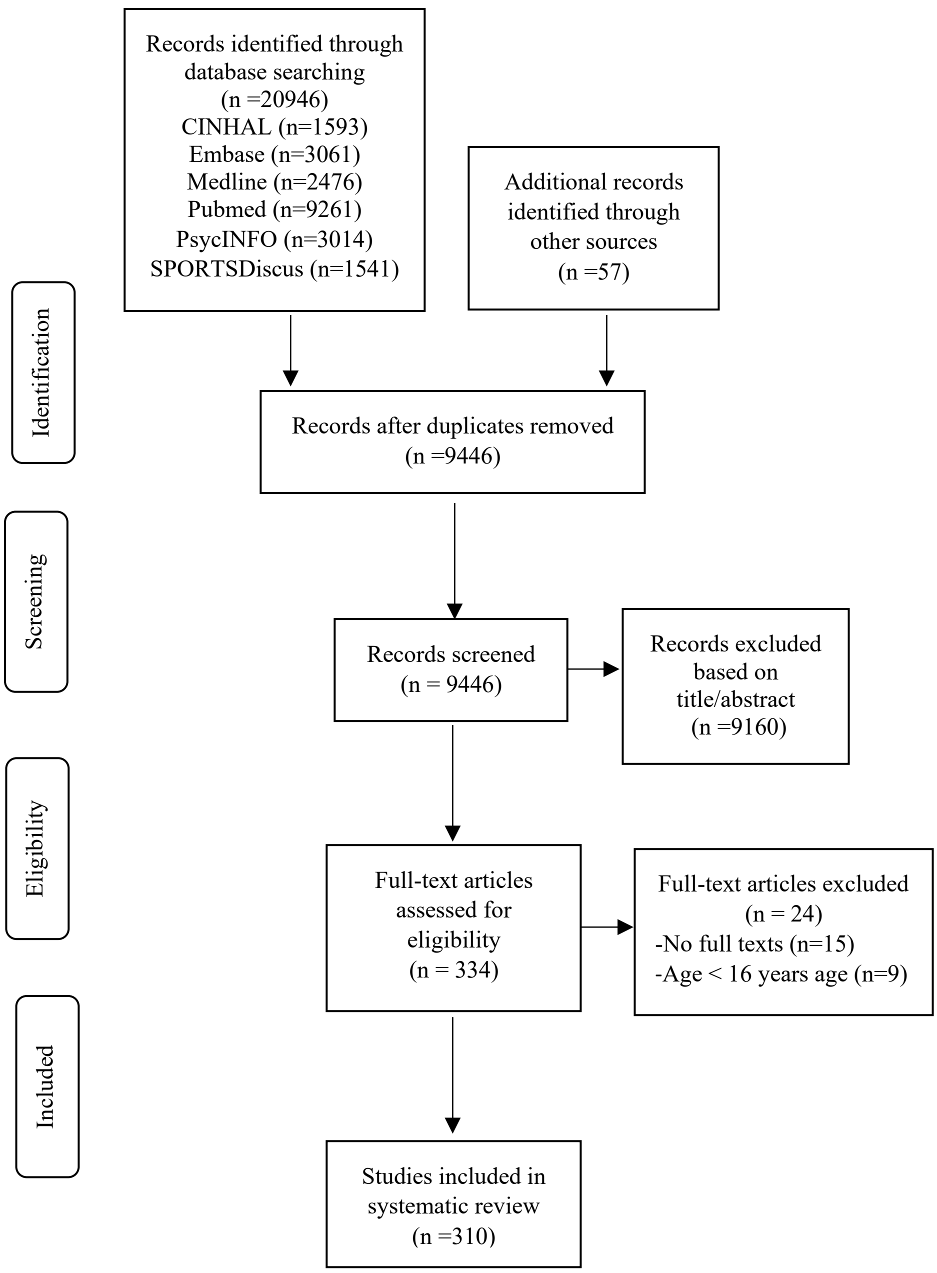


Figure 4. 1: Preferred reporting for systematic reviews (PRISMA) flow diagram part one (Moher et al., 2009).

Most common AROMs

Multiple item

Table 4.1 summarizes the most common (used more than twice) multiple item and single item AROMs. Several AROMs have been subject to modifications, therefore we subdivided into original version and modified/partial. The top three AROMs were the Profile of Mood State (POMS) (19.1%, n=90), Recovery-Stress Questionnaire Athletes (RESTQ-Sport) (11.6%, n=55) and the Daily Analysis of Life Demands for athletes (DALDA) (4.0%, n=19).

Single item

Wellness items (including Hooper questionnaire) (16.9%, n=80), muscle soreness or pain (5.9%, n=28), and recovery (5.3%, n=25) were the most common single item measures. Further analysis of the single item domains presented in Table 4.2 and demonstrates the variation of terminology used within athlete monitoring literature.

Table 4. 1: Instruments used more than twice (% of the total amount of AROMs).

Most commonly used athlete monitoring questionnaires	**Frequency (n)**	**Percentage (%)**
Multiple item		
Profile of Mood States	90	19.1
Original Profile of Mood States	53	11.2
Partial Profile of Mood States	37	7.9
Recovery Stress Questionnaire for Athletes	55	11.6
Original Recovery Stress Questionnaire for Athletes	44	9.3
Partial Recovery Stress Questionnaire for Athletes	11	2.3
Daily Analysis of Life Demands for Athletes	19	4.0
Original Daily Analysis of Life for Athletes	18	3.8
Modified Daily Analysis of Life for Athletes	1	0.2
Overtraining questionnaire of the Societe Francaise de Medecine du Sport	12	2.5
Short Recovery and Stress Scale	11	2.3
Athlete Burnout Questionnaire	11	2.3
Acute Recovery and Stress Scale	8	1.6
Multi-Component Training Distress Scale	5	1.1
Fatigue questionnaire*	4	1.1
Perceived Stress Scale	3	0.6
Sleep log	2	0.4
Lake Louise Acute Mountain Sickness Questionnaire	2	0.4
The 24-hour history	2	0.4
Single item		
Wellness items	80	16.9
Wellness items	75	15.9
Original Hooper Questionnaire	5	1.1
Muscle soreness or pain	28	5.9
Recovery	25	5.3
Total quality of recovery scale	14	2.9
Other	11	2.3
Sleep	23	4.8
Fatigue	22	4.6
Other	25	4.6

Multiple item: includes multiple or combined score (e.g., traditional psychometric instruments). Single item: includes single items or self-developed multiple item measures with or without a combined score. Hooper questionnaire also referred to as training diary, Hooper index, and Hooper score, Subjective wellbeing, Perceived ratings of wellness, Psychometric measures of wellness, Wellness psychometric questionnaire, Daily subjective questions, Well-being questionnaire, Questionnaire adapted by McLean et al, Athlete Self-Report Measures, Perceptual Measures, Hooper-Mackinnon questionnaire, Training log, Ratings of well-being, Player wellness questionnaire, Players wellbeing questionnaire, Wellbeing monitoring, Welfare Questionnaire, Wellness measure/Hooper index, Daily well-being scale, Wellbeing indices, Mclean Perceived Welfare Questionnaire, Perceived scaled questionnaire. *Fatigue questionnaire also called short questionnaire of fatigue based on several different authors (Atlaoui et al., 2004; Chatard et al., 2003). Single item 'other' includes training difficulty, mood state, energy levels, stress, health status, physical self-efficacy, well-being, Hardy and Rejeski scale, body perception and anxiety.

Table 4. 2: Single items (% of the total amount of AROMS).

Domain	Frequency (n)	Percentage (%)
Muscle soreness or pain	**28**	**5.9**
Muscle soreness	19	4.0
Delayed onset muscle soreness	2	0.4
Leg perceptions	1	0.2
Pain	1	0.2
Perceived muscle pain scale	1	0.2
Perceived muscle soreness	1	0.2
Perceived pain	1	0.2
The muscle perception scale	1	0.2
The muscle sluggishness scale	1	0.2
Fatigue	**22**	**4.6**
Fatigue	18	3.8
General fatigue	1	0.2
Perceived fatigue	1	0.2
Subjective feeling of accumulated fatigue	1	0.2
Tiredness	1	0.2
Recovery	**25**	**5.3**
Total quality recovery scale	14	2.9
Perceived recovery status scale	3	0.6
Recovery	2	0.4
Mental recovery	1	0.2
Modified total quality recovery action scale	1	0.2
Perceptual recovery	1	0.2
Physical recovery	1	0.2
Sleep	**23**	**4.8**
Sleep quality	13	2.7
Karolinska sleepiness scale	2	0.4
Lack of sleep per week	1	0.2
Self-reported sleep perception	1	0.2
Sleep diary	1	0.2
Sleep restfulness	1	0.2
Epworth sleepiness scale	1	0.2
Other	**25**	**4.6**
Hardy and Rejeski scale	2	0.4
Physical self-efficacy scale	2	0.4
Stress	2	0.4
Well-being	2	0.4
Energy levels	1	0.4
Mood state	1	0.2
Self-reported health status	1	0.2
Self-reported body perception	1	0.2
Sport competition anxiety test	1	0.2
Training difficulty	1	0.2

Part two

Search results

In total 73 studies were included in this review (Figure 4.2) reporting measurement properties for 26 AROMs: RESTQ-Sport, Acute Recovery Stress Scale (ARSS), Short Recovery Stress Scale (SRSS), POMS, DALDA, Athlete Burnout Questionnaire (ABQ), Perceived Stress Scale (PSS), Total Quality Recovery Scale (TQR), Overtraining questionnaire of the Societe Francaise de Medecine du Sport (OFSM), Multi-Component Training Distress Scale (MTDS), and wellness and single items. It should be noted that many of these instruments are modified or shortened versions of the original, which according to the COSMIN guidelines should be treated as a new AROM. Of the 26 AROMs, the RESTQ-Sport had one shortened version, the POMS had seven shortened versions, DALDA had one modified version and the ABQ had two shortened versions. The quality of the evidence for measurement properties of the AROMs, developed mainly for the athlete population, together with the wellness items are presented in Table 4.3. The quality of evidence for the remaining AROMs is presented in the Appendix D and the results of each AROM version can be found in the Appendix E in order to reduce the length of the paper. For each AROM, a summary for each measurement property examined has been presented.

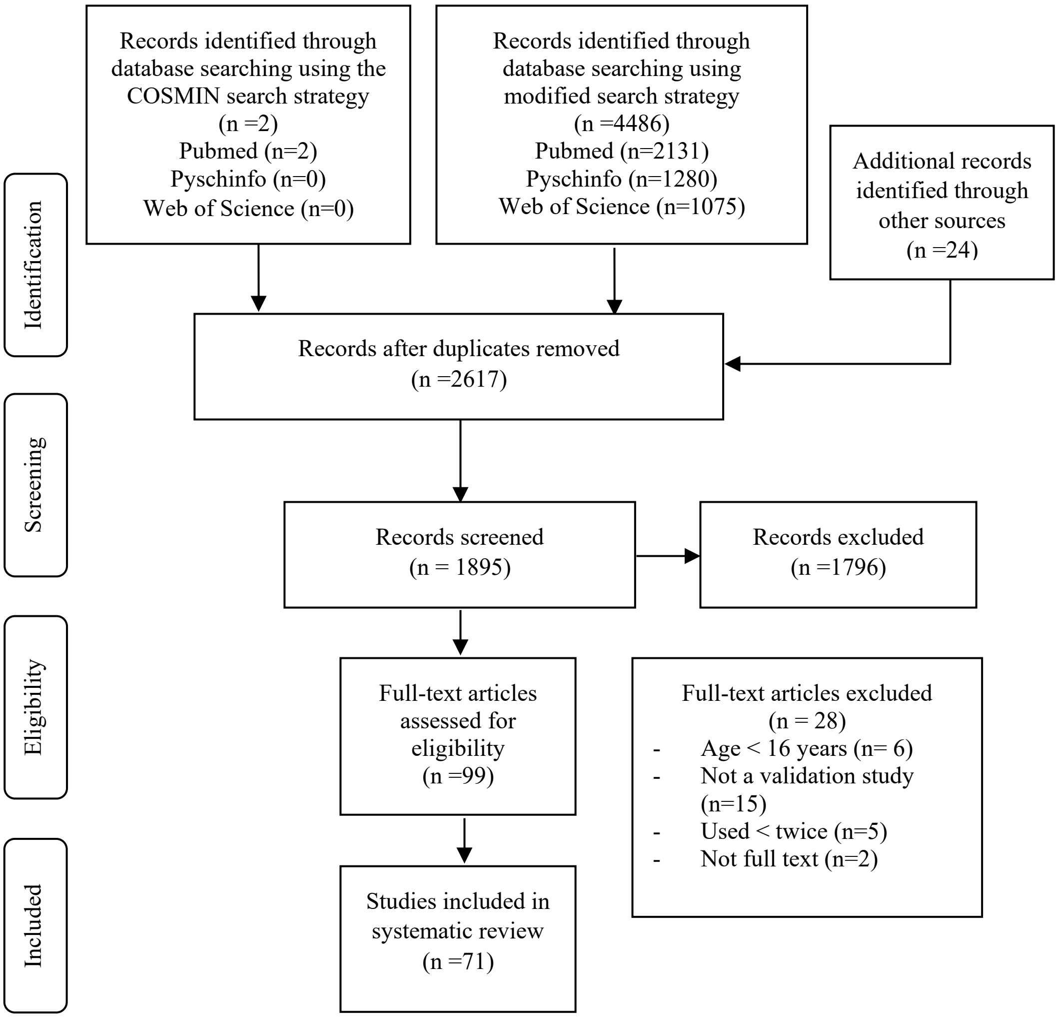


Figure 4. 2: Preferred reporting for systematic reviews (PRISMA) flow diagram part two (Moher et al., 2009).

Table 4. 3: Quality of the evidence for measurement properties of the AROMs

	RESTQ 76 (Kallus & Kellmann, 2016)		RESTQ 36 (Nicolas, Vacher, Martinent, & Mourot, 2019)		ARSS (Kellmann, Kolling, & Hitzschke, 2016)		SRSS (Hitzschke et al., 2015)		Single Item-muscle soreness (Impellizzeri & Maffiuletti, 2007)		MTDS (Main & Grove, 2009)		Wellness (Crowcroft, McCleave, Slattery, & Coutts, 2017; Fitzpatrick, Hicks, Russell, & Hayes, 2019; Roe et al., 2016)	
	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low
Content Validity	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low		
Relevance	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low		
Comprehensiveness	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low		
Comprehensibility	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low		
Structural validity	+	High	+	High	+	High	+	High			-	Moderate		
Internal consistency	+	High	?	Moderate	+	High	+	High			+	High		
Cross-cultural validity			+	Low	+	High	+	High						
Measurement invariance														
Reliability	-	Very low											?	Low
Measurement error													?	Low
Construct validity	+	Moderate			?	Moderate	+	High	+	Low	+	High		
Responsiveness			+	Moderate	+	High	+	High					+	Low

“ +” = sufficient, “ - “ = insufficient, “ ±” = inconsistent,” ?” = indeterminate.
RESTQ: Recovery-Stress Questionnaire Athletes-Sport; ARSS: Acute Recovery and Stress Scale; SRSS: Short Recovery and Stress Scale; MTDS: Multi-Component Training Distress Scale. RTT-Q: Readiness to train questionnaire.

Overall quality of the studies

Across all included studies, measurement properties were collectively assessed 120 times. Generally, the methodological quality was rated very good across the range of measurement properties. Structural validity and internal consistency were the more commonly assessed properties across the studies, however evaluations for measurement invariance, measurement error and responsiveness (comparison with other instruments) were lacking. Criterion validity was not assessed due to lack of a ‘gold standard’, according to the COSMIN definition and instead was assessed as construct validity. The quality of the AROM development is in Table 4.4. The quality of the evaluated measurement properties is in the Appendix F.

Table 4. 4: Quality of the AROM development.

AROM	AROM design							Cognitive interview (CI) study [2]				TOTAL PROM DEVELOPMENT
	General design requirements					Concept elicitation[1]	Total PROM design	General design requirements	Comprehensibility	Comprehensiveness	Total CI study	
	Clear construct	Clear origin of construct	Clear target population for which the PROM was developed	Clear context of use	PROM developed in sample representing the target population			CI study performed in sample representing the target population				
RESTQ 48 (Kallus, 1995)	V	V	V	V	I	I	I	I			I	I
RESTQ 77 (Kallus & Kellmann, 2016)	V	V	V	V	V	I	I	I			I	I
RESTQ 36 (Nicolas et al., 2019)	V	V	V	V	V	I	I	I			I	I
ARSS (Kellmann et al., 2016)	V	V	V	V	V	I	I	I	D	D	I	I
SRSS (Hitzschke et al., 2015)	V	V	V	V	V	I	I	I	D	D	I	I
POMS 65 (McNair, Lorr, & Droppleman, 1971)	V	V	V	V	I	I	I	I			I	I
POMS 63 (Andrade et al., 2010)	V	V	V	V	V	I	I	I			I	I
POMS 48 (Fernández, Fernández, & Pesqueira, 2002)	V	V	V	V	V	I	I	I			I	I
POMS 44 (Andrade, Arce, de Francisco, Torrado, & Garrido, 2013)	V	V	V	V	V	I	I	I			I	I
POMS 40 (Grove & Prapavessis, 1992)	V	V	V	V	V	I	I	I			I	I
POMS 36 (Viana, de Almeida, & Santos, 2001)	V	V	V	V	V	I	I	I			I	I
POMS 30 (Bourgeois, LeUnes, & Meyers, 2010)	V	V	V	V	V	I	I	I			I	I

BRUMS (Terry, Lane, Lane, & Keohane, 1999)	V	V	V	V	V	I	I	I	I	I
DALDA (Rushall, 1990)	V	V	V	V	V	I	I	I	I	I
DALDA (Nicholls, Backhouse, Polman, & McKenna, 2009)	V	V	V	V	V	I	I	I	I	I
ABQ 15 (Raedeke & Smith, 2001)	V	V	V	V	V	I	I	I	I	I
ABQ 18 (Arce, De Francisco, Andrade, Seoane, & Raedeke, 2012)	V	V	V	V	V	I	I	I	I	I
ABQ 12 (Isoard-Gautheur, Oger, Guillet, & Martin-Krumm, 2010)	V	V	V	V	V	I	I	I	I	I
OFSM (Legros, 1993)	I	D	I	D	V	I	I	I	I	I
MTDS (Main & Grove, 2009)	V	V	V	V	V	I	I	I	I	I
PSS (Cohen, Kamarck, & Mermelstein, 1983)	V	V	I	D	I	I	I	I	I	I
PRSS (Laurent et al., 2011)	V	V	I	I	V	I	I	I	I	I
TQR (Kenttä & Hassmén, 199.)	V	V	V	V	A	I	I	I	I	I
Fatigue (Chatard et al., 2003)	I	D	V	D	V	I	I	I	I	I
Fatigue (Atlaoui et al., 2004)	I	D	V	D	V	I	I	I	I	I

Measurement properties were rated on a four-point scale according to COSMIN: "V" = very good, "A" = adequate, "D" = doubtful and "I" = inadequate. RESTQ: Recovery-Stress Questionnaire Athletes-Sport; ARSS: Acute Recovery and Stress Scale; SRSS: Short Recovery and Stress Scale; POMS: Profile of Mood States; BRUMS: Brunel Mood Scale; DALDA: Daily Analysis of Life Demands for Athletes; ABQ: Athlete Burnout Questionnaire; OFSM: Overtraining questionnaire of the French Society for Sports Medicine; MTDS: Multi-Component Training Distress Scale; PSS: Perceived Stress Scale; PRSS: Perceived Recovery Stress Scale; TQR: Total Quality Recovery Scale.

Discussion

Athlete-reported outcome measures are commonly used to quantify the training response and inform decision making in sport. Plausible AROMs and measurement tools are vast and a detailed scrutiny of their key measurement properties according to established guidelines is thus far lacking. Therefore, the first objective of this systematic review was to identify the most commonly used AROMs for athlete monitoring in sport. We identified 310 studies that used AROMs in sport over the last 50 years, with the most common multiple item scales being the POMS, RESTQ-Sport and DALDA. For single item measures the most common were the ‘wellness items’, muscle soreness and recovery. Interestingly, we noted many modifications from the original versions. For example, RESTQ-Sport 77 was originally designed to have a single score for the various scales/constructs measured. However, many studies had combined scales into one overall score, which was not the author’s original intention. Furthermore, the Hooper/wellness items have been vastly modified across the years from the original single items proposed by Hooper et al (Hooper & Mackinnon, 1995; Hooper, Mackinnon, Howard, Gordon, & Bachmann, 1995) into a combined single score of ‘wellness’. Therefore, we categorised studies with the original intended version and by modifications. Overall, there has been a rise in the numbers of AROMs used, particularly in the last 5 years.

The second objective of this systematic review was the evaluation of the quality of measurement properties of identified AROMs according to the COSMIN standards. We identified 71 studies that reported measurement properties of 26 AROMs. The information obtained from the studies included, revealed that no AROM met all COSMIN criteria standards for measurement properties with evaluations on measurement invariance, reliability and measurement error (not examined at all) most lacking.

Additionally, no study reported complete assessment of all the measurement properties of the AROMs identified as per the COSMIN criteria.

For brevity, a summary with a brief explanation of each measurement property as examined according to the COSMIN, together with assessment criteria (Mokkink et al., 2010) is available in the Appendix G. We invite the readers who are not familiar with these concepts to consult the Online Supplementary Files before reading the following paragraphs. The individual discussion of the AROMs (RESTQ-Sport, ARSS, SRSS, POMS, DALDA, ABQ, PSS, TQR, OFSM, MTDS) is provided in the Appendix H. Below we provide a general collective comment on the results of these AROMs. We recommend readers who are interested in this topic to consult the Appendices.

Athlete Reported Outcome Measures

The overall results of AROMs (RESTQ-Sport, ARSS, SRSS and MTDS) specifically developed for athlete populations were as follows. The content validity across all AROMs was rated very low quality of evidence with an overall insufficient rating. Both structural validity and internal consistency were mostly high quality of evidence with overall positive rating. Cross cultural validity was generally high-quality evidence with positive overall rating. Reliability was rated very low quality of evidence with insufficient rating; however, we did not identify many studies examining this measurement property. Construct validity was moderate to high quality of evidence with mostly positive overall rating. Finally, responsiveness was moderate to high quality of evidence with positive overall rating.

Collectively, we summarise the results for the remaining AROMs: POMS, DALDA, ABQ, PSS, TQR and the OFSM, including any modified versions. The content validity across all AROMs was rated very low quality of evidence with an overall insufficient rating. Generally structural validity rated moderate to high quality of evidence with mixed of overall ratings (e.g., sufficient, indeterminate and insufficient). Internal consistency was high quality of evidence with positive overall rating. Very few studies have examined cross cultural validity with moderate quality of evidence and positive overall rating. Reliability ranged from very low to high quality of evidence with mostly insufficient rating. Construct validity was generally high quality of evidence with overall positive rating. Responsiveness ranged from very low to moderate quality of evidence with mostly indeterminate overall ratings.

Single items and the so called 'wellness' items

Traditional psychometric instruments usually contain multiple items that are exhibited (observable) in nature but are intended to measure a specific underlying, latent (unobservable) construct (Sloan et al., 2002). However, the use of multiple item instruments can cause participant boredom, fatigue and frustration because of item redundancy and repetition. Long instruments are also more time consuming, have a higher administrative burden and are not suitable for daily and frequent use (Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011; Sloan et al., 2002). For these reasons the use of single items is attractive as they are simple, may enable a high completion rate and are easier to implement (Sloan et al., 2002). However, the validity of single items is a matter of debate and psychometrists usually discourage their use because of some limitations. For example, internal consistency cannot be computed, are more vulnerable to random measurement errors and to unknown biases in meaning and interpretation (Hoeppner et al., 2011). Nevertheless, the use of single items may be acceptable under certain conditions. First, the construct of interest should be sufficiently unidimensional and unambiguous. Indeed, a single item necessitates that a respondent “consider all aspects of a phenomenon, ignore aspects that are not relevant to their situations, and differentially weigh the relevant aspects to provide a single rating” (de Boer et al., 2004). In this circumstance a single item may possess acceptable face validity. In addition, single items may reduce the chance of common method variance (state dependence) (Hoeppner et al., 2011).

From our systematic search the single items plus the Hooper/wellness items (from which most of the single items are developed) accounted for 46.1% of all instruments used in previous studies. These items are also frequently used in practical settings and are included in athlete management systems of national organisations, commercial software, etc. Surprisingly, other than two studies addressing reliability and responsiveness (low level of evidence) (Crowcroft et al., 2017; Fitzpatrick et al., 2019), we could not find any development and validation studies specifically addressing their measurement properties neither for each item used in isolation nor when combined in a single score (Abbott, Brownlee, Harper, Naughton, & Clifford, 2018; Chamari, Haddad, Wong, Dellal, & Chaouachi, 2012). In this review, we restricted the search and rating to studies purposely developed to examine the psychometric properties of these instruments and not studies, for example, where relations between these items and other variables were calculated (Hills & Rogerson, 2018; Thorpe et al., 2017). The lack of a conceptual framework specifying why specific associations were assumed is a serious limitation since it is prone to HARK-ing (hypothesised after results are known). This approach is acceptable in an explorative phase where associations can be used for hypothesis generation or model development; but if this is the aim, it should be declared and results not used as “evidence” of validity. Instead, the hypothesis generated should be subsequently used as a framework for developing, confirming and examining the associations hypothesised. Indeed, the COSMIN suggests using a hypothesis confirmation approach where the researchers develop specific hypotheses that are subsequently examined. However, even for measures used in the exploration, the validity of the items is a necessary condition to avoid developing theories or hypotheses based on spurious associations or lack of association due to invalid or flawed measures.

It seems the origin of the wellness items is based on the work of Hooper et al, (Hooper & Mackinnon, 1995; Hooper et al., 1995) for the monitoring of overtraining, in which initially four items (fatigue, sleep quality, muscle soreness and stress) were proposed for swimmers (Hooper et al., 1995) and subsequently an additional four items (enjoyment of training, irritability, health causes of stress and unhappiness) for athletes in general (Hooper & Mackinnon, 1995). However, the wellness items were only first presented in the methods section of a paper examining overtraining in swimmers and subsequently in a review paper. In addition, over the years there have been many modifications to these items even including the summation of a single score of the items known as the "Hooper index or Hooper score" of wellness. These adjustments have appeared without any attempt at validation and have presented arbitrary modifications in the methods section. To further confuse the issue, these 'wellness' items have been referred in different ways such as the Hooper questionnaire (Baklouti et al., 2017; Boukhris et al., 2019; Clemente, Martinho, Calvete, & Mendes, 2019; Clemente et al., 2019; Clemente et al., 2019; Clemente et al., 2019; Dergaa, Fessi, Chaabane, Souissi, & Hammouda, 2019; Matos et al., 2019; Mendes et al., 2018; Moalla et al., 2016), Hooper index or score (Chamari et al., 2012; Clemente et al., 2019; Clemente et al., 2019; Clemente et al., 2019; Goncalves et al., 2020; Matos et al., 2019; Mendes et al., 2018; Moalla et al., 2016; Oliveira et al., 2019; Rabbani, Baseri, Reisi, Clemente, & Kargarfard, 2018; Rabbani, Clemente, Kargarfard, & Chamari, 2019; Romdhani et al., 2019; Selmi et al., 2019), Hooper-Mackinnon questionnaire (Gathercole, Sporer, & Stellingwerff, 2015), subjective or self-reported wellness ratings (Gathercole et al., 2015; Govus, Coutts, Duffield, Murray, & Fullagar, 2018; Hills & Rogerson, 2018), perceived ratings of wellness (Buchheit et al., 2013; Thorpe et al., 2016), psychometric measures of wellness (Buchheit, Cholley, & Lambert, 2016), perception of well-being (Matos et al., 2019; Noon, James, Clarke, Akubat, & Thake, 2015), subjective/self-reported well-being measures and questionnaire (Abbott et al., 2018; Buchheit et al., 2013; Cullen, Thomas, Webb, Phillips, & Hughes, 2017; Doeven, Brink, Huijgen, de Jong, & Lemmink, 2019; Lacome, Carling, Hager, Dine, & Piscione, 2018; Marrier et al., 2018; McGuinness et al., 2018; McLean et al., 2010; Oliver, Lloyd, & Whitney, 2015; Ryan et al., 2018; Tibana et al., 2019), daily well-being (Sawczuk, Jones, Scantlebury, & Till, 2018), wellbeing indices (Selmi et al., 2019), daily subjective questions (Crowcroft et al., 2017), subjective perceptions (or perceptual) of well-being (Fowler, Duffield, & Vaile, 2014; Marrier et al., 2018), training diary (Decroix, Lamberts, & Meeusen, 2018; Sargent, Lastella, Halson, & Roach, 2014), training log (Hooper et al., 1995). The lack of consistency in a definition of the construct that supposedly measures these items indicates the absence of a theoretical framework, thereby creating a lack of clarity regarding what are the domains of interest (and also why that domain was useful within the training process). Assuming each item measures one sub dimension of a higher order construct, it is however not clear what is the reference higher order construct. Indeed, all the definitions used in the literature actually reflect different constructs (wellness, wellbeing, welfare, etc.). Within research, several single items have been used to measure various dimensions of this undefined domain: general health, fatigue, energy level, recovery, happiness, desire to train, stress, enjoyment, mood and irritability etc (Abbott et al., 2018; Crowcroft et al., 2017; Gathercole et al., 2015;

Hooper & Mackinnon, 1995; Robey et al., 2014). There is no research explaining why these single item measures should reflect subdomains of a variety of constructs, and it is unreasonable that the same subdomains belong to so many different constructs. Therefore, it appears researchers just use these definitions to categorise the items, but these definitions are conceptually empty.

This also raises concern regarding the combination of the items in a single score, that would assume that the items reflect the same construct (reflective model) or form the same construct (formative model). Again, the COSMIN standards indicate items should be "consistent with the theory, conceptual framework or disease model that was used to define the construct of interest". It also means that the relevance of the items within the training process should also be presented in such a way that specific hypotheses (e.g., directions and strength of the associations) can be generated and tested. To understand the complexity of defining a construct like well-being and how it should be addressed, we refer the reader to a recent narrative review specifically addressing this issue from a sport psychology (psychometric) perspective (Giles, Fletcher, Arnold, Ashfield, & Harrison, 2020). By reading this review, it seems that the single items used to monitor training responses have little to do with "well-being". Overall, it is highly concerning that the most commonly used items in both research and practice are not validated and some lack of underlying conceptual validity (single items used to measure multidimensional constructs). Even the choice of the response options has never been justified. It appears that they have been liberally changed from one study to another, while the selection of the response options should also be based on conceptual and computational considerations. Furthermore, being single items, the nature of the response scale should be also considered, or acknowledged, when analysing the data (e.g., ordinal versus interval, i.e., parametric vs nonparametric analysis).

Reliability was purposely examined in only two studies, one of which also assessed responsiveness (Crowcroft et al., 2017; Fitzpatrick et al., 2019). Other studies just reported the reliability in the methods section without providing any details that allow evaluation of the appropriateness of the methods used. Further, the results are not encouraging (SEM as coefficient of variation from 5% to 31.9% and ICC of 0.58) (Gastin, Meyer, & Robinson, 2013; Hills & Rogerson, 2018; Malone et al., 2018; Thorpe et al., 2017). Based on our systematic search there is no evidence supporting the use of the items in both research and practice. Given the interest regarding single items, proper instruments should be developed by following established guidelines and psychometric, clinimetric methods. While it may be suggested (by practitioners) that these items are useful in a practical setting (self-awareness of the athletes, 'warning' signals, etc.), these arguments are worth considering but move the discussion from a scientific to an anecdotal point of view if not properly investigated. Instead, these opinions can be used to develop precise hypotheses and reference frameworks for scientifically examining whether these 'benefits' really occur in practical setting.

Critical issues and limitations

In carrying out this review, the use of the COSMIN checklists on the selected studies proved challenging at times. It was evident that many studies on AROMs, particularly those studies published more than 15 years previously, were not reported in sufficient detail. The COSMIN initiate was established in 2005, with the original COSMIN checklist developed in 2006/2007. On this basis, studies published post this date potentially should have had superior methodological quality, together with higher level of evidence. Although the majority of studies (80%) included in our Part Two review were published post 2007, it is clear the clinimetrics guidelines did not constitute a reference. Further, we are aware that the COSMIN is not well-known in sport science. Nevertheless, the COSMIN is based on established and well- known methodology that should be used by those investigating this area.

In this review we used a clinimetric approach, which is the methodological discipline focusing on the quality of measurements in clinical and medical practice. Although clinimetrics uses psychometric methods, most of these instruments have been developed within the psychological area and therefore standards may differ. However, AROMs are commonly used for assessing the responses to training and competition, mainly symptoms and physical signs which are the targets of clinimetric instruments. For this reason, we rely on clinimetric terminology and methodology. That is, the instruments were examined more than in their role of exploring subjective responses associated to the training process, as attributes (perceived symptoms) influenced by training (de Vet, Terwee, & Bouter, 2003). More psychological-oriented interpretations of AROMs, or any other psychological instruments, are not a matter of expertise for coaching and sport science staff to whom this review is targeted. This necessitates the involvement of experts (i.e., psychologists) for a correct interpretation and implementation. Our systematic review should be interpreted from an applied perspective that is, instruments which monitor perceptual responses to optimise physical training.

A limitation of the current review is the modification of the COSMIN search strategy and filters. When the original was applied, we only obtained two results. We believe this issue may be in part due to the fact that many studies do not adequately report the essential information (e.g., instrument name, inclusion of instrument or constructs, psychometric property) within their abstract. Therefore, we developed our own filters and search strategy and performed a manual search of the literature. As such, we may have potentially missed the inclusion of some studies due to this issue. Finally, the application of the COSMIN necessitates several subjective decisions as for any risk of bias tool. However, considering these guidelines are not standard in sports science, we were more liberal in the application of the criteria than in previous studies (Impellizzeri et al., 2020). Therefore, given some subjectivity of the evaluation, other reviewers may rate the selected instruments slightly differently (with clinimetricians potentially taking a stricter approach), but we believe it is unlikely the findings would be substantially different.

Conclusion and practical recommendations

Monitoring training responses with subjective measures is common and many instruments are available. Our systematic review provides to date, a unique and comprehensive appraisal of the risk of bias and quality of the psychometric (measurement) properties of the most frequently used AROMs according to established guidelines. Categorising the results of the current review in terms of “traditional” psychometric instruments and sport science generated questionnaires (e.g., single items), we can conclude that while some of the most commonly used psychometric instruments do not show adequate quality in all measurement properties, most psychometric attributes have been at least examined. Among these AROMs, a few have been developed specifically for athletic populations such as MTDS, RESTQ-Sport, ARSS and SRSS with some showing good evidence, at least in some measurement properties (e.g., ARSS and SRSS). The main area requiring improvement is certainly the content validity, in which all studies were rated inadequate. The finding of the current review provides suggestions of properties that require further assessment and exploration.

Conversely, the AROMs that are most commonly used in sport science have not been validated, despite often presented as validated. Of further concern, is the finding that these

AROMs (e.g., the so-called wellness items) have often been modified from the original version (e.g., adding/removal of items or combining scores) without justification. Any modification should go through a validation process, but in sport science it seems common to use 'not validated modified versions of the originally non-validated versions'. Until proper validation studies are carried out, conclusions based on these AROMs remain questionable and difficult to interpret. The paucity of methodologically adequate validation or development studies, especially for single items proposed within the sport science area, is likely related to the lack of specific contextual and background knowledge on psychometry and clinimetrics. We commend previous authors for having raised interest in the area of subjective symptoms and perceptions in training monitoring. However, there is an urgency to adopt proper methodology considering that reference methods (including guidelines) already exist, and that this area (subjective measures) has a long history (e.g., clinimetrics).

Acknowledgments

The authors would like to thank Dr Michael Kellmann for additional information provided for this systematic review.

Chapter Five

Development of a revised conceptual framework of physical training for measurement validation and other applications

Jeffries, A.C., Marcora, S., Coutts, A. J., Wallace, L., McCall, A., & Impellizzeri, F. M. (*Under revision, minor modifications requested*). Development of a revised conceptual framework of physical training for measurement validation and other applications. *Sports Medicine*.

Preface

Given the lack of studies showing the validity of the single items commonly used in research (*Study Two*), *Study Four* was planned to select and examine the validity of items that can be useful to monitor training effects in dancers. However, the validation process necessitates a framework and for this reason *Study Three* was conducted. *Studies Two* and *Three* were developed encompassing athletes in general, not only dancers. A conceptual framework may be broader and not necessarily discipline/sport specific (unlike the validation process), as long as it is adequate (generalizable) to the requirements and training process of each disciplines. From an examination of the literature of relevant conceptual or theoretical frameworks developed specifically for the training process, we identified several concepts that could be further modified or extended that are important in the conceptualization of the process and outcomes of physical training.

Abstract

A conceptual framework has a central role in the scientific process. Its purpose is to synthesize evidence, assist in understanding phenomena, inform future research and act as a reference operational guide in practical settings. We propose an updated conceptual framework intended to facilitate the validation and interpretation of physical training measures. This revised conceptual framework was constructed through a process of qualitative analysis involving a synthesis of the literature, analysis and integration with existing frameworks (Banister and PerPot models). We identified, expanded and integrated four concepts that are important in the conceptualization of the process and outcomes of physical training. These were: 1) formal introduction of a new measurable component 'training effects', a higher order construct resulting from the combined effect of four possible responses (acute and chronic, positive and negative); 2) explanation, clarification and examples of training effect measures such as functional (performance-based), physiological, subjective and other measures (cognitive, biomechanical, etc.); 3) integration of the sport performance outcome continuum (from performance improvements to overtraining); 4) extension and definition of the network of linkages (uni and bidirectional) between individual and contextual factors and other constructs. Additionally, we provided examples of theoretical and practical applications of the conceptual framework such as validation and conceptualization of constructs (e.g., performance readiness), and understanding of higher order constructs, such as training tolerance when monitoring training to adapt it to individual responses and effects. This conceptual framework provides an overarching model that may help understand and guide the development, validation, implementation, and interpretation of measures used for athlete monitoring.

Keywords: Training response, contextual factors, performance, subjective measures, acute, chronic, athlete monitoring.

Aims of the conceptual framework

What is a conceptual framework?

A conceptual framework synthesizes evidence, assists in understanding the phenomena under investigation, informs future research and acts as a reference guide in practical settings. In the scientific process, conceptual frameworks allow more precise specification of hypotheses, thus increasing their degree of falsifiability, including when auxiliary or primary hypotheses of the main theory are modified if predictions fail (Maxwell, 2005; Victora, Huttly, Fuchs, & Olinto, 1997). This prevents ad hoc modifications of the theories, i.e., bias. A conceptual framework may aid in the explanation of particular phenomena providing the background knowledge, based on which the potential causal structure is formally presented as directed acyclic graphs for statistical examination (counterfactual framework) (Greenland & Brumback, 2002; Hernán, Hernández-Díaz, & Robins, 2004). This also allows the identification of confounders, moderators, mediators, colliders etc. (Cole et al., 2009; VanderWeele & Robins, 2007). The second purpose of a conceptual framework is to guide the validation of measurements (Impellizzeri & Marcora, 2009). Lack of conceptualization and articulation of causal phenomena weakens research in a number of ways including poor methodological arguments, lack of conceptual clarity and undermining of the foundation of solid empirical research (Gimeno-Santos et al., 2011).

In the context of patient-reported outcome measures, a conceptual framework is a representation of the relationships between the construct to be measured (e.g. using reflective or formative models) (de Vet, Terwee, Mokkink, & Knol, 2011). This explanation is consistent with a generic definition, which describes a framework as the structure which the researcher believes to best explain the phenomenon to be studied (Camp, 2001). It may take the form of a visual representation, arranged in a logical structure, illustrating presumed relationships between key concepts, constructs or variables. The availability of a reference conceptual framework, together with operational definitions of the constructs, is therefore essential for the measurement validation process (Rothman et al., 2007). This provides a basis for researchers to propose exactly what and why specific variables (objective or subjective) are measured and act as a reference for validation studies. Of particular importance is understanding "why" since a measure can be valid for a specific context and goal but not in another.

Why develop a conceptual framework of physical training?

In the last decade, there has been an increased interest in subjective measures within sports science, with new constructs (e.g. wellness, wellbeing, performance readiness etc.) introduced but never clarified (Jeffries et al., 2020). Moreover, while the use and development of reference theoretical and conceptual frameworks is common in disciplines such as psychometrics and clinimetrics (Rothman et al., 2007; Wilson & Cleary, 1995), they are not commonly used when validating measurements used in sport science. Unfortunately, the absence of a reference framework presenting hypothesized relations between these constructs and other measures (e.g. training load), makes the examination of the validity and usefulness of these items introduced in the sport science literature confusing and prone to bias (e.g. post hoc "theories" explaining relations, that are not subsequently tested and that decrease the degree of falsifiability) (Smoliga & Zavorsky, 2017). For example, some common subjective measures used in research and practice such as "wellness items" have never been validated following appropriate and established methods such as those presented by the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) (Jeffries et al., 2020). This may also be due to the fact that available frameworks (Impellizzeri et al., 2019; Impellizzeri et al., 2005) do not include all components necessary to provide a suitable reference for these kinds of validation studies. For this reason, we re-examined available frameworks and proposed a conceptual framework that can facilitate the validation of physical training measures. The current conceptual framework refers to physical training, which is the training involving physical activities with the goal of improving sport performance. Physical training encompasses both process and outcomes. Training process is the systematic repetition of physical exercises involving external and internal load (Viru & Viru, 2000), and its outcome may include physiological, biochemical, anatomical, and functional changes i.e. training effects and sport performance outcomes (Impellizzeri et al., 2005). The conceptual framework excludes other kinds of training such as tactical training, watching videos and pure psychological skills training that can improve sport performance but do not require physical exertion.

Development process of the conceptual framework

The development of our conceptual framework (Figure 1) was constructed through a process of qualitative analysis. This was an integrative and evolving progression, consisting of a synthesis of the literature and analysis of existing frameworks, coupled with an assimilation of information and concepts from previously developed models and theories. We developed the conceptual framework for the training process using conceptualization, a process in which imprecise constructs and their constituent components are defined in precise terms (Bhattacherjee, 2012). The version presented in the current paper is the final result of more than 15 versions and elaborations, that were conceptually tested using a red team approach (Lakens, 2020), where the authors challenged conceptually the framework by proposing (worst case) scenarios to test whether it could reasonably fit common measures and training strategies. By conceptualizing and adding other constructs (individual and contextual factors, and training process), we also defined their relationships within the conceptual framework. Below we have documented the main stages in the development process in which we explicitly justify and elucidate decisions about key elements of the conceptual framework.

The previous version of the physical training framework

We decided to re-evaluate and refine a previous physical training framework (Impellizzeri et al., 2019; Impellizzeri et al., 2005) integrating or modifying, together with the examination of other frameworks or models available in the literature. This original framework was developed to define and conceptualize the essential measurable components of the training process and its outcomes. Briefly, a fundamental notion of this physical training framework was that when delivered appropriately, and according to the training goal and plan, exercise will induce psychophysiological responses leading to adaptations. These responses *during* the exercise are the stimuli initiating the adaptations determining the training outcomes. The training load construct was differentiated into two components, external and internal training load, by redefining and elaborating two terms (external and internal) that were already present, but not operationally conceptualized, in some non-peer-reviewed coaching literature (Sassi, 1997) and seminal textbooks on training periodization (Matveyev, 1977). In this previous framework, the external training load, and its interaction with various individual and contextual factors, determines the internal training load, which ultimately will produce changes in the training outcome. For the operational definitions see Table 1 and for more details we refer the reader to a recent publication (Impellizzeri et al., 2019). This previous theoretical framework has since been also adapted to other fields such as physical activity for health or other contexts such as biomechanics (Herold, Torpel, Hamacher, Budde, & Gronwald, 2020; Vanrenterghem, Nedergaard, Robinson, & Drust, 2017).

Identification of other frameworks

A first attempt to systematically search the literature failed because, in sport science, frameworks for physical training are not presented as such and/or they are not clearly indicated in titles and abstracts. We therefore decided to locate suitable frameworks based on the expert knowledge of the authors, scoping searches and publication references. This may have introduced bias, although this is inevitable when developing conceptual frameworks given the qualitative nature of the process. From the examination of the literature, we identified and selected two other potentially relevant models specifically proposed for the training process: the Banister Impulse-Response (IR) (Calvert et al., 1976; Morton et al., 1990) and the Performance Potential Metamodel (PerPot) (Perl, 2001, 2004). The model of Banister considered athletic performance to be measurable as the net outcome of two key training responses, also called fitness and fatigue, from the application of a training impulse. The original works by Banister et al. (Calvert et al., 1976; Morton et al., 1990) applied a systems theory to evaluate the response to physical training using a mathematical function. In this mathematical model, performance is the output, with the athlete regarded as the system, and the training impulse as the input (Calvert et al., 1976; Morton et al., 1990). The functional relationship between the training impulse and the system response is expressed by two differential first order equations attributed to the antagonist effects that were called fitness and fatigue (Calvert et al., 1976; Morton et al., 1990). The PerPot metamodel (Perl, 2001, 2004), simulates the interaction between load and performance in adaptive physiological processes in sport, by means of antagonist dynamics (response flow and strain flow). The components of the

PerPot (Perl, 2001, 2004) have some similar theoretical characteristics to Banister et al, (Calvert et al., 1976; Morton et al., 1990) such as the concept of antagonist role of strain and response potential, whilst using a different computational approach. Finally, we also referred to the joint consensus statement of the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM) on overtraining (Meeusen et al., 2013). Although it was not formally presented as such, it can be seen as a theoretical framework presenting the continuum of the effects of training and individual/contextual factors on sport performance.

Integration with other frameworks

While these two models (Calvert et al., 1976; Morton et al., 1990; Perl, 2001, 2004) included, albeit with different names, a negative and positive component, the theoretical framework of physical training by Impellizzeri et al. (Impellizzeri et al., 2019; Impellizzeri et al., 2005) did not include these two elements. Therefore, we initially attempted to combine the two components (fitness and fatigue) of the IR-Banister model into a new framework. However, there was a conceptual problem related to the definitions, or lack thereof, for the terms ‘fitness’ and ‘fatigue’. Originally, these two terms were arbitrarily used to define two mathematical components reflecting the positive and negative training responses. They were computationally and not conceptually defined since they just represented an assigned name for two “model components of performance ability” and not two theoretical constructs. However, the interpretation of these two terms may result in confusing overlapping with the more generic definitions reported in the literature, making it challenging to achieve a clear understanding of these constructs for conceptualizing into a general framework. For example, the ACSM defines fitness as “a set of attributes that people have or achieve that relates to the ability to perform physical activity" (Wilder et al., 2006). Fitness has been also operationalized as ''[a set of] measurable health and skill-related attributes'' that include cardio-respiratory fitness, muscular strength and endurance, body composition, flexibility, balance, agility, reaction time and power (American College of Sports Medicine, 2002). These definitions refer, however, to a set of attributes which are outcome measures (output) and do not reflect the fitness component of the IR Banister model (input). Similarly, this problem also relates to the many definitions of fatigue (Coutts et al., 2017; Edwards, 1983; Lewis & Wessely, 1992), none reflecting the second input variable of the Banister model and even the strain of the PerPot. Indeed, fatigue is a complex construct that can be interpreted in terms of an acute reduction in muscle force/power, higher perception of effort, task failure, a mood dimension or as a symptom of disease depending on the context (Enoka & Duchateau, 2016; Morgan, Brown, Raglin, O'Connor, & Ellickson, 1987). Whilst some kinds of fatigue can be considered negative effects of training, the term fatigue cannot be used as a synonym of negative effects because fatigue is not the only negative effect of training. For example, training can induce muscle damage and delayed-onset muscle soreness (DOMS) (Peake, Neubauer, Della Gatta, & Nosaka, 2017) and other negative alterations in mood such as an increase in depression (Morgan et al., 1987). Therefore, whilst we considered it worth adding these two competing components, we decided not to use the terms fitness and fatigue due to the above issues. We then replaced "fitness" and "fatigue" in Banister's model, with positive and negative training effects. This was still consistent with the original meaning of these two components that were defined as positive or negative based on their direct effects on performance (Clarke & Skiba, 2013; Morton et al., 1990). Additionally, it was also compatible with the PerPot model where Perl described the antagonist dynamics as: "*the same load input has two contradictory effects, namely the performance increasing response flow* [our positive training effects] *and the* *performance decreasing strain flow* [our negative training effects]" (Perl, 2004). However, we did not want to use the term strain and response, as in the PerPot model, since these terms are loaded and may create confusion in the physical training context. We also specify that the PerPot model can be applied to various "performances" and not necessarily competitive performance. In the PerPot model, performance is a generic term that can include biological and physiological outcomes (e.g., hemoglobin and heart rate) (Endler, Hoffmann, Sterzing, Simon, & Pfeiffer, 2017; Perl, 2001, 2004).

In a further evolution of our conceptual framework, we expanded the training effect construct with the additional acute and chronic dimensions. Below, each section of the conceptual framework is described in further detail. Finally, we integrated the ECSS/ACSM (Meeusen et al., 2013) overtraining framework to define the possible sport performance outcomes resulting from the balance between positive and negative training effects.

New and expanded concepts

Above we identified the lack of the two positive and negative training effects in the previous framework (Impellizzeri et al., 2019; Impellizzeri et al., 2005). We therefore adapted the previous framework to overcome this limitation and to integrate, elaborate and clarify concepts of existing frameworks. Specifically, we identified four concepts that could be integrated or extended. An iterative process was utilized to determine the final version of the conceptual framework.

The first concept we expand and integrate is the *training effect*, which in the previous framework (Impellizzeri et al., 2019; Impellizzeri et al., 2005) was mixed with the generic concepts of 'adaptations' and of training outcome. However, it did not constitute a formal and defined construct of the framework. We acknowledge that the lack of the training effect construct may have contributed to confusion, where measures of effects occurring *after* the training session have been misinterpreted as measures of internal training load (Halson, 2014), which is by definition the psychobiological response *during* the exercise(s) constituting the training session (Impellizzeri et al., 2019).

The second aspect we more formally clarified is training effect *measures*. The training effect is a construct for which there is no gold standard measure however it can be assessed in several ways, including using proxy measures when the training effect being considered cannot be directly quantified. This is an important clarification since the validation process may also refer to proxy measures, e.g., whether measures are adequate reflections of the construct of interest.

The third area we further articulate is sport performance outcomes. In the previous frameworks of (Impellizzeri et al., 2005) and (Impellizzeri et al., 2019) used the term 'training outcomes' which does not properly differentiate between the chronic training effect and sport performance outcomes. In the current conceptual framework, these two constructs are differentiated and sport performance outcomes further elaborated by including the potential range, i.e. from improvement to overtraining.

The fourth area we expand on is the concept of individual and contextual factors (e.g., training status, health, nutrition and environmental factors). In the previous frameworks, (Impellizzeri et al., 2019; Impellizzeri et al., 2005) contextual factors were only mentioned as aspects influencing internal load and thereby training outcomes. Here, we clarify that modifiable and non-modifiable individual/contextual factors have a more widely integrated relationship with all components of our conceptual framework, including bidirectionality with training effects. These four areas are further discussed in the next section.

Conceptual Framework

This conceptual framework for physical training is intended to illustrate the relationship between stimulus (internal training load), training effects and their measures and sport performance. Similar to the previous versions, we wanted to develop a parsimonious conceptual framework presenting the essential measurable components such that it may be elaborated and expanded for specific applications and needs. The internal and external training load constructs have been presented in earlier sections and in previous publications (Impellizzeri et al., 2019; Impellizzeri et al., 2005) and although essential parts of the current conceptual framework, will not be further discussed here. For the measurement of external and internal load we direct readers to the following reference (McLaren et al., 2018). We now provide details of the additional or modified constructs: training effects, sport performance outcomes, and individual/contextual factors. The conceptual framework (constructs and their relationships) is visually depicted in Figure 5.1. Operational definitions of each construct of the conceptual framework are presented in Table 5.1.

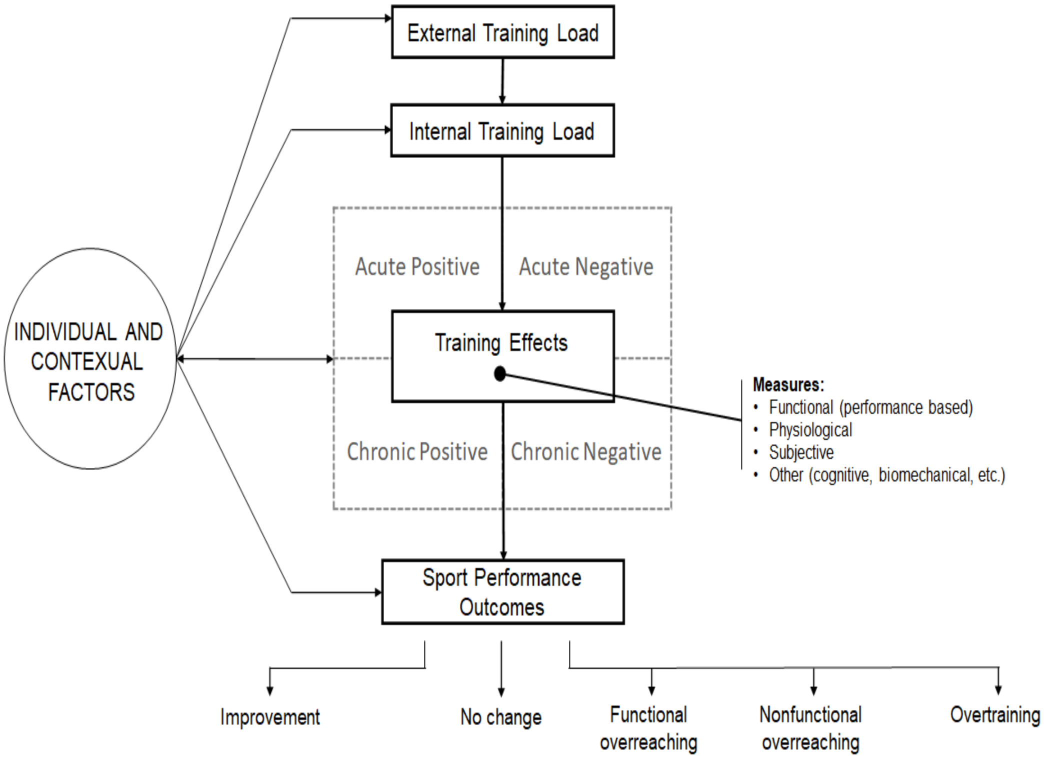


Figure 5. 1: Conceptual framework of physical training

Table 5. 1: Operational definitions

Construct	Operational definition
Training load	*Training load* refers to a construct reflecting the input variable that is manipulated to elicit a desired training response in athletes. Within this context, *load* is a generic term which is qualified by the term *training* in a fashion similar to other areas of research that have adopted the term *load* within a variety of contexts (i.e., allostatic load, cognitive load, mechanical load etc.). Accordingly, *training load* does not specifically refer to the forces experienced, as is typical in physics, or any other physical quantity. Training load, as a generic construct, accommodates a variety of measures of various nature (spatio-temporal, mechanical, psycho-physiological etc.) which can be described as being external or internal depending on whether the measurable aspect in question is internal or external to the athlete.
External training load	In the context of training load, the term *external load*, implicitly refers to the *external training load* undertaken by an athlete. External load has been defined as the physical work prescribed in the training plan. Notably, this does not refer to 'work' in the physics sense (force x distance) but more so in a generic manner. Accordingly, the term e*xternal load* accommodates quantification and prescription in a variety of manners, enabling the use of a diverse range of measures and metrics. Some common measures of external load include GPS derived units (speed, accelerations, etc.) and level of resistance.
Internal training load	As per external load, the term *internal load* implicitly refers to internal *training* loads, where the term training is the qualifier of load. In the context of training load, internal load typically refers to the psycho-physiological stress experienced by an athlete *during* the exercise or the training session. Therefore, the concept of internal load incorporates all the psycho-physiological responses initiated to cope with the requirements elicited by the external load, i.e., the execution of the exercise (single or sequence) according to the modalities prescribed by the coach. Despite internal loads typically being psycho-physiological in nature (essentially due to relative ease of application and quantification), the stress and strain experienced by specific tissues in response to an applied force is also internal and should therefore also fall within this category.
Training effects	Effects caused and occurring after the training session, that can be assessed using functional, physiological, subjective, biomechanical and other measures.
Acute effects	Effects induced by one training session to one week of training, that requires brief time (e.g., less than one week) to occur and to return to baseline once training is stopped.
Chronic effects	Effects induced by more than one week of training, that requires longer time (e.g., more than one week) to occur and to return to baseline once training is stopped.
Positive effects	Acute or chronic response that directly improves the sport performance outcome.
Negative effects	Acute or chronic response that directly impairs the sport performance outcome.
Sport Performance outcomes	The outcome as a result of the balance between the positive and/or negative training effects.
Contextual factors	All the factors not part of the main process (physical training) such as environmental, social, cultural factors, etc that can influence the training process or the training outcome. These factors have an integrated relationship with all components of the conceptual framework, including bidirectionality with training effects.
Individual factors	Characteristics of the individual athlete such as genetics, psychological traits and states, training background, etc. that can influence the training process or the training outcome These factors have an integrated relationship with all components of the conceptual framework, including bidirectionality with training effects.

Training effects

The main new construct added to this conceptual framework is the construct "training effect", which is the combination of four possible effects denoted by four quadrants (Figure 5.1). These four quadrants represent the possible acute/chronic and positive/negative effects that can influence sport performance and interact with each other, i.e., all permutations. We decided to operationally define acute and chronic training effects according to two related criteria: the amount of training required to induce a measurable effect and the time needed to return to baseline once training is stopped. Clearly, it is not possible to dichotomise the criteria since they should be seen as a continuum. However, from an operational point of view, we suggest that a training effect occurs after a single training session or few training sessions (up to one week of training) to be classified as acute, whilst the training effects occurring after more than one week of training as chronic. A similar time frame can be used with regards to return to baseline: training effects requiring less than one week to disappear should be classified as acute, whilst those lasting more than one week as chronic. Importantly, these are approximate criteria, and acute and chronic should be seen more as a continuum than a strict dichotomy.

The training effect can be both positive or negative and, as indicated by the conceptual framework, can occur concurrently. In other words, a positive effect (increased muscle protein synthesis after resistance training) (Coffey & Hawley, 2007) may be accompanied by a negative effect (neuromuscular fatigue) (Thomas et al., 2018). With the terms positive and negative, we indicate an acute or chronic effect that directly improves (positive) or directly impairs (negative) the sport performance outcome. This definition is consistent with the original Banister (and PerPot) model, where effects referred directly to performance (Calvert et al., 1976; Morton et al., 1990; Perl, 2001, 2004). More specifically, however, our use of the terms negative and positive qualifies the direct effects but does not necessarily imply that their indirect effects are also positive or negative. For example, glycogen depletion has a *direct* negative effect on endurance performance (Knuiman, Hopman, & Mensink, 2015). However, training in a glycogen depleted state can enhance activation of key cell signaling kinase transcription factors and transcriptional co-activators, resulting in a coordinated up-regulation of nuclear and mitochondrial genomes (Baar & McGee, 2008; Bartlett, Hawley, & Morton, 2015) (*indirect* positive effect). Therefore, whilst the *direct* effect of glycogen depletion is negative, its *indirect* effect via enhanced mitochondrial biogenesis can have a positive effect on endurance performance (Baar & McGee, 2008; Bartlett et al., 2015).

As mentioned previously, we have introduced acute and negative training effects to conceptually integrate the two competing components from previous models (fatigue and fitness in the IR Banister model and strain and response flow in the PerPot model) (Calvert et al., 1976; Morton et al., 1990; Perl, 2001, 2004). However, a fundamental difference with these models is that while we integrated the concepts of two competing effects, we did not identify these effects with the input measures as in their models, where inputs are measures of external load. Indeed, we provided a conceptual framework, while the aforementioned models aimed to predict performance outcomes using computational methods. Nevertheless, the current conceptual framework is coherent with these two models since they tried to predict performance outcomes by using external training load measures as proxies (surrogates) to quantify the negative and positive effects, with the balance between these two competing effects determining the performance. The use of these proxy measures is conceptually (and tautologically) supported by the causal relation between *training* effects and *training* load. The two models differed in the way the negative and positive effects were computationally determined, but they both fit into the current conceptual framework, further supporting the plausibility of our proposal.

Measures of training effects

The components included in the training framework are constructs for which, commonly, there are no gold standards, but there are various measures, including surrogate (proxy) measures reflecting the construct. New candidate measures can be proposed but they necessitate proper validation before implementing in practice and research, and the current framework provides the conceptual model to develop testable hypotheses to support their validity. Here we present and discuss some examples of training effect measures that we have classified, for simplicity, in four main categories as presented in Figure 5.1: 1) functional measures (performance-based), 2) physiological measures; 3) subjective measures, 4) other measures (cognitive, biomechanical, etc.).

Functional measures

Functional measures are measures of performance during a task that is either related to competitive performance or that is thought to measure a specific fitness component (e.g., strength). An example of such measures commonly used in sport is the countermovement jump (CMJ) test. As many other measures of training effects, the CMJ test can be used in two ways. Firstly, to quantify neuromuscular fatigue (Ade, Drust, Morgan, & Bradley, 2020; Fitzpatrick, Akenhead, Russell, Hicks, & Hayes, 2019) which is an acute negative effect of training. In this case, the CMJ test is performed after the training session and a reduction compared to the values in rested conditions is a measure of neuromuscular fatigue (Clarke, Farthing, Lanovaz, & Krentz, 2015; Sparkes et al., 2020). The second use of the CMJ test is to measure the chronic effect of training on lower body power (proxy measure) or jumping performance. For this purpose, conditions need to be standardized to limit as much as possible the influence of acute training effects and contextual/individual factors. For example, by reducing the training load and allowing recovery from neuromuscular fatigue and by controlling time of testing and sleep, nutrition, motivation before and/or during test (Tanner, Gore, & Australian Institute of Sport, 2013). Standardization has the dual roles of increasing reliability (de Vet et al., 2011) and providing a measure reflective of a chronic training effect (Tanner et al., 2013). Other examples of functional measures that are used to assess acute or chronic training effects are agility tests, sprints and multistage fitness tests (Behm et al., 2017; Fanchini et al., 2015; Moran et al., 2020; Speirs, Bennett, Finn, & Turner, 2016).

Physiological measures

There are various physiological measures that may be suitable for monitoring acute or chronic training effects. The better validated and most widely used are the physiological measures used to assess the positive chronic training effects of endurance training. These measures include maximal oxygen consumption, running economy and lactate threshold (Jones & Carter, 2000; Joyner & Coyle, 2008; Tolfrey, Hansen, Dutton, McKee, & Jones, 2009). Although critical power or critical velocity are often considered measures of a physiological construct (Jones & Carter, 2000), we have classified as functional measures as they are based on performance during time to exhaustion tests or time trials (i.e. functional measures representing the observable outcomes of a physiological construct).

Another variable often monitored in athletes is creatine kinase (CK) because the elevated CK levels that may occur after eccentric and/or unaccustomed exercise are used as a proxy measure for muscle damage which is an acute negative training effect (Baird, Graham, Baker, & Bickerstaff, 2012; Berriel et al., 2020; Marin, Bolin, Campoio, Guerra, & Otton, 2013). Another example of physiological measures are the assessment of central and peripheral fatigue using electrical or magnetic stimulation. However, contrary to the CMJ test mentioned earlier, these tests are less frequently utilized in practical settings due to the expertise required and need for specialized equipment (Coutts & Cormack, 2014; Tofari, Kemp, & Cormack, 2020).

A physiological measure that can be used to measure both acute or chronic training effects is heart rate variability (HRV). Indeed, HRV can indicate acute post-exercise perturbation and recovery of the cardiac autonomic system (acute training effect). Additionally, HRV can signify chronic adaptations to training by means of vagally mediated indices (chronic training effect) (Plews, Laursen, Stanley, Kilding, & Buchheit, 2013; Stanley, Peake, & Buchheit, 2013).

Subjective measures

Subjective measures are commonly used for monitoring the training process. However, whilst the use of subjective measures is considered sound (Saw et al., 2016), this area is characterized by the widespread use in sport research and practice of improperly validated instruments (Jeffries et al., 2020). Among the most utilized in research and practice (also implemented in athlete management systems and commercial software), are the so-called "wellness" items (Jeffries et al., 2020). However, it is difficult to understand why the term "wellness" is used, given the lack of clarity surrounding this construct, and why it would be relevant in the training process. For example, is it intended to measure training effects (acute or chronic) or individual/contextual factors? The wellness items, depending on the versions, are supposed to measure fatigue, sleep quality, stress, muscle soreness, enjoyment of training, irritability, health causes of stress and unhappiness (Hooper & Mackinnon, 1995; Hooper et al., 1995). Even ignoring that these items have not been properly validated, it is not clear whether these items are supposed to measure acute effects, chronic or individual and contextual factors; that can be influenced by the acute training effects. Nevertheless, among the items, fatigue and muscle soreness are good candidates as measures of acute training effects. Constructs such as stress, irritability and unhappiness appear more as individual/contextual factors or chronic effects. However, this is speculative as no framework has been previously provided or used as reference. Therefore, we are using a "reverse engineering" approach, starting with the measures used and going back to understand what components of the training process they are supposed to measure. As such, it is difficult to suggest practical subjective measures of acute training effects other than muscle soreness, which is a validated measure, and fatigue for which validated single item instruments do not exist (at least in athletic populations), however multiple items instruments are available (Schnelle et al., 2012). However, this highlights that a reference conceptual framework for validating these measures is needed (and one of the reasons why we developed it).

Other measures

There are other potential measures that can be used to assess acute or chronic training effects. These include biomechanical (e.g. force-velocity profile, acceleration load) (Akenhead, Marques, & Paul, 2017) or other more area specific measures such as cognitive tests (Decroix, Piacentini, Rietjens, & Meeusen, 2016; Hurdiel et al., 2015). For instance, assessment of acute training effects has been examined using cognitive function tests, such as the Stroop and response time tests, in elite cyclists and ultramarathon runners (Decroix et al., 2016; Hurdiel et al., 2015). We have provided examples of measures that can be relatively easily implemented in practice (e.g., jump tests, CK, HRV). However, there are other measures that theoretically can be used but because of their complexity are relegated to the research area such as mitochondrial biogenesis, muscle protein synthesis and markers of various signal transduction pathways (Lysenko, Popov, Vepkhvadze, Sharova, & Vinogradova, 2019). Whilst these cannot be measured in a practical manner (e.g., in a field setting), it does not mean they should be disregarded, as they play an integral role within the physical training process. Frameworks can be used to organize and contextualize fundamental science when thinking about training and its possible effects.

Sport performance outcomes

Sport performance outcomes are the result of the balance between positive and negative training effects, again consistent with the IR-Banister and PerPot models (Calvert et al., 1976; Morton et al., 1990; Perl, 2001, 2004). We provide a possible range of sport performance outcomes according to the ECSS/ACSM consensus on overtraining (Meeusen et al., 2013). These include improvement, no change, and then a progression from functional to non-functional overreaching and finally over training (Meeusen et al., 2013). Functional overreaching is used to improve performance and often applied during normal training cycles. It involves intensified training, which generally results in reduced performance, however after adequate rest an athlete's performance will improve relative to baseline levels. Conversely, if intensified training continues, without adequate rest, this may result in non-functional overreaching causing decrease in performance or stagnation potentially lasting for several weeks or more (Meeusen et al., 2013). Despite this, with appropriate rest an athlete should fully recover. Towards the end of the continuum is overtraining in which performance decrement may last for months.

Individual and contextual factors

Contextual factors can be defined as all the factors not part of the main process (training) such as environmental, social, cultural factors, etc. (Coles et al., 2017). Individual factors are characteristics of the individual athlete such as genetics, psychological traits and states, training background, etc. These factors interact, influence, alter, facilitate or constrain all the components of the training process (Coles et al., 2017). In the conceptual framework, paths are depicted graphically by uni or bidirectional arrows. We conceptualized arrows as representing lines or directions of influence (potential causality). The bidirectional arrow represents a reciprocal nature of interactions between variables. We have included a bidirectional arrow between individual and contextual factors and the training effects. For example, a negative training effect (e.g., increased fatigue) can act as an individual factor subsequently influencing the internal training load in the following training session (indicated by the unidirectional arrow). On the other hand, other individual or contextual factors may also influence the training response (e.g. causing higher or lower negative effects) (Durand-Bush & Salmela, 2002). The unidirectional arrows indicate that the internal training load resulting from a specific external load will vary based on individual or contextual factors. For example, individual or contextual factors can moderate the effect of residual fatigue on the internal training load elicited by the planned external load. Similarly, modifications of individual and contextual factors can influence sport performance outcomes. For example, psychological status, nutrition (interventions or cultural aspects) or recovery strategies may impact sport performance outcomes, in either a negative or positive way (Beck, Thomson, Swift, &

von Hurst, 2015; Chtourou et al., 2011; Judge et al., 2016; Nédélec et al., 2015). Hypoxia induced by altitude training is an example of a contextual factor that can influence all the training process components, and that can be influenced by other contextual factors and interventions (Burtscher et al., 2018; Caris & Santos, 2019). Clearly, some individual and contextual factors are modifiable (health, nutritional status, environmental conditions, etc.) and others non modifiable (e.g., genetics, training history, etc.). We clarify that forms of training other than physical may be classified as contextual factors, that is strategies altering individual factors that can influence performance (e.g., mental or cognitive training) (Bühlmayer, Birrer, Röthlin, Faude, & Donath, 2017; Gould, Damarjian, & Greenleaf, 2002).

Examples of practical and theoretical applications of the conceptual framework

Training monitoring

One practical utility of this conceptual framework is its application to training monitoring, that is the identification of the measurable components and their role within the training process. The conceptual framework together with the operational definitions allows us to understand what and why to measure these components, thus also facilitating the interpretation of the measures. As another practical example of application, it can be used to adjust and adapt the training process and periodization plans. Typically, periodization for athletic preparation focuses on the exercise component, by developing a program using external load, eventually estimating the induced internal training load using background knowledge and available evidence. In parallel, the planned external training load can be adjusted based on athlete responses and feedback (Kiely, 2012). This is an approach that may also require coaches to use heuristic methods (Gigerenzer & Gaissmaier, 2011). Our conceptual framework does not present any innovative strategy in this regard (Nosek, Brownlee, Drust, & Andrew, 2020), however it may assist in accommodating training based on individual responses by taking advantage of the measures that are nowadays possible to collect. We provide a model according to which the measures (type and timing) are selected based on what construct we want to quantify and for what reason (to measure an acute or chronic effect, an individual factor, etc.).

Overall, the measures of the component of the training process can assist in identifying and developing measures useful for informing decisions, controlling training load and determining if responses are progressing as intended, and/or the program needs to be modified. For example, a negative effect that is planned is not necessarily a problem. However, when a negative but unplanned effect occurs, this may be a signal of something amiss or something is changing (e.g., in contextual and individual factors). Therefore, modulation of external training load using feedback from the training effects, which are influenced by individual and contextual factors (bidirectional), may provide a supplementary "optimization" of the training process. With the term "optimization" we mean a training process which is flexible enough to be modified based on individual training responses and effects.

Differentiation between acute and chronic effects is also possible. The monitoring of acute training effects over time (time series) allows for the evaluation of trends to see if an acute effect is becoming chronic (e.g., fatigue). By adding a recovery short period (e.g., a few days, one week) may help identify if this trend indicates a chronic condition or whether this is just an acute and transient training effect (i.e., according to one of the criteria we proposed for the differentiation between acute and chronic). Other strategies to assess acute and not chronic negative effects can be to measure fatigue before and after training, with the difference used as acute fatigue induced by the training. The measure before training can be seen as an integration of an acute-chronic effect with individual and contextual factors. For example, fatigue due to recent and accumulated negative effects (chronic) combined with other factors such as nutrition and psychologically stressful situations outside of training. The difference between pre- and post-training measures of fatigue is the acute training effect.

Training tolerance

It is common to refer in the training context to the ability of an athlete to cope with the demands of the training load. For example, to understand whether or how to adjust the training program, i.e., proceeding as planned or needing modification. The "ability to cope" may be defined as training tolerance, which is a "field" term commonly used and can be interpreted as a higher order construct, difficult to directly quantify. The proposed conceptual framework provides the opportunity and method to better define and assess training tolerance. For example, higher-than-normal fatigue (quantified using subjective or performance-based measures) caused by a typical internal training load may indicate lowering in training tolerance. Also, for consideration is the external training load, where for example, high levels of muscle soreness are to be expected when training includes eccentric exercise or in unaccustomed athletes (Baird et al., 2012; Cheung, Hume, & Maxwell, 2003). However, if high levels of muscle soreness are reported after training, including primarily concentric exercise in accustomed athletes, this acute negative training effect may be indicative of poor training tolerance.

Performance readiness

As mentioned in previous sections, subjective instruments may be used to assess training effects. Unfortunately, commonly used subjective measures have not been properly validated and as a consequence it is unclear whether they are actually measuring what they are supposed to measure, and what their role is within the training process. However, using a reference framework assists in understanding how, why and what constructs we wish to measure to support and understand the training process. Not without reason, the COSMIN guidelines require a reference framework for evaluating the validity of an instrument.

Understanding of the necessity and application of frameworks in the validation process is a conceptual matter grounded in research methodology. To explain how a framework can be used in the validation process, we provide an example of performance readiness, a concept widely used in sport settings, although never properly or clearly conceptualized (Australian Institute of Sport, 2020). The lack of conceptualization is evident from the measures used to assess this construct, and variations of the descriptors (e.g., player readiness, readiness to train). Furthermore, the same items for measuring wellness are also used for assessing performance readiness (McGahan, Burns, Lacey, Gabbett, & O'Neill, 2018, 2020). It is questionable how it is possible that the same items can measure two constructs. In an effort to provide an explanation, we speculate they implicitly represent two different but related constructs. Differentiating between these would require separate operational definitions and/or concept elicitation studies. Unfortunately, no such information or research exists. For exemplificative purposes, we refer to "wellness" to identify items commonly used in the sport science literature under this name (Jeffries et al., 2020). The link between the two constructs can be hypothesized using the conceptual framework. Wellness items may reflect training effects, with performance readiness as a higher order construct influenced by or "incorporating" these subjective training effects. In one of the few attempts to provide an operational definition, Ryan et al. (Ryan, Kempton, Impellizzeri, & Coutts, 2019) proposed that optimal player readiness is a condition where an athlete has no impairment of physical performance, no mental fatigue or excessive psychological distress, and "the athlete's capacity to complete training activities and perform during competition". This is consistent with the use of this construct in the literature that refers to subsequent performance or physical training (Cullen, McCarren, & Malone, 2020; Mason, McKune, Pumpa, & Ball, 2020). Using this definition (others may be possible), physical performance seems to be the criterion for validation. Therefore, the lack of association between candidate training response effects, assumed to be related to performance readiness and hence subsequent training or competition physical performance, would not support the validity of these items as measures of performance readiness (Cullen et al., 2020). However, they still may be valid measures of training effects. For example, fatigue perception can be a valid measure of acute training effects, although may not be related to subsequent physical performance (e.g., high fatigue corresponding to lower external load) and therefore cannot be used for assessing performance readiness, or readiness to train.

To further illustrate, below we provide hypothetical examples for items measuring subjective training effects (e.g., wellness items) that may be related to performance readiness. We developed two hypotheses presented in Figure 5.2 A and B.

A) The sum of single item scores are used to measure a training effect that is supposed to reflect performance readiness. Therefore, an association between their summary score, eventually plus measures of contextual factors, and performance is expected.

B) Only two of these items are supposed to measure performance readiness. Convergence of these two items (correlation with physical performance) and divergence for others (no or weaker correlation with physical performance) are expected.

Clearly, depending on the hypothesized relations, the presence or lack of associations provide or do not provide evidence for validity of these items as measures of performance readiness. In example A, there is no interest in the correlations between physical performance and the single items but in the summary score (assuming the summary score has been validated). In example B, our interest is specifically the single items of which some are not expected to be correlated. In that case, even if the items combined measure a specific training effect construct, our focus is on single items as we are interested in their isolated association to another construct, which is performance readiness. Simply calculating correlations would not be useful without a reference framework, and associations would not be interpretable. Exploratory correlational studies are clearly possible, but at least two issues should be considered. First, the finding can only be used to generate hypotheses that need to be subsequently tested and are not findings that can suggest how or whether to implement these items in the monitoring process. Second, the multiple associations increase the risk of alpha inflation and this should be considered when interpreting the results and/or building hypotheses based on the findings.

These are hypothetical examples and although based on background knowledge and theories, different kinds of relations and/or causal structures can be hypothesized and tested. However, in validation studies these predictions should be explicitly declared and the conceptual framework (or theory) supporting these hypotheses should be presented. Similarly, the nature and real goal of the study should be explicitly reported (e.g. confirmatory, exploratory), otherwise a fishing expedition in the data lake could be generated (adding confusion). Again, we underline it is important to differentiate whether a measure is an attempt to quantify an effect, an individual or contextual, or an effect influencing individual or contextual factors. It is also important to understand how single items, or composite scores, are supposed to influence the associated constructs. This can have a profound effect on the validation process and interpretation of the results. Validating measures is theoretically complex and for this reason necessitates a reference conceptual framework.

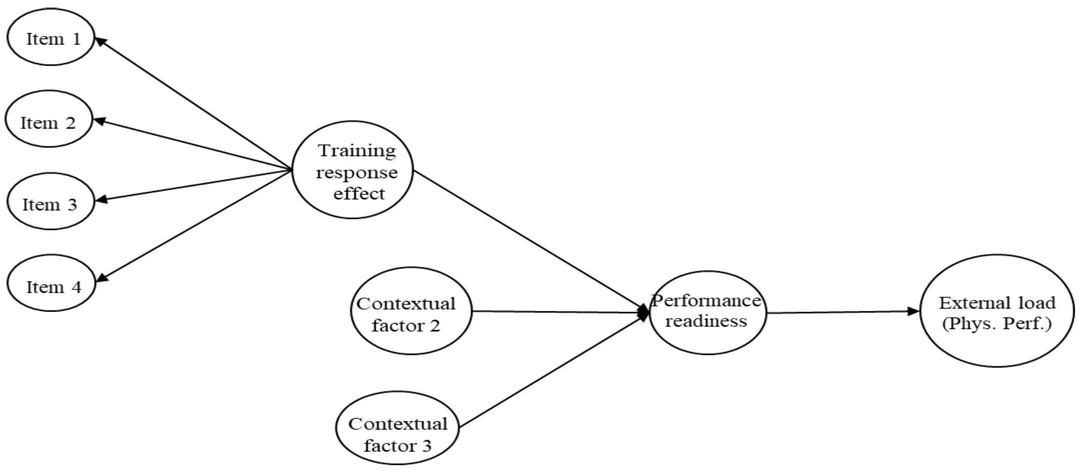


Figure 5. 2: Hypothetical examples. Example A shows 4 items reflecting and used to obtain a summary score for a training response effects. This response measure, together with other two measures of contextual factor, are supposed to reflect or determine (depending on the model used) performance readiness. A study may want to see whether training effect measures and contextual factors belongs to the same construct (reflective model): performance readiness. And/or, whether this summary score (combining the three factors) that is supposed to measure performance readiness, is related to physical performance, which is the criterion for the validation.

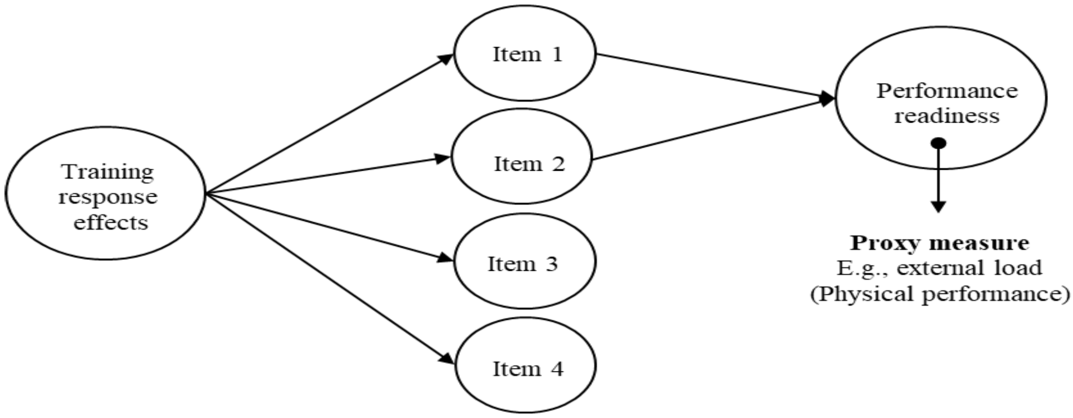


Figure 5.2: Example B shows, instead, that 2 out of 4 items measuring training effects are supposed to also reflect performance readiness. A study may test this hypothesis, where item 1 and 2 are hypothesized to be associated (converge), and 3 and 4 don't associate (diverge). If there is not a validated measure of performance readiness, physical performance can be used as proxy measure (surrogate), when performance readiness is, by operational definition, associated to the amount of physical activity completed in the training or competition.

Limitations

Given the qualitative approach of this study, researcher bias is a potential limitation that should be acknowledged. We have assumed, as generally accepted, that optimizing training using measures of the constructs of the training process is superior in terms of sport performance outcomes. However, this assumption is based on existing training theory, experience and inductive reasoning, but not experimental studies. Furthermore, our conceptual framework provides a tentative theory of the phenomena under investigation (physical training), and may need refinement, verification or further explanation.

Summary and conclusion

A conceptual framework synthesizes evidence by conceptualizing a phenomenon. Furthermore, it allows for a rigorous, valid and reliable research process, and may serve as a practical tool for interpretation of measures collected in training monitoring. In the development of our conceptual framework, we have built on existing frameworks and models to better explain the training process and its outcomes. In this paper, we have described the conceptual framework development process and presented operational definitions. We introduced and expanded on four concepts: training effects and its possible measures, sport performance outcomes, and individual and contextual factors. Additionally, we explained the relevance and interconnections of these concepts within the training process. The resulting conceptual framework is coherent and fits conceptually available models such as IR-Banister and PerPot. Finally, we presented examples of theoretical (role in the validation process) and practical applications (monitoring, training tolerance, periodization) based on the current conceptual framework. These examples were clearly hypothetical to practically demonstrate how the use of a framework may clarify and explicitly present precise and theoretically grounded testable hypotheses. This provides a useful overarching model for understanding the scientific literature regarding training and additionally guides the development, implementation, and evaluation of a more comprehensive and transparent approach to athlete monitoring and validation of measures.

Acknowledgements: We would like to thank Professor Fabio Nakamura and Carlo Buzzichelli for their assistance.

Chapter Six

Development and validation of single items for fatigue and recovery in dancers

Jeffries, A.C., Novak, A.R., Coutts, A.J., McCall, A., McLaren, S.J, Impellizzeri, F.M. (*Under review*). Development and validation of single items for fatigue and recovery in dancers. *Journal of Sports Sciences.*

Preface

One of the findings from *Study One* was that professional dancers experience high training loads, together with very few recovery periods within and between macrocycles. Considering that professional dancers may need to undertake high training loads in order to achieve optimal performance, the monitoring fatigue and recovery may be useful single items for the measurement of acute or chronic training effects. Furthermore, no instruments are available that assess both absolute value and the perceptions in relation to personal standards, which may be highly individual. However, given the lack of studies examining the validity of single items as identified in *Study Two,* we then planned *Study Four*, based on our conceptual framework (*Study Three*), to examine the validity and reliability of single items (absolute and relative) fatigue and recovery in dancers.

Abstract

Purpose: Examine the construct validity and reliability of two single items for fatigue and recovery in dancers. The construct validity was assessed using reference instruments; the fatigue items of the Brunel Mood Scale (BRUMs) and the Short Recovery Stress Scale (SRSS). A secondary aim was to explore the respondent interpretation of these two items using a concept identification approach.

Methods: Two hundred and forty-three dancers completed single items fatigue and recovery (unipolar and bipolar), BRUMs and SRSS once for construct validity. For reliability (n=49) completed the questionnaires twice, one week apart. Using a concept identification approach (n=49) dancers were also asked comprehension and interpretation of fatigue and recovery.

Results: Construct validity, reliability and agreement for the single item of fatigue were confirmed. The recovery item, despite acceptable reliability and agreement, was only partially confirmed in terms of construct validity, when using the SRSS recovery items as reference. Tiredness, muscle soreness and fatigue were identified as the most common concepts reported for fatigue and recovery.

Conclusion: We provide preliminary confirmation of the validity and reliability of the single item fatigue in dancers. Further research is warranted further exploring other measurement properties.

Keywords: COSMIN, monitoring, training effects, measurement properties, self-report, artists

Introduction

Athlete Reported Outcome Measures (AROMs) are frequently used in practice and research to measure the training response with a dramatic increase in the past five years (Jeffries et al., 2020). Whilst there are many AROMs and measurement tools available, the most commonly used are the single or 'wellness' items, but most have not been properly validated or validated at all (Jeffries et al., 2020). Only multiple item AROMs derived from psychometrics possess adequate levels of evidence for measurement properties, with the exclusion of content validity and measurement error. Therefore, there is a critical need to develop and examine the validity of instruments by adopting appropriate methodology using established reference methods when developing AROMs. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) is an established reference set of recommendations that we recently suggested to guide development and validation of measures aimed to assess symptoms and physical signs (Jeffries et al., 2020).

Given the importance of a reference framework in the validation process, as highlighted by the COSMIN standards (Mokkink et al., 2012), we developed a conceptual framework to facilitate the interpretation and validation of measures of training process components (Jeffries et al., 2020). This was an adaptation of a previous framework used in sport science (Impellizzeri et al., 2019; Impellizzeri et al., 2005), introducing a new measurable construct (training effect) and expanding three concepts: possible measures of training effects, performance outcomes, and contextual modifiable and non-modifiable individual factors. In defining the reference framework, it is possible to indicate what components of the training process a measure is purported to assess. Additionally, a conceptual framework may reduce bias, such as post hoc and ad hoc justification of the results. For brevity, our reference framework is not included in this paper but can be viewed elsewhere (Jeffries et al., 2020).

There are many AROMs available for monitoring the training response including multiple and single item scales. Although traditional multiple item scales may have superior psychometric properties, there are several reasons for the use of single items including minimizing respondent burden, reducing criterion contamination and increasing face validity. For example, it has been shown that survey length has detrimental effects on response rates (Crawford, Couper, & Lamias, 2001). Therefore, reducing survey length may increase overall response rate and minimise negative consequence of nonresponse bias (Rogelberg & Stanton, 2007). While multiple items are overall superior, single items are attractive for their easy implementation, especially when used frequently (e.g., daily) as is typical in sport. If the construct is unambiguous and sufficiently unidimensional, single items may be psychometrically acceptable alternatives to multiple item measures. Additionally, single-item measures have been used in a variety of research fields (Milton, Bull, & Bauman, 2011; Mitchell, 2007; Patrician, 2004; Reissmann, John, Kriston, & Schierz, 2013; van Hooff, Geurts, Kompier, & Taris, 2007).

Subjective symptoms can be used to measure the training effect construct to assess how an athlete is responding to the demands imposed by training. Fatigue measures are the most commonly used and can be considered unidimensional (Michielsen, De Vries, Heck, Van de Vijver, & Sijtsma, 2004). Recovery is another common measure, such as the total quality recovery scale (Kenttä & Hassmén, 1998), which is not validated according to psychophysics or psychometric methods. Other instruments assessing recovery are the recovery scale of the REST-Q (Kellmann & Kallus, 2001) and shortened versions (ARSS and SRSS) (Hitzschke et al., 2015; Kellmann et al., 2016), however given they include multiple items, their implementation for routine assessment is limited (e.g. on a daily or weekly basis). One item included in both abridged versions of the REST-Q concerns 'overall recovery'. However, as a measure of acute training effects to single sessions may not be temporally specific. In addition, the items include examples (i.e., descriptors) which may increase the risk associated with multiple-barrelled questions, potentially adding confusion. While we do not detract its utility, we believe a variation may be more suitable to measure training effects as a single item, by restricting the reference time frame and avoiding examples (under the assumption of unidimensionality).

Within dance, it is common to rely on sport instruments, however this should require independent validation specific to this population. Given that dancers undertake high training loads, together with potentially inadequate recovery (Jeffries et al., 2020), a purposefully validated instrument to assess these constructs may be useful to quantify and monitor training effects. Therefore, the primary purpose of the research was to examine construct validity (convergence and divergence using a hypothesis testing approach) and reliability of two single items for fatigue and recovery, the former taken from the literature and latter purposely created. The construct validity was assessed using reference instruments. A secondary aim was to explore the comprehension and the interpretation of these two items using a concept identification approach.

Methods

Study design and recruitment of participants

This was a cross sectional web-based study. After ethical approval (ETH20-4675), data were collected between April 2020 and August 2020. Participants were provided with an information page describing the purpose, length of survey (10 minutes to complete), data storage security and the investigators involved in the study. The survey data were collected and managed using Research Electronic Data Capture (REDCap) (Harris et al., 2019). The survey was advertised through social media and industry contacts. The survey was voluntary, with no monetary incentives and to prevent bias, the order of items randomized. Inclusion criteria were: ≥ 18 years old, a dancer of any level (professional, amateur etc.) and any style (contemporary, ballet, jazz). Exclusion criteria was dancers < 18 years. The COSMIN guidelines and reporting of results of e-surveys was followed (Eysenbach, 2004). For brevity, we had to omit some methodological detail as per COSMIN guidelines due to word limits.

Participants and data collection

A total of 243 participants completed an online survey to investigate construct validity. Of these participants, 49 completed the survey a second time under the same conditions to investigate reliability, with each measure separated by seven days. These 49 participants were not involved in performances or interventions and were therefore considered stable across the seven-day testing period. However, some variations in their state may have occurred and this cannot be completely ruled out. Forty-nine participants completed two additional questions regarding comprehension and interpretation of the fatigue and recovery items for concept identification. Unfortunately, the time frame for this study was defined before the emergence of the COVID-19 pandemic. As such, some data was collected during lockdown and home confinement, with dancers training at home. The reliability and concept identification subsample completed the questionnaires in June and July 2020, after the peak emergency and when reassuming training. Participant characteristics are provided in Table 6.1. Participants' data were excluded entirely if any items were incomplete.

Instruments

Fatigue subscales of the Brunel Mood Scale

Participants completed 4 items from the fatigue subscale (worn out, exhausted, sleepy, tired) of the Brunel Mood Scale (BRUMs) (Terry et al., 1999). The four items were rated on a 5-point intensity scale (0 = not at all; 4 = extremely). Total subscale scores range from 0 to 16. Additionally, three items (worn out, sleepy, tired) included in the Multiple Component Training Distress Scale (Main & Groove, 2009) were also summated into a total subscale, however the data are not presented, as is very similar to the BRUMS (i.e., it consists of 3 out of 4 BRUMS items).

Short Stress and Recovery Scale

The Short Stress and Recovery Scale (SRSS) consists of eight items, four related to stress, and four to recovery (Hitzschke et al., 2015). The stress-related scales are Muscular Stress, Lack of Activation, Negative Emotional State, and Overall Stress. The recovery-related scales are Physical Performance Capability, Mental Performance Capability, Emotional Balance, and Overall Recovery. Each item contains a supporting list of adjectives that describe the stress or recovery construct. The participants were instructed to rate how much each expression applied to them at the moment, with responses ranging from 0 (does not apply at all) to 6 (fully applies).

Single item measures

We developed two single-item measures for fatigue and recovery with two contrasting scaling types. The single item measures were a simple statement of 'how fatigued do you currently feel?’ and ‘How well did you recover from the last training session?’. Both items were rated on a unipolar (absolute) response option (1 “not at all” to 10 “completely”) and bipolar (relative) response option (-3 “much worse than normal”, 0 “normal” to +3 “much better than normal”) (Figure 6.1 and 6.2). The fatigue item has previously been validated in other areas (van Hooff et al., 2007). We used two response options (uni and bipolar scales), with the bipolar scale used as a comparative scale.

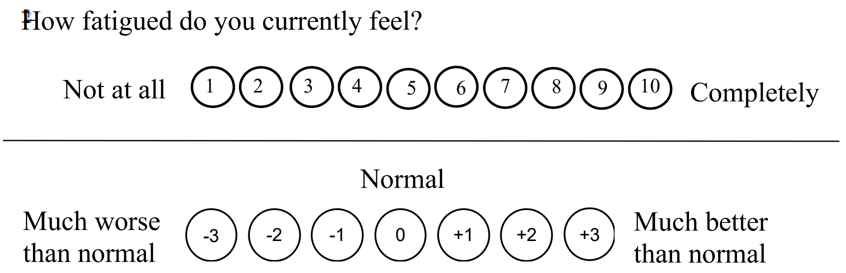


Figure 6. 1: Unipolar and bipolar scale for fatigue

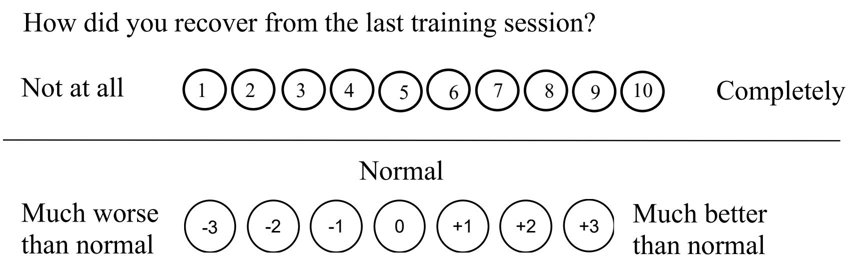


Figure 6. 2: Unipolar and bipolar scale for recovery

Preliminary concept identification

Directly following each of the single item questions for fatigue and recovery, participants were asked to provide short statement(s) on why they gave that number/rating. A saturation table organized by concept code for fatigue and recovery was developed to systematically document concepts emerging from the participant open-ended questions. This tracks the appearance of new concepts appearing in each of the participant responses and enables the comparison of emerging concepts (Patrick et al., 2011). Saturation was confirmed when no new important or relevant concepts emerged in the final set of participant responses. A total of 49 participants completed the concept identification section, however some responses were removed due to lack of clarity (fatigue n=3 and recovery n=13).

Statistical analysis

Statistical analysis was performed using SPSS software, version 25. Examination of normality of residuals and the distribution of the single items, SRSS and BRUMS was performed, and descriptive statistics reported as mean and median scores, standard deviations (SD), interquartile range (IQR) and range (calculated for all the items). Correlational analysis was undertaken using ggplot 2 in R statistical computing software version 3.6.2 (R Core Team, 2019) and visualised using correlograms.

Construct validity

To evaluate construct validity, we used the hypotheses testing guidelines of COSMIN. Construct validity is defined by COSMIN as "the degree of which the scores of a measurement instrument are consistent with the hypotheses, with regard to internal relationships, relationships with scores of other instruments or differences between relevant groups" (Mokkink et al., 2010). The COSMIN guidelines recommend that hypotheses are formed a priori with specific and clearly defined direction, magnitude and rationale (Mokkink et al., 2010). A priori, we developed 24 independent hypotheses. For each of these formulated hypotheses, the anticipated (spearman) correlation direction, criteria and rationale for which the hypothesis is based were clearly defined (Table 6.2). The basis of the hypotheses were formed to reflect that the single items of fatigue evaluate the construct of fatigue in the BRUMS and stress of the SRSS. Although the latter was considered related but not a measure of the same construct. Additionally, our single item of recovery would evaluate the construct of recovery from the SRSS. To confirm this, we expected that the single item of fatigue would be positively correlated as it measures a similar construct (fatigue of BRUMS), and the recovery item would be negatively correlated as it measures a similar construct (recovery of SRSS). We also provided in the hypothesis the criteria in relation to the confidence intervals and not only to the mean estimates. Confidence intervals were interpreted as compatibility intervals, that is the set of all effects that if hypothesized to be the true unknown value of a parameter, the statistical test at that "null" hypothesis would not be rejected at the 0.05 level (95% CI).

Agreement and reliability

According to the COSMIN manual, we calculated agreement and reliability as percent agreement and weighted kappa for the new single items and (for comparison) the SRSS (Mokkink et al., 2012). Percent agreement assumes that the rater made a deliberate choice and not a random guess (Cohen, 1960). While a random guess seems to be unlikely in the context of this study (i.e., participants volunteered), it cannot be excluded. However, the probability of providing the same rating in two repeated measures by "guessing" is, in case of 7 to 10 ordinal categories, negligible (1 to 2%). For ordinal data "agreement" can be calculated as the difference below a certain level, e.g., minimal clinically important change. This figure is not available for these single items. We therefore reported the raw observed percent of agreement and we discussed the results qualitatively [discussion combining <1% and cut off >80% as good]. Observed percent agreement can help interpret the results or repeated measures on the same subjects. We also calculated index bias and differences between median test and retest values (systematic bias) using the Mann Whitney U test.

Contrary to raw percent agreement, Cohen's kappa and variations account for the expected agreement. In this study we used linear weighted kappa that consider two ratings can vary by a small or large amount (Cohen, 1968). As generic benchmarks for interpreting kappa, we referred to Landis and Koch (Landis & Koch, 1977) (0.41 to 0.60, moderate; 0.61 to 0.80, substantial; 0.81-1.00, almost perfect), Fleiss (Fleiss, 1971) (0.40 to 0.75, good; >0.75, excellent) and Altman (Altman, 1991) (0.41 to 0.60, moderate; 0.61 to 0.80, good; 0.81-1.00, very good).

Results

Demographics and distribution

A STROBE flow-chart of participants is presented in Figure 6.3, indicating the eligibility and number of participants at each stage of the study. Five questionnaires of the 248 were excluded from the analysis due to incomplete data and one error and therefore unlikely to have introduced survey response bias. Table 6.1 outlines the participant characteristics for construct validity and reliability. In total 243 participants were eligible for the study (female: 69.1%; n=168, male: 29.2%; n=71) with the mean age 25.4 ± 8.1 years (range 18-61). The main dance style was contemporary (53%; n =129), with most professional or international level (57.6%; n=140) or semi-professional (26.3%; n=64). The distribution of the scores of the items (median, interquartile, minimum and maximum) is presented in Table 6.3.

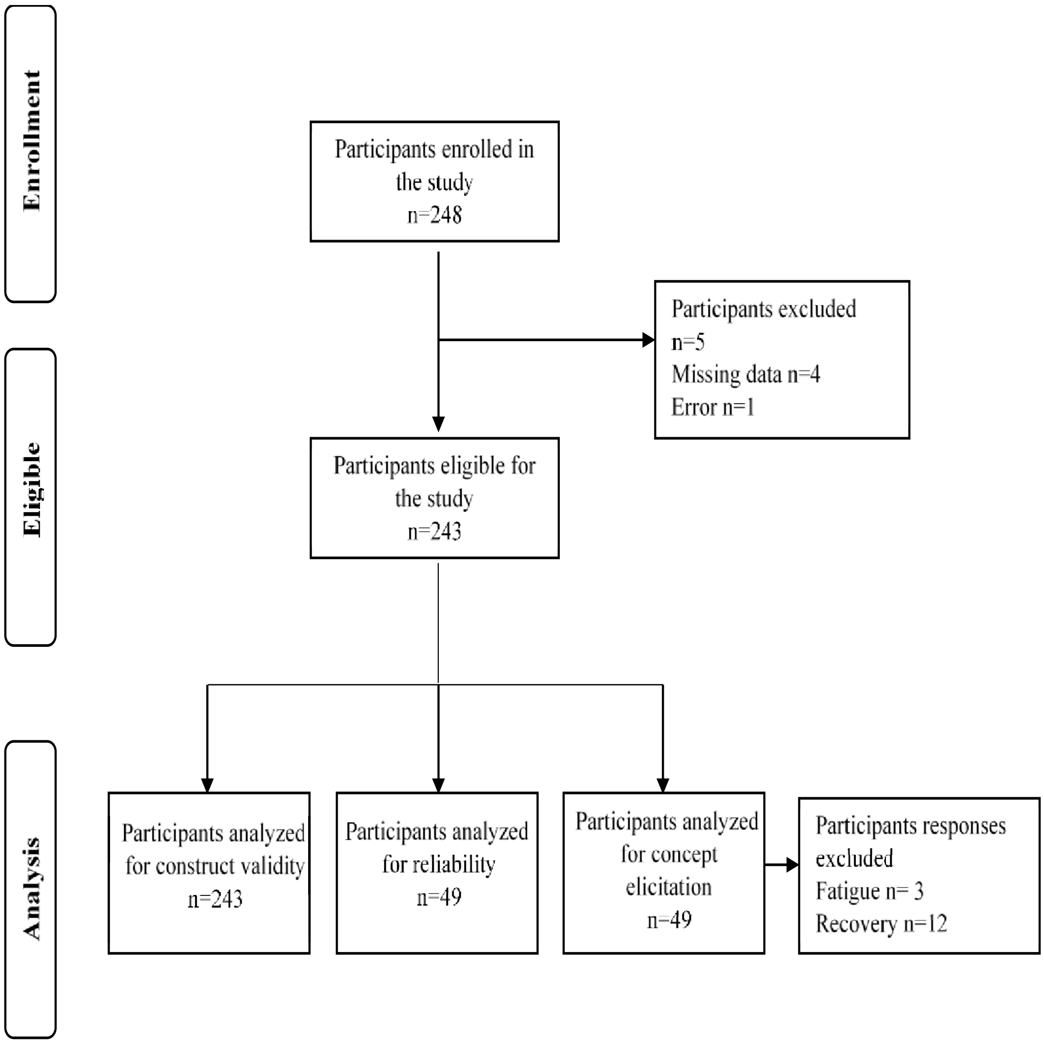


Figure 6. 3: Flow chart showing participant eligibility and available sample sizes for assessing psychometric properties of the Short Recovery and Stress Scale (SRSS), Brunel Mood Scale (BRUMS), Multiple Component Training Distress Scale (MTDS) and single items.

Table 6. 1: Participant characteristics for construct validity and reliability samples.

Characteristics	Construct Validity (n=243)	Reliability (n=49)
Age (Mean, SD)	25.4 ± 8.1	25.4 ± 5.0
Range	18-61	18-35
Gender (n,%)		
Female	168 (69.1)	26 (53.1)
Male	71 (29.2)	22 (44.9)
Other	3 (1.2)	1 (2.0)
Prefer not to say	1 (0.4)	
Primary dance style (n, %)		
Contemporary	129 (53.0)	38 (76.0)
Classical ballet	69 (28.3)	11 (22.0)
Jazz	24 (9.8)	1 (2.0)
Modern	7 (2.8)	
Hip hop	5 (2.0)	
Latin dance	5 (2.0)	
Other*	4 (1.6)	
Secondary dance style (n, %)		
Contemporary	13 (5.3)	1 (2.0)
Classical ballet	7 (2.8)	4 (8.0)
Jazz	3 (1.2)	
Hip hop	3 (1.2)	
Level of dance (n, %)		
Professional/international	140 (57.6)	42 (85.7)
Semi professional	64 (26.3)	6 (12.2)
Amateur	21 (8.6)	1 (2.0)
Recreational	8 (3.2)	
National level	3 (1.2)	
Other*	7 (2.8)	

*Other= Flamenco, physical culture, tap and Ukrainian folk dance
SD: standard deviation

Construct validity

Results for the hypotheses tests are presented in Table 6.2. Hypotheses testing for the single item of fatigue (absolute) resulted in 4 out of 5 hypotheses confirmed and 1 out of 5 partially confirmed; single item fatigue (relative) 5 out of 7 hypotheses confirmed and 2 out of 7 partially confirmed. For the single item recovery (absolute) 2 out of 5 hypotheses were confirmed, 2 out of 5 partially confirmed and 1 out of 5 not confirmed; single item recovery (relative) 4 out of 7 hypotheses were confirmed and 3 out of 7 only partially confirmed. A correlation matrix between BRUMS with single items, SRSS with single items and BRUMS with SRSS are presented in Appendices (I-K).

Table 6. 2: Hypotheses and criteria

Hypotheses	Criteria	Criteria rationale	Confirmed
Fatigue item (absolute)			*(yes/no/partially)* *Bold: yes; normal: no*
Positive correlation with BRUMS and MTDS fatigue scores (two hypotheses were combined since MTDS is calculated using 3 out of 4 items of the BRUMS).	r > 0.50, with lower limit for confidence intervals > 0.50 indicating strong evidence, and > 0.7 suggesting very strong evidence of convergent validity.	According to COSMIN, to confirm convergence large correlations are necessary and considering the uncertainty (CI), instead of mean estimate only, is a stricter (not included in the COSMIN criteria) but stronger criterion.	*Yes* **BRUMS: ρ 0.66 (95% CI 0.56 to 0.74)** **MTDS: ρ 0.61 (95% CI 0.51 to 0.70)**
Positive correlation with all the fatigue items of the BRUMS	r > 0.50, with lower limit for confidence intervals > 0.50 indicating strong evidence, and > 0.7 suggesting very strong evidence of convergent validity.	Same as above.	*Partial* **Tired: ρ 0.64 (95% CI 0.54 to 0.72)** **Worn out: ρ 0.63 (95%CI 0.53 to 0.72)** **Exhausted: ρ 0.64 (95% CI 0.55 to 0.72)** Sleepy: ρ 0.28 (95% CI 0.14 to 0.41)
Positive correlation with SRSS stress items	Mean estimate between 0.30 and 0.50.	Stress is a related but different construct than fatigue, so a moderate relation is expected. COSMIN suggests for related but different constructs a relation between 0.30 and 0.50. Criteria for confidence limits cannot be applied since, given the sample size, this would require a mean estimate exactly of 0.4.	*Yes* **Muscular stress: ρ 0.44 (95% CI 0.30 to 0.55)** **Lack activation: ρ 0.37 (95% CI 0.23 to 0.51)** **Neg.emo.state: ρ 0.39 (95% CI 0.26 to 0.51)** **Overall stress: ρ 0.51 (95% CI 0.39 to 0.61)**
Negative correlation with SRSS recovery items	Mean estimate between -0.30 and -0.50.	As above but referred to recovery (i.e. opposite direction relation).	*Yes* **Physical perf: ρ -0.45 (95% CI -0.57 to -0.34)** **Mental perf: ρ -0.44 (95% CI -0.55 to -0.33)** **Emot bal: ρ -0.50 (95% CI -0.60 to -0.39)** **Overall recov: ρ -0.52 (95% CI-0.62 to -0.41)**
Negative correlation with the single recovery item	Mean estimate between -0.30 and -0.50.	As above referred to recovery (i.e. opposite direction relation).*	*Yes* **Recov ABS: ρ -0.48 (95% CI -0.58 to -0.37)** **Recov REL: ρ -0.51 (95% CI -0.60 to -0.40)**

Recovery item (absolute)			*(yes/no/partially)*
Negative correlation with BRUMS and MTDS fatigue scores (two hypotheses were combined since MTDS is calculated using 3 out of 4 items of the BRUMS).	Mean estimate between -0.30 and -0.50.	Fatigue is related (opposite direction) but different construct than recovery, so a moderate negative relation is expected.	*Yes* **BRUMS: ρ -0.34 (95% CI -0.46 to -0.22)** **MTDS: ρ -0.31 (95% CI -0.42 to -0.18)**
Positive correlation with all the SRSS recovery items	r > 0.50, with lower limit for confidence intervals > 0.50 indicating strong evidence, and > 0.7 suggesting very strong evidence of convergent validity.	According to COSMIN, to confirm convergence large correlations are necessary and considering the uncertainty (CI), instead of mean estimate only, is a stricter (not included in the COSMIN criteria) but stronger criterion.	*No* The results failed to meet the criteria Physical perf: ρ 0.41 (95% CI 0.28 to 0.54) Mental perf: ρ 0.40 (95% CI 0.27 to 0.53) Emot balance: ρ 0.39 (95% CI 0.27 to 0.52) Overall recov: ρ 0.43 (95% CI 0.43 to 0.56)
Negative correlations with BRUMS fatigue items	Mean estimate between -0.30 and -0.50.	Fatigue is a related (opposite direction) but different construct than recovery.	*Partial* **Tired: ρ -0.32 (95% CI -0.43 to -0.210)** **Worn out: ρ -0.36 (95% CI -0.47 to -0.23)** **Exhausted: ρ -0.37 (95% CI -0.49 to -0.24)** Sleepy: ρ -0.09 (95% CI -0.22 to 0.03)
Negative correlations with SRSS stress items	Mean estimate between -0.30 and -0.50.	Stress is a related (opposite direction) but different construct than recovery.	*Partial* **Overall stress: ρ -0.34 (95% CI -0.45 to -0.23)** Muscular stress: ρ -0.29 (95% CI -0.42 to -0.16) Lack activation: ρ -0.26 (95% CI -0.4 to -0.14) Neg emo state: ρ -0.29 (95% CI -0.41 to -0.17)
Negative correlation with the single fatigue absolute item & and positive correlation with single fatigue relative	Mean estimate between -0.30 and -0.50 and between 0.30 to 0.50	Fatigue is a related (opposite direction for absolute, and same direction for relative) but assumed to be a different construct. *	*Yes* **Fatigue ABS: ρ -0.48 (95% CI -0.58 to -0.37)** **Fatigue REL: ρ 0.34 (95% CI 0.21 to 0.47)**

Fatigue item (relative)			*(yes/no/partially)*
Positive correlation with single recovery (relative) item	Mean estimate between 0.30 and 0.50.	We expected a positive correlation with recovery since related even if a different construct (higher fatigue than normal, <0, associated to worse recovery, <0, than normal). Recovery related to fatigue but different construct.*	*Yes* **Recovery REL: ρ 0.47 (95% CI 0.36 to 0.59)**
Positive correlation with single recovery (absolute) item	Mean estimate between 0.30 and 0.50.	We expected a positive correlation with recovery (higher fatigue than normal, <0, associated to lower, worse, absolute recovery).*	*Yes* **Recovery ABS: ρ 0.34 (95% CI 0.21 to 0.47)**
Negative correlation with single fatigue (absolute) item	Mean estimate between -0.30 and -0.50. Moderate correlation is expected since the same absolute value may not correspond to the same fatigue level considered normal between subjects.	We expected a negative correlation with absolute fatigue (worse fatigue than normal, <0, associated to higher absolute fatigue).*	*Yes* **Fatigue ABS: ρ -0.55 (95% CI -0.66 to -0.45)**
Negative correlations with SRSS single stress items.	Mean estimate between -0.30 and -0.50.	We expected a negative correlation with SRSS stress (higher fatigue than normal, <0, associated to higher stress).	*Partial* **Muscular stress: ρ -0.30 (95% CI -0.42 to -018)** **Lack activation: ρ -0.32 (95% CI -0.44 to -0.21)** **Overall stress: ρ -0.40 (95% CI -0.50 to -0.27)** Neg emo state: ρ -0.28 (95% CI -0.39 to -0.15)
Positive correlations with SRSS single recovery items.	Mean estimate between -0.30 and 0.50.	We expected a positive correlation with SRSS recovery (higher fatigue than normal, <0, associated to lower recovery scores). Stress related to fatigue but different construct.	*Yes* **Physical perf: ρ 0.32 (95% CI 0.18 to 0.46)** **Mental perf: ρ 0.38 (95% CI 0.25 to 0.51)** **Emot balance: ρ 0.32 (95% CI 0.19 to 0.44)** **Overall recov: ρ 0.31 (95% CI 0.19 to 0.45)**
Negative correlations with BRUMS/MTDS single fatigue items.	Mean estimate between -0.30 and 0.50.	We expected a negative correlation with BRUMS/MTDS fatigue (higher fatigue than normal, <0, associated to higher BRUMS fatigue items scores).	*Partial* **Tired: ρ -0.42 (95% CI -0.52 to -0.30)** **Worn out: ρ -0.41 (95% CI -0.51 to -0.31)** **Exhausted: ρ -0.41 (95% CI -0.51 to -0.31)** Sleepy: ρ -0.20 (95% CI -0.33 to -0.07)
Negative correlations with BRUMS/MTDS total scores.	Mean estimate between -0.30 and 0.50.	We expected a negative correlation with BRUMS/MTDS total scores (higher fatigue than normal, <0, associated to higher BRUMS total scores).	*Yes* **BRUMS: ρ -0.43 (95% CI -0.53 to -0.33)** **MTDS: ρ -0.41 (95% CI -0.50 to -0.30)**

Recovery item (relative)			***(yes/no/partially)***
Positive correlation with single fatigue (relative) item	Mean estimate between 0.30 and 0.50.	We expected a positive correlation with recovery since related even if a different construct (worse recovery than normal, <0, associated to worse fatigue, <0, than normal). Recovery related to fatigue but different construct.*	*Yes* **Fatigue REL: ρ 0.47 (95% CI 0.36 to 0.59)**
Negative correlation with single fatigue (absolute) item	Mean estimate between -0.30 and -0.50.	We expected a negative correlation with recovery (worse recovery than normal, <0, associated to higher, worse, absolute fatigue).*	*Yes* **Fatigue ABS: ρ -0.51 (95% CI -0.60 to -0.40)**
Positive correlation with single recovery (absolute) item	Mean estimate between -0.30 and -0.50. Moderate correlation is expected since the same absolute value may not correspond to the same recovery level considered normal between subjects.	We expected a positive correlation with absolute fatigue (worse recovery than normal, <0, associated to lower absolute recovery level).*	*Yes* **Recovery ABS: ρ 0.66 (95% CI 0.58 to 0.73)**
Negative correlations with SRSS single stress items.	Mean estimate between -0.30 and -0.50.	We expected a negative correlation with SRSS stress (worse recovery than normal, <0, associated to higher stress scores).	*Partial* **Muscular stress: ρ -0.31 (95% CI -0.42 to -0.19)** **Neg emo state: ρ -0.30 (95% CI -0.41 to -0.18)** **Overall stress: ρ -0.36 (95% CI -0.47 to -0.25)** Lack activation: ρ -0.27 (95% CI -0.40 to -0.14)
Positive correlations with SRSS single recovery items.	Mean estimate between -0.30 and 0.50.	We expected a positive correlation with SRSS recovery (worse recovery than normal, <0, associated to lower recovery scores). Stress related to fatigue but different construct.	*Yes* **Physical perf: ρ 0.45 (95% CI 0.33 to 0.56)** **Mental perf: ρ 0.35 (95% CI 0.23 to 0.47)** **Emot balance: ρ 0.35 (95% CI 0.22 to 0.47)** **Overall recov: ρ 0.45 (95% CI 0.33 to 0.56)**
Negative correlations with BRUMS/MTDS single fatigue items.	Mean estimate between -0.30 and 0.50.	We expected a negative correlation with BRUMS/MTDS fatigue (worse recovery than normal, <0, associated to higher BRUMS fatigue items scores).	*Partial* **Tired: ρ -0.34 (95% CI -0.46 to -0.23)** **Worn out: ρ -0.33 (95% CI -0.44 to -0.21)** **Exhausted: ρ -0.39 (95% CI -0.49 to -0.28)** Sleepy: ρ -0.07 (95% CI -0.20 to -0.06)
Negative correlations with BRUMS/MTDS total scores.	Mean estimate between -0.30 and -0.50.	We expected a negative correlation with BRUMS/MTDS total scores (worse recovery than normal, <0, associated to higher BRUMS total scores).	*Partial* **BRUMS: ρ -0.34 (95% CI -0.45 to -0.22)** MTDS: ρ -0.29 (95% CI -0.41 to -0.17)

*, lower limit CI > 0.70 may also indicate redundancy. P: spearman rank correlation coefficient; CI: confidence interval; BRUMS: Brunel Mood Scale; MTDS: Monitoring Training Distress Scale; Neg.emo.state: negative emotional state; Emot bal: emotional balance; Overall recov: overall recovery; ABS: absolute; REL: relative.

Table 6. 3: Distribution of the scores

Questionnaire	Item	Construct validity (n=243)				Reliability (n=98)			
		Median	IQR	Minimum	Maximum	Median	IQR	Minimum	Maximum
SRSS Recov	Physical performance	4	2	0	6	4	2	0	6
SRSS Recov.	Mental performance	4	2	0	6	4	2	0	6
SRSS Recov.	Emotional balance	3	2	0	6	4	2	0	6
SRSS Recov.	Overall recovery	4	2	0	6	4	2	0	6
SRSS Stress	Muscular stress	3	2	0	6	3	3	0	6
SRSS Stress	Lack activation	3	3	0	6	2	3	0	6
SRSS Stress	Negative emotional	3	3	0	6	3	3	0	6
SRSS Stress	Overall stress	2	3	0	6	2	3	0	6
BRUMS	Tired	2	2	0	4	2	2	0	4
BRUMS	Sleepy	1	2	0	4	2	2	0	4
BRUMS	Worn out	1	3	0	4	1	3	0	4
BRUMS	Exhausted	1	3	0	4	1	3	0	4
BRUMS	Total score[b]	6	7	0	16	7	6	0	16
MTDS	Total score[c]	4.5	6	0	12	6	5	0	12
Single item	Fatigue relative[d]	0	2	-3	3	0	2	-3	3
Single item	Fatigue absolute[e]	5	3	1	10	5	4	1	10
Single item	Recovery relative[d]	0	2	-3	3	0	2	-3	3
Single item	Recovery absolute[e]	6	3	1	10	6	3	3	10

[a]IQR: inter quartile range; SRSS= short recovery and stress scale. Recov=recovery. BRUMS= Brunel Mood Scale. [b]BRUMS total: the sum of all four fatigue items; MTDS= Multi-component training distress scale [c] MTDS total is the sum of the first three fatigue items of the BRUMS. [d] relative=bipolar scale; [e] absolute=unipolar scale

The correlogram in Figure 6.4 shows the correlation between the single items fatigue and recovery (relative and absolute) with the SRSS stress items (lack of activation, negative emotional state, muscular stress and overall stress) and the BRUMs (tired, sleepy, worn out, exhausted and the total score) with the whole group (n=243). Figure 6.5 shows a subset analysis with participants that completed the reference concept identification (n=49). Figure 6.6 shows the correlogram between the single item fatigue (absolute and relative) with the BRUMS (n=243). The correlations between each pair of variables is visualized through a numerical symbol and highlighted by colours. Appendix L shows the correlogram between the single item recovery (absolute and relative) with the SRSS recovery items.

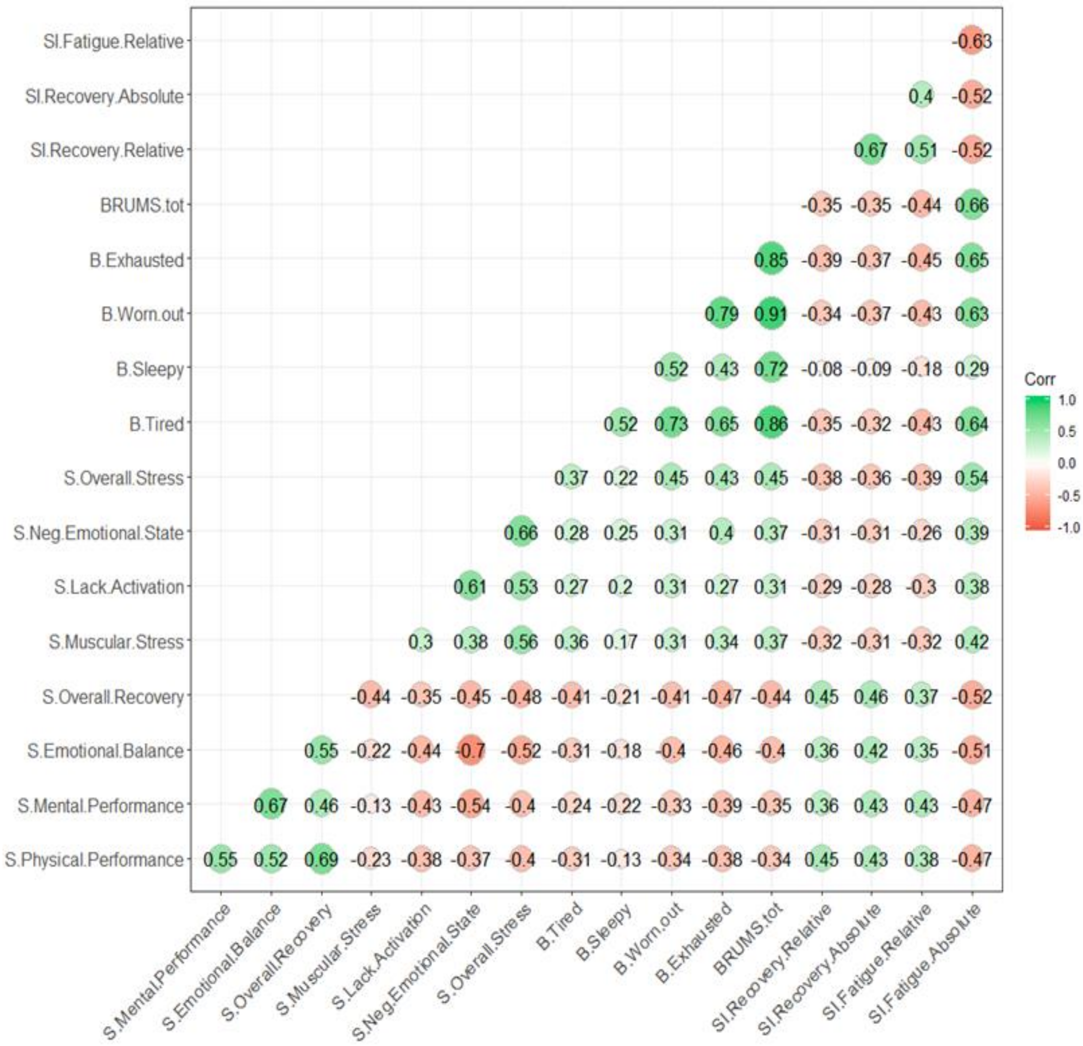


Figure 6. 4: Correlogram matrix of the single items, Short Recovery and Stress Scale and the Brunel Mood Scale for the whole sample (n=243). SI: single item; B: Brunel mood scale; BRUMS tot: Brunel Mood Scale total score; S: Short Recovery and Stress Scale.

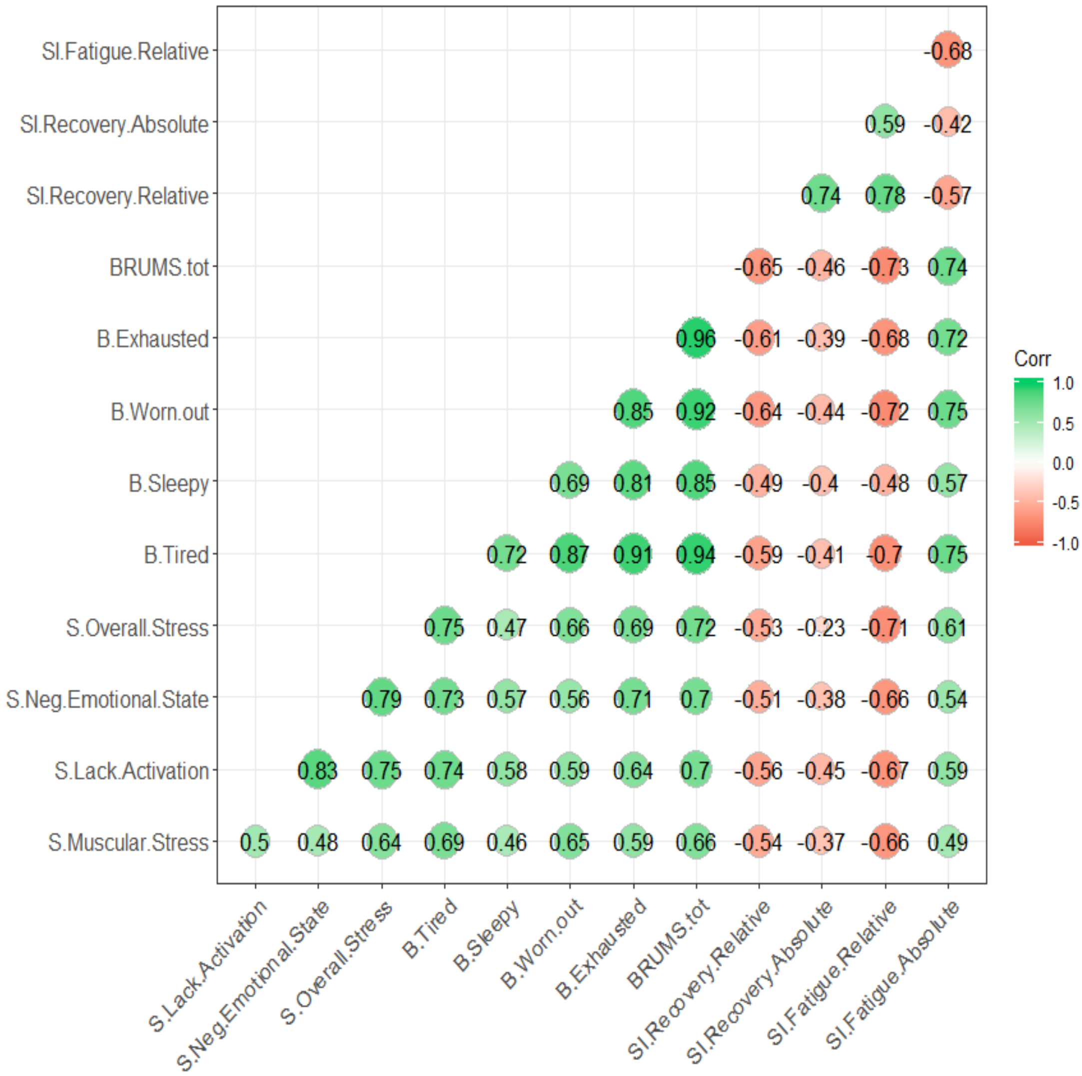


Figure 6. 5: Correlogram matrix of the single items, Short Recovery and Stress Scale and the Brunel Mood Scale for the concept elicitation sample (n=49). SI: single item; B: Brunel mood scale; BRUMS tot: Brunel Mood Scale total score; S: Short Recovery and Stress Scale.

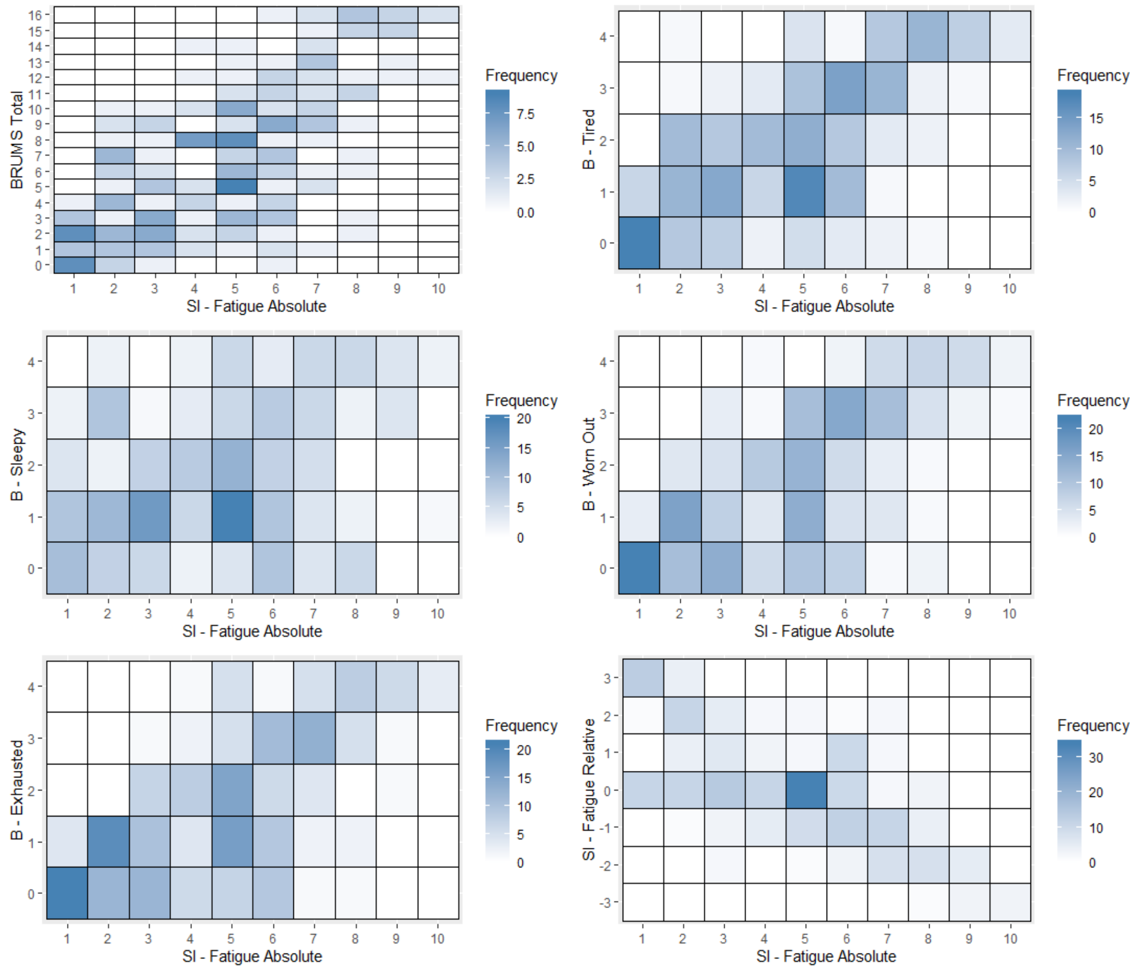


Figure 6. 6: Combined correlgram for single item (SI) fatigue absolute and realtive and Brunel Mood Scale (B).

Agreement and Reliability

Table 6.4 presents the percentages of agreement (<=1 point) and the weighted kappa coefficients for the single items of fatigue (relative and absolute) and recovery (relative and absolute), and the SRSS items overall stress and overall recovery. The level of agreement for the single items of fatigue (absolute and relative) were 90% and 96%, respectively. For the single items of recovery (absolute and relative) level of agreement was 84% and 88%, respectively. For the SRSS, the level of agreement of the overall recovery item was 84% and for the overall stress item 68%. For the single items the weighted kappa ranged from substantial to excellent depending on the refence benchmarks (from 0.71 to 0.78). The weighted kappa of the two single items for recovery and stress of the SRSS (for comparison being used as single items) were smaller (0.53 and 0.66, for stress and recovery respectively), ranging from moderate to good depending on the reference benchmarks.

Table 6. 4: Agreement and reliability

Questionnaire	Item	Agreement					Reliability	
		Absolute difference	Frequency	Valid %	Cumulative %	Agreement	$Kappa_{LW}$	CI
SRSS Recov	Overall recovery	0	28	56	56	Perfect	0.66	0.53, 0.78
		1	14	28	84	Acceptable		
		2	8	16	100			
SRSS Stress	Overall stress	0	26	52	52	Perfect	0.53	0.37, 0.68
		1	8	16	68	Acceptable		
		2	13	26	94			
		3	2	4	98			
		4	1	2	100			
Single item	Recovery absolute[a]	0	29	58	58	Perfect	0.77	0.66, 0.87
		1	3	26	84	Acceptable		
		2	5	10	94			
		3	2	4	98			
		5	1	2	100			
Single item	Recovery relative[b]	0	33	66	66	Perfect	0.71	0.58, 0.83
		1	11	22	88	Acceptable		
		2	6	12	100			
Single item	Fatigue absolute[a]	0	28	56	56	Perfect	0.78	0.69, 0.87
		1	17	34	90	Acceptable		
		2	2	4	94			
		3	3	6	100			
Single item	Fatigue relative[b]	0	34	68	68	Perfect	0.77	0.66, 0.88
		1	14	28	96	Acceptable		
		2	2	4	100			

LW: linearly weighted, CI: confidence interval (lower and upper limit); [a]unipolar item; [b]Bipolar item; SRSS=short stress and recovery scale; % percentage.

Reference concept identification

Figure 6.7 presents the percentage of participants response for concept elicitation of fatigue and recovery. The main responses for the concept of fatigue were tiredness (34.7%), muscle soreness (17.3%) and energy (13.0%). The main responses for the concept of recovery were muscle soreness (43.0%), tiredness (27.9%) and fatigue (24.0%). One of the main responses for the concept of fatigue (26%) was fatigue, indicating a circular answer/tautology. Additionally, attached in the Appendix MA and MB is a saturation table organized by concept code developed to systematically document concepts emerging from successive questionnaires in each of the 49 participant responses. The saturation table tracks the appearance of new concepts emerging in each set of participant response results and facilitates the comparison of concepts emerging in the current set of response results to those that appeared in the previous set of interview transcripts (Patrick et al., 2011). Participant responses were divided into quartiles with an "×" marking representing the first time that a concept was mentioned during the participant responses.

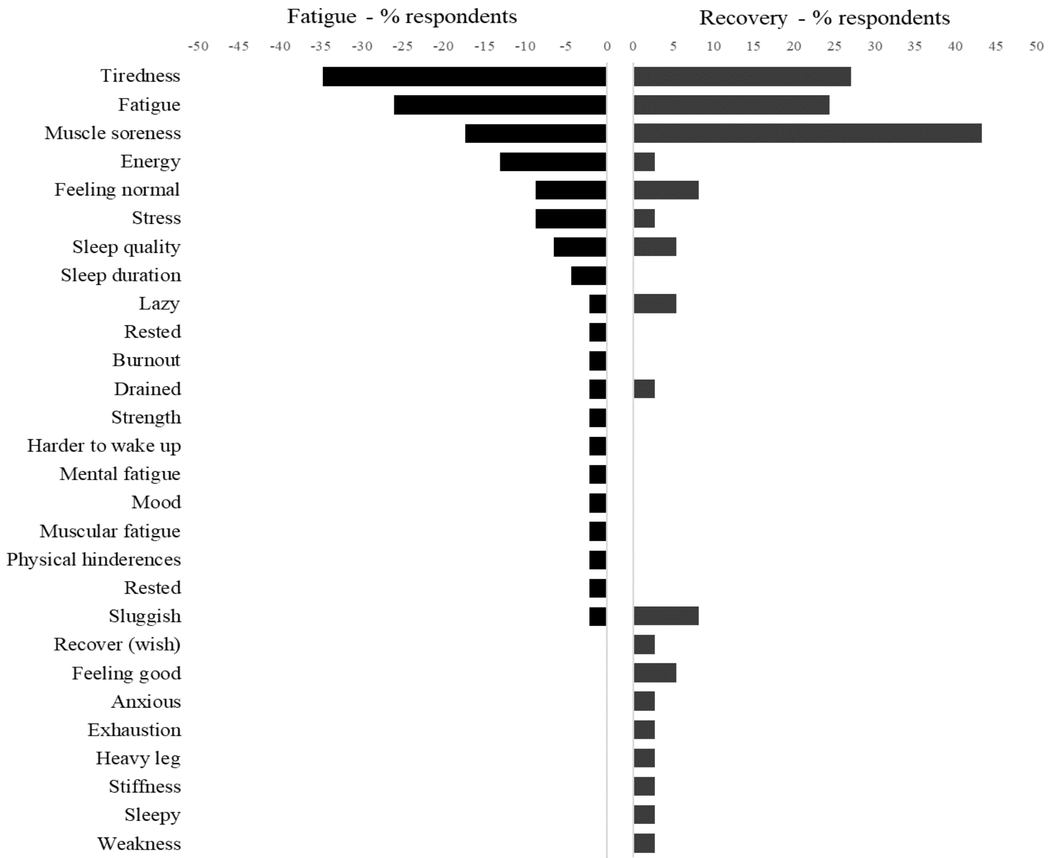


Figure 6. 7: Participant response for reference concept identification of fatigue and recovery.

Discussion

In this study we examined the validity of single items using established methodological guidelines, together with a reference and explicit framework. Our aim was to develop and validate two single items as measures of acute training effects that could be used to monitor training. For fatigue we used a previously validated item from another field (occupational medicine) (van Hooff et al., 2007), asking about the current perception of fatigue, in order to assess an acute training effect. Similarly, we developed a recovery item referring to a defined and short period that is in relation to the last training session. Our findings provide evidence confirming the construct validity, reliability and agreement for the single item of fatigue in dancers. Despite acceptable reliability and agreement, the recovery item was only partially confirmed in terms of construct validity, when using the SRSS recovery items as reference.

Construct validity

Four out of 5 hypothesis tests were confirmed for the single item fatigue (absolute) with only the sleep item of the BRUMS not meeting the predefined criteria. Similarly, results of the single item fatigue (relative) 5 out of 7 hypotheses were confirmed and 2 out of 7 partially confirmed, with weak correlations for the SRSS 'negative emotional state' and the sleep item of the BRUMS. The single item recovery (absolute) produced mixed results, with 2 out of 5 hypotheses confirmed, 2 out of 5 partially confirmed and 1 out of 5 not confirmed, with all the SRSS recovery items, 3 out of 4 SRSS stress items (muscular stress, lack activation and negative emotional state) and the sleep item of BRUMS, failing to meet the predefined criteria for confirmation of hypothesis. Finally, 4 out of 7 hypotheses were confirmed and 3 out of 7 partially confirmed for the single item recovery (relative), with ‘lack of activation’ SRSS Stress item, sleep item of the BRUMS and total score of the MTDS all failing to meet the predefined criteria for hypothesis confirmation. The sleep item of the BRUMS failed to meet all predefined criteria for hypothesis testing in all items.

The relative (bipolar) items of fatigue and recovery met the majority of the predefined hypothesis criteria. In contrast, the absolute (unipolar) items for recovery only partially met the predefined criteria. A possible issue with using unipolar single items to measure broad, bipolar constructs lies in the item design, with participants having to make a unidirectional judgement. The use of a bipolar scale addresses this issue and is characterized by a continuum between two opposite end points, with pairs of contrasting descriptors presented at each end of the pole. A central property of the bipolar scale is that it measures both directionality (side of the scale) and intensity (distance from centre) (Lavrakas, 2008). In summary, the fatigue items met all the criteria for confirming its construct validity, while the recovery items resulted partially confirmed using the SRSS as reference.

The addition of comparative scales (i.e., compared to normal) was chosen because symptoms such as fatigue and recovery are normal acute responses to training. From a practical perspective, knowing whether these ’normally fluctuating’ symptoms are perceived as normal or not, may allow better adjustments of training load and evaluation of training tolerance (e.g., as warning signals). However, the usefulness and validity of these comparative scales should be further evaluated in longitudinal studies. From a practical perspective our results suggest that the bipolar scale for fatigue would be suitable for monitoring training affects in dancers.

Agreement and reliability

In this study we examined the percentage of agreement and reliability of the single items fatigue and recovery (relative and absolute) and for comparison the SRSS overall stress and recovery items. Since the minimal clinical or practical important changes are not yet known, we used the lowest possible change in the response scale (one point) to calculate agreement. Overall, the percentages for our single items were good (>80% level of agreement), indicating raters often choose the same response option (within one point). The Kappa weighted coefficients for the single items ranged from substantial to excellent depending on the reference benchmark (Altman, 1991; Fleiss, 1971; Landis & Koch, 1977) (0.71 to 0.78) suggesting acceptable interrater reliability. A potential reason why reliability is so high for our single-item measures may relate to the singular conceptual clarity of the statement for each item. When comparing with the reference instruments whose single items are supposed to be used in isolation (SRSS), the new fatigue and recovery items showed similar or higher agreement than the overall recovery (84%) and the overall stress item (68%) of the SRSS. Similarly, the Kappa weighted coefficients for the SRSS overall stress and recovery item were lower than those of the new items (Table 6.4).

Reference concepts of respondents

The secondary aim of this study was to explore the comprehension and interpretation of fatigue and recovery using a concept identification approach. In other words, we were interested on what the respondent thought when rating each construct. We believe this is an important aspect to investigate particularly as fatigue is a complex and multifaceted concept for which no established definition exists, and these definitions tend to be context specific (clinical, physiological, psychological, etc.) (Pattyn, Van Cutsem, Dessy, & Mairesse, 2018). However, in order to interpret the scores, it is important to understand what concepts the respondent focus on when providing a rating. The main concepts that emerged for fatigue were 'tiredness' (34.7%), muscle soreness (17.3%) and energy (13%). One of the main concepts that emerged for fatigue was 'fatigue' (26%) however we removed this result from the figure presenting the proportion given the circular nature of the answer. The second most referred concept was tiredness. This is not unexpected and confirms the strong relation with the BRUMS item on tiredness as shown in the subgroup analysis (Figure 6.5) and also in the whole population correlations (Appendix J). Tiredness and fatigue are commonly associated; however, their interpretation is not straightforward. An interesting reconceptualization was proposed by (Olson, 2007) who suggested tiredness, fatigue and exhaustion as three levels of responses to stress, each characterised by different degree of alterations in sleep quality, cognition, stamina, emotional reactivity, control over body processes and social interactions. Interestingly, this author proposed fatigue not as continuum, but a state along a continuum with tiredness and exhaustion as distinct states (the Fatigue Adaptation model). Our results highlight that the respondents perceived the concept of 'fatigue' as related but different than tiredness.

Another notable finding is that 17% of respondents refer to muscle soreness, suggesting that when rating fatigue, dancers also consider and value physical symptoms. For recovery, none of the respondents referred to recovery (as for fatigue), but they mainly referred to tiredness (27%), fatigue (24%) and muscle soreness (43%). This explains the negative correlation with the fatigue items and suggests that perception of recovery may be affected by other constructs. The high proportion of dancers indicating muscle soreness, again, underlies that physical symptoms are important components of the perception of both fatigue and recovery (more than other psychological aspects). Within our reference conceptual framework, we were interested in measures of acute effects of physical training sessions, i.e., physical related symptoms. The concepts referred by the respondents seems to be more "physical" with psychological aspects reported but less frequently. This supports their conceptual suitability to measure acute training effects, but also highlights that these single items cannot grasp more psychological-related responses. For the latter, however, there are several psychometric instruments that can be used and eventually validated in athletic or performance artist population. While these findings are preliminary and partial (this was not a proper concept elicitation), they clearly show that further research is warranted to better understand what we are measuring.

Limitations

A limitation was the use of the BRUMS which was not developed for performing artists but still possesses acceptable psychometric properties (Jeffries et al., 2020) and additionally three of the BRUMS items have also been confirmed with factor analysis in the MTDS (Main & Grove, 2009). Another limitation, for the concept elicitation results, is asking participants to interpret two separate questions (fatigue and recovery). We may have obtained a different result had we only asked one question (e.g., fatigue only). Finally, although we examined associations with the single items of the BRUMS, this instrument provides a summary score reflecting the fatigue construct and the single items are not supposed to be used in isolation (contrary to the SRSS). However, for explorative purposes and to better understand the items under validation, correlations with the single BRUMS items were used as they all reflect the same, unidimensional, construct. We consider the present results as preliminary, since the goal was to examine construct validity and reliability, however future studies should confirm these results and examine the association with measures of external or preferably internal load, as further validation. However, without evidence of construct validity it would be illogical to explore associations with other constructs.

Conclusion

We provide preliminary evidence of construct validity, reliability and agreement for the single item of fatigue in dancers. The recovery item, despite acceptable reliability and agreement, was only partially confirmed in terms of construct validity, when using the SRSS recovery items as reference. Additionally, we explored the concepts on which respondents focused when providing a rating, with tiredness, muscle soreness and fatigue identified as the most common concepts/symptoms reported for fatigue and recovery. While these results are encouraging, further studies are needed to confirm their validity and to examine other measurement properties.

Acknowledgements

We would like to thank Dr Nico Kolokythas, the English National Ballet School, Juncal Roman Pastor, Dr Rachel Ward, the Dance Research Collaborative Australia, Dr Jason Lam, Alexander McKinven, the Royal Academy of Dance and Ashlea Mary Cohen for their generous help assisting with data collection for this research project.

Chapter Seven

General Discussion

Main findings

Dance is a physical discipline involving rigorous and repetitive training requiring both artistic and extreme athletic capacity which increases dancer's susceptibility to injury. However, detailed information about the nature of injury is lacking and illnesses prevalence and incidence completely absent in the literature. Similarly, it is unknown if training load may contribute to injury incidence or illness, as has been postulated in sport. Many methods are available to monitor the training process and its effects, however the quality and level of evidence of subjective measures, which are commonly used in sport, are not yet examined and within dance no validated measures exist for monitoring training effects. The validity of the measurements in essential otherwise the results of studies cannot be interpreted and risk of misclassification high. However, there is a need for a physical training framework which provides an overarching model that may help understand and guide the development, validation, implementation, and interpretation of measures used for athlete monitoring, particularly subjective measures. Therefore, a series of research studies were conducted to develop a greater understanding of health-related issues and measures in dancers with particular reference to subjective responses.

In summary, the main findings of *Study One* was that dancers experience high training loads relative to other sports and concomitantly high injury and illness incidence and risk. In *Study Two,* no validated measures exist for the most frequently used AROMs in sport science. *Study Three* provides an updated conceptual framework to help guide the future development, validation and implementation of measures used for athlete monitoring. *Study Four* was the validation of single items fatigue and recovery for measuring acute training affects in dancers. The main findings of each study are addressed in detail below.

Injury, illness, and training load in a professional contemporary dancers

Dance requires a multifaceted interaction of aesthetic, artistic and technical components (Kuno et al., 1996) and involves extreme physical demands together with a high level of athletic ability. Professional dancers routinely undertake long hours of training and high training loads, forming the foundation for the refinement and development of both technique and performance. Inappropriate balance between training load and recovery has been suggested to induce fatigue, abnormal training responses (Meeusen et al., 2013), and increased risk of injury and illness (Schwellnus et al., 2016). This thesis provides new insight into health-related issues affecting professional dancers. *Study One* provides a detailed description of injuries occurring in professional dancers. The main finding of *Study One* is that professional dance is associated with high injury and illness incidence rates. Our entire cohort of participants (100%) acquired at least two medical attention injuries during the one-year surveillance period, and several participants experienced more than seven new injuries. Whilst the medical attention injury rate was similar to previous reports, our time loss injury rate was higher (1.4 per 1000 hours) than in several evaluations of similar contemporary dance studies (0.16–0.22 per 1000 hours) (Bronner et al., 2018; Bronner & Wood, 2017). Within the one-year surveillance period there were 261 days for time loss injuries and 447 days of training (in a modified form) for medical attention injuries. Despite this, only four dancers missed performances due to new injuries and one due to reinjury. This highlights that despite the presence of injury or medical complaints, dancers persist in training and performance.

Despite overuse injuries being commonly cited as the most frequent type of injury in dancers (Bronner et al., 2003), in *Study One* acute injuries were responsible for 75% of all injury type. This difference may be due in part to the injury definitions used, as no consensus currently exists. For example, overuse injury may be considered as a mechanism of injury (Meeuwisse et al., 2007), or a diagnosis-based definition (Nilsson et al., 2001). An alternative explanation for the high percentage of acute injuries is the lack of recovery periods noted during the surveillance period of this study and potential resultant fatigue. It has been suggested acute fatigue may affect intrinsic modifiable risk factors altering injury risk profile in athletes (Verschueren et al., 2020). As such, using subjective measures to monitor the training effect in dance may be useful. However, the quality and level of evidence of subjective measures used in sport (and dance) are yet to be examined. Consequently, this formed the basis for *Study Two*.

Whilst ligaments and joints were most commonly injured, there was a variation across body region and tissue type. Overall, we identified a wide range of injury types, which should be considered when injury-prevention programs are developed or for examining causal associations in future etiologic studies. For example, we reported one metatarsal stress reaction–fracture, which is characteristic of excessive stress experienced by bone (Melvin, 1993) and was accordingly classified as an overuse injury. It has been proposed overuse injuries can be modelled as mechanical fatigue phenomena occurring when a tissue is exposed to repetitive cyclic loading exceeding reparability (Edwards, 2018). Although training load has been proposed as a factor contributing to overuse injury, there are many factors involved (e.g. energy deficiency through diet, disordered eating, low body mass index, hormonal imbalance, low bone mineral density) (Freslon et al., 2004; Romani, Gieck, Perrin, Saliba, & Kahler, 2002), some of which are common in dance (e.g. disordered eating) (Arcelus, Witcomb, & Mitchell, 2014). The variety of injury types and nature in this cohort indicates that even though the final mechanical causes are likely similar (stress and strain superior to the strength of the structure), the antecedent causes may be different. Therefore, in order to establish associations with predictive factors (e.g., training load) and given the unique responsiveness of tissues and structures, specific classifications are needed to examine associations with both specific injuries and tissues (Impellizzeri et al., 2020; Impellizzeri et al., 2020).

Another major finding of *Study One* was the high incidence of upper respiratory tract infections, with an illness incidence rate of 9.1 per 1000 hours and 134 illness episodes. The average duration of upper respiratory tract infections symptoms was 40 days, with a range from one to approximately 3 months. Furthermore, whilst dancers experienced illness symptoms, no modifications to training or time loss were reported. This indicates, dancers persist with training and performance despite the presence of illness symptoms. Given persistent illness can negatively impact health and performance, particularly when undertaking high levels of strenuous exercise (Gleeson & Pyne, 2016), this finding is very concerning from a health perspective and further highlights the high-risk nature of dance.

In recent years, many studies have examined the association between training load and injury in various sport disciplines (Eckard et al., 2018; Ehrmann et al., 2016; Timoteo et al., 2018). Indeed, the initial aim of this thesis was to examine associations of injury and illness with training load and subjective measures (such as wellness) in professional dancers (*Study One*). When the initial project was designed this was a common approach in research and therefore exploration of these associations formed the initial goal of the first study. Given the high injury rate in dance, this seemed a reasonable approach to injury reduction and a potential intervention strategy. However, most of these studies (Eckard et al., 2018; Ehrmann et al., 2016; Timoteo et al., 2018) are underpowered because the number of injuries required to indicate an association with training load is much higher than typically used (van Smeden et al., 2016). Whilst we reported a similar number of injury events (n=79) to previous studies (Ehrmann et al., 2016; Timoteo et al., 2018), this was still inadequate to reliably determine associations; hence, we did not perform any predictive or associative analyses. Apart from the above issues, there are many critical conceptual and methodological issues in current training load injury research (Impellizzeri et al., 2020; Impellizzeri et al., 2020). These include lack of conceptual framework, different measures of exposure, inappropriate use of ratios, unclear analyses, inadequate sample size, missing data, variety of injury definitions used, and poor standard and quality of reporting (Impellizzeri et al., 2020; Impellizzeri et al., 2020). These critical issues highlight why we excluded the analysis of training load from *Study One.* Additionally, there is no underlying rationale together with the possibility of increased researcher degrees of freedom leading to inflated rate of false-positive findings and over estimation of effect sizes. Indeed, we demonstrated this scenario in the results of *Study One.* As such, we concentrated on accurately describing the training and utilised categorisation methods such as individual and group percentiles (low, medium, and high).

An interesting finding from *Study One* was the injury and illness incidence rates in relation to training load were dependent on the categorization and reference method. For example, when based on group data, dancers in high training load group experienced the highest injury incidence rate (5.7 per 1000 hours) compared with the medium and low groups (3.6 and 3.2, respectively, per 1000 hours). In contrast, when using individual training loads, the injury incidence rate was opposite in direction to the group training load injury rate, with dancers in the highest training load category experiencing lower injury rates. Most previous authors have used group data to create training load categories. For example, a study participant is placed in the high, medium, or low category based on the training load completed by the whole group. However, no justification exists for using the group training load as a reference. Given the individual nature of injuries, it would be reasonable to use individual data for categorization. Additionally, in contemporary dance, the mechanism of injury is likely to be unique, given that choreography and repertoire are individualised for each dancer. Therefore, a clear association of training load level and injury rate was not present. Rather, we accurately describe the training load to help develop a conceptual framework and a causal structure for use in future epidemiologic studies that may require international and multicentre initiatives to obtain appropriate sample sizes.

Whilst the main aim of *Study One* was to contribute to the body of knowledge for identification of potential mechanisms and prognostic factors, from a practical point of view, our results suggest that a possible area of intervention is training load distribution by application of basic training principles. Indeed, few recovery periods occurred within and between macrocycles. It may be that professional dancers need to undertake high training loads to achieve optimal performance. Therefore, modifying the training load distribution and including recovery days and short periods of reduced training load are potential immediate interventions. As such, subjective measures such as fatigue and recovery are areas that could be easily monitored in dancers. As yet, no validated subjective measures for fatigue and recovery exist in dance and this formed the basis for *Study Four*.

The collective findings of *Study One* demonstrate professional dancers undertake very high training loads, with little recovery periods together with high injury and illness incidence rates. Dancers missed very few performances, despite injury and illness and kept training even when affected by medical attention injuries and illness. This is concerning from a health perspective and underlines the need for further studies to understand how to decrease the risk and develop preventive interventions and educational initiatives.

Athlete-Reported Outcome Measures for monitoring training responses

As mentioned earlier the original aim of this thesis was to examine associations of injury and illness and subjective measures (such as wellness) in professional dancers (*Study One*). When the initial project was designed this was a common approach in research and therefore exploration of these associations formed the initial goal of the *Study One*. However, in order to prove associations, potential explanatory or predictive variables need to be valid. As such we began to examine the validity of AROMs (*Study Two*).

Subjective measures are commonly used in sport science and have been reported to have better sensitivity and consistency than objective measures and provide a more accurate reflection of changes to acute and chronic training loads (Saw et al., 2016). Given the training process is athlete centred and the ultimate goal is understanding how athletes are tolerating training demands, subjective measures seem to provide the ability to measure constructs and dimensions that are not objectively measurable. However, any measure used in research and practice should be valid in order to provide accurate, meaningful, useful, clear, and specific information. These properties are especially critical when using these measures to develop models and understand phenomena. While a previous review (Saw et al., 2016) provided initial examination of athlete self-report measures, we extended and complemented this research with the first comprehensive assessment of AROM measurement properties.

The major finding of *Study Two* is that the most commonly used AROMs in sport science have not been validated, despite often being presented as validated. Additionally, these AROMs (e.g., wellness items) have often been modified from the original version (e.g., adding/removal of items or combining scores) without justification. In contrast, the measurement properties of multiple-item AROMs derived from psychometrics were acceptable (with the exclusion of content validity and measurement error). The single items plus the Hooper/wellness items (from which most of the single items are developed) accounted for 46.1% of all instruments used in previous studies. These items are also frequently used in practical settings and are included in athlete management systems of national organizations, commercial software, etc. Until proper validation studies are completed, all conclusions based on these AROMs are questionable. Indeed, the results of *Study Two* confirm we why excluded the analysis of these variables from *Study One*. Therefore, established reference methods, such as those of clinimetrics or psychometrics, should be used to develop and assess the validity of AROMs (*Study Four*).

Other than the major issue of lack of validation, there have been many modifications to the 'wellness' items, including the summation of a single score of the items known as the "Hooper index" or "Hooper score" of wellness. These adjustments have appeared without any attempt at validation and have been presented as arbitrary modifications in the methods section. To further confuse the issue, these "wellness" items have been referred to in different ways (e.g. Hooper questionnaire (Baklouti et al., 2017; Clemente et al., 2019; Mendes et al., 2018), subjective or self-reported wellness ratings (Gathercole et al., 2015; Govus et al., 2018; Hills & Rogerson, 2018), perceived ratings of wellness (Buchheit et al., 2013; Thorpe et al., 2016), perception of well-being (Matos et al., 2019; Noon et al., 2015). The lack of consistency in a definition of the construct that supposedly measures these items indicates the absence of a theoretical framework, thereby creating a lack of clarity regarding what the domains of interest are (and also why that domain was useful within the training process). Assuming each item measures one subdimension of a higher order construct, it is, however, not clear what is the reference higher order construct. Consequently, it became evident there was a lack of a reference framework on which subjective measures are based (*Study Three*). The COSMIN standards indicate items should be "consistent with the theory, conceptual framework or disease model that was used to define the construct of interest." It also means that the relevance of the items within the training process should also be presented in such a way that specific hypotheses (e.g., directions and strength of the associations) can be generated and tested (*Study Four*).

Monitoring training responses with subjective measures is common and many instruments are available. *Study Two*, provides a unique and comprehensive appraisal of the risk of bias and quality of the psychometric (measurement) properties of the most frequently used AROMs according to established guidelines. Categorizing the results of the review in terms of "traditional" psychometric instruments and sport science generated questionnaires (e.g., single items), we conclude that while some of the most commonly used psychometric instruments do not show adequate quality in all measurement properties, most psychometric attributes have at least been examined. In order to effectively evaluate health outcomes in dancers, there is a need for validated items, specific to the population under investigation together with a reference framework (*Study Three*). This provides a basis to propose exactly what and why specific variables (objective or subjective) are measured and thus act as a reference for validation studies (*Study Four*).

Development of a revised conceptual framework of physical training

Based on the results of *Study Two*, it was evident there was an absence of a reference framework presenting hypothesized relations between constructs (e.g., 'wellness') and other measures (e.g., training load), making the examination of the validity and usefulness of these items introduced in the sport science literature confusing and prone to bias. Whilst theoretical frameworks and models for physical training are available, we identified they did not include all the components necessary to provide a suitable reference for validation studies of subjective measures. As such, we re-examined available frameworks and proposed a conceptual framework that can facilitate the validation of physical training measures.

Based on an examination of the literature, in *Study Three* we identified four concepts that could be integrated or extended that are important in the conceptualization of the process and outcomes of physical training. This updated conceptual framework is intended to facilitate the validation and interpretation of physical training measures, with particular reference to subjective measures. The main concept was the formal introduction of a new measurable component 'training effects', a higher order construct resulting from the combined effect of four possible responses (acute and chronic, positive and negative). In previous frameworks (Impellizzeri et al., 2019; Impellizzeri et al., 2005) this concept was mixed with the generic concept of 'adaptations' and training outcome. However, it did not constitute a formal and defined construct of the framework. Potentially, the lack of the training effect construct may have contributed to confusion, where measures of effects occurring after the training session have been misinterpreted as measures of internal training load (Halson, 2014), which is by definition the psychobiological response during the exercise(s) constituting the training session. The second aspect we formally clarified as training effect measures. The training effect is a construct for which there is no gold standard measure however it can be assessed in several ways, including using proxy measures when the training effect being considered cannot be directly quantified. This is an important clarification since the validation process may also refer to proxy measures, e.g., whether measures are adequate reflections of the construct of interest. The third area we further articulate is sport performance outcomes. Previous frameworks used the term 'training outcomes' which does not properly differentiate between the chronic training effect and sport performance outcomes (Impellizzeri et al., 2019; Impellizzeri et al., 2005). In the current conceptual framework, these two constructs are differentiated and sport performance outcomes further elaborated by including the potential range, i.e., from improvement to overtraining. Finally, we expanded on the concept of individual and contextual factors (e.g., training status, health, nutrition and environmental factors). In the previous frameworks (Impellizzeri et al., 2019; Impellizzeri et al., 2005), contextual factors were only mentioned as aspects influencing internal load and thereby training outcomes. Here, we clarify that modifiable and non-modifiable individual/contextual factors have a more widely integrated relationship with all components of our conceptual framework, including bi-directionality with training effects. In addition, we provided examples of theoretical and practical applications of the conceptual framework such as validation and conceptualization of constructs (e.g., performance readiness), and understanding of higher order constructs, such as training tolerance when monitoring training to adapt to individual responses and effects. This conceptual framework, together with operational definitions of the constructs, provides an overarching model that may help understand and guide the development, validation, implementation, and interpretation of measures used for athlete monitoring. The development of this framework (*Study Three*) formed the foundation for *Study Four*, validation of health related (single item) measures fatigue and recovery in dancers.

Development and validation of single items for fatigue and recovery: preliminary results in dancers

Athlete Reported Outcome Measures are commonly used in practice and research for measurement of the training effects. Whilst there are many AROMs available, the most commonly used are the ‘wellness items’, which have not been properly validated, despite often being presented as so (*Study Two*). As such, there is a critical need to develop and examine the validity of instruments by adopting appropriate methodology using established reference methods when developing AROMs. The COSMIN is an established reference set of recommendations that can guide the development and validation of measures aimed to assess symptoms and physical signs. The foundation for an appropriate methodological approach, is the necessity of a conceptual framework to define the construct of interest (*Study Three*). Furthermore, development of a framework permits the hypothesis of relationships between constructs, allowing the examination of validity (*Study Four and Study Five ongoing*).

The primary findings of *Study Four* confirm the construct validity of the single items, fatigue and recovery, in dance using established methodological guidelines, together with a reference and explicit framework (*Study Three*). To our knowledge this is the first study examining the validity of single items used in dance or in sport supported by a reference framework. Only a multiple item instrument specifically developed for athletes have used a reference framework, i.e., the REST-Q and derivatives. However, the framework is only mentioned and explained in the manual (Kellmann & Kallus, 2001), which is unfortunate, since the validity process and methodology are driven by a conceptual framework. The aim of our study was to develop and validate two single items as measures of acute training effects that could be used to assess training effects. We refer to these results as preliminary since the goal was to examine construct validity and reliability, so that future studies can examine the association with measures of external or preferably internal load, as further validation. Overall, the percentages of agreement for the single items were high (>80% level of agreement), indicating raters often choose the same response option. The Kappa weighted coefficients for the single items ranged from substantial to excellent depending on the reference benchmark (Cohen, 1968; Fleiss, 1971; Landis & Koch, 1977) (0.71 to 0.78) suggesting high interrater reliability.

An interesting finding of *Study Four* was the main concepts that emerged for fatigue were 'tiredness' (34.7%), muscle soreness (17.3%) and energy (13%). Interestingly, one of the main concepts that emerged for fatigue was 'fatigue' (26%). Whilst we removed this result from the analysis given the cyclic nature (i.e., counter response) of the response, this may highlight that the concept of 'fatigue' is ambiguous, lacks clarity or is confusing in this population. This is not surprising given the plethora of definitions of fatigue available in the literature (Coutts et al., 2017; Edwards, 1983; Lewis & Wessely, 1992). Furthermore, the main concept identified for recovery was muscle soreness (43%), with tiredness and fatigue respectively (27%, 24%). Given that the response of 'tiredness' and 'fatigue' were identified by participants for both fatigue and recovery, this may potentially indicate a conceptual overlap of these constructs. Further research is warranted investigating concept elicitation and conceptual overlap or distinctiveness, of these constructs (fatigue and recovery) within dancer and athlete monitoring.

To extend and complement the research of *Study Four*, in *Study Five (ongoing)* we will further explore content validity, specifically the relevance and clarity of commonly used wellness and readiness to train items currently used in sport. Using qualitative methods, we will aim to identify, describe and substantiate measurement concepts that are important and relevance to the target population (dancers) and inform the content of these AROMs. Given content validity is the most important measurement property of any AROM or patient reported outcome measure, which was highlighted in the results of *Study Two*, we hope that the results of *Study Five* will establish if the current wellness and readiness to train items, are relevant and clear to dancers and athletes.

Limitations

Unfortunately, it was not feasible to conduct *Study Two* prior to *Study One*, as the dance injury, illness and training load study (*Study One*) was a two-year data collection process and commenced almost immediately at the start of this thesis due to timing and compatibility issues with the dance company. Consequently, we discarded several of the variables that were initially going to form part of the analysis (e.g., training load and wellness). Additionally, due to time constraints with thesis preparation, we have analysed only one year of data from *Study One*, however we plan to publish the second year of data as a further study. Additionally, *Study Two* took almost 15 months to complete due to the large amount of articles involved. Furthermore, it was only once we were in the midst of *Study Two*, we identified the need for an updated conceptual framework that would necessitate the validation of health measures in dancers with particular reference to subjective measures. Another limitation is the small sample sizes in *Study One* in which we were restricted to the number of participants in the dance company. Accordingly, our goal was descriptive, and associations between training load and illness and injury explored to show the dependence of the results on the categorization criteria. However, in the future we hope to conduct further investigations, possibly as international, multicentre collaborative studies.

A major limitation of *Study Four* was the use of the BRUM scale, which in *Study Two* was identified as having poor content validity. However, as there was insufficient time (due to time constraints of thesis submission) to develop a new multiple item instrument, we selected the fatigue subscale of the BRUMs as it is one of the most commonly used. Additionally, three of the BRUMs items have also been confirmed with factor analysis in the Multi-Component Training Distress Scale (Main & Grove, 2009). This provides additional support, despite the lack of content validity. Nevertheless, a future study may further examine the content validity of single items such as the relevance of concepts included in the wellness items from final user perspective (*Study Five ongoing*). Another limitation of *Study Four* was due to the COVID-19 pandemic in which conducting qualitative in person interviews (that are typical with concept elicitation research) was not possible. Therefore, we provide preliminary information that may be further explored in future research. Additionally, due to the COVID-19 situation the unusual training condition of dancers may have influenced the results.

Practical implications

The findings of this thesis have identified practical recommendations regarding health outcomes and measures for dancers, together with a revised conceptual framework that is intended to facilitate the validation and interpretation of physical training measures, with particular reference to subjective measures:

- Professional dancers experience high training loads and concomitantly high injury and illness incidence and risk. Therefore, modification of the training distribution, including recovery days, and reducing load for short periods are potential areas of intervention together with injury and illness surveillance systems.
- Dancers continue with training, albeit modified, even when affected by medical-attention injuries and illness. Consequently, for optimal health outcomes practitioners may need to carefully monitor the illness incidence, actively modify training load accordingly or prescribe complete rest.
- Inconsistent results were present for the incidence of injury and illness based on individual and group categorizations of training load. Therefore, future

investigation is warranted to develop a conceptual framework that may explain particular causal phenomena and provide a basis for interpreting results.

- No validation studies could be found for single item AROMs (including the so-called wellness, wellbeing, welfare, etc. items). As such, all conclusions based on these AROMs are questionable. There is an urgent need to adopt proper methodology using established reference methods when developing AROMs such as the COSMIN guidelines.
- For the development, validation, implementation, and interpretation of measures (particularly subjective) used in athlete monitoring should be based on their fit within a conceptual framework.
- For monitoring acute or chronic training effects in dancers there is preliminary evidence of construct validity, reliability, and agreement for the single item of fatigue. Despite acceptable reliability and agreement, the single item of recovery was only partially confirmed in terms of construct validity.

Chapter Eight

Summary and Future Directions

Summary

Optimal performance in dance requires a complex interaction of aesthetic, artistic and technical components (Kuno, Fukunaga, Hirano, & Miyashita, 1996), involves extreme physical demands and a high level of athletic ability. Whilst high injury rates have been reported there is often inadequate detail reported in injury typology that is necessary to investigate risk factors and provide causal pathways for injury. Little is known about health outcomes such as illness or its relationship with the training demands of professional contemporary dancers. Additionally, valid measures for monitoring subjective responses appear scarce in sport science and even more so in dance. This thesis aimed to investigate health outcomes and measures in professional contemporary dancers through a series of studies; a longitudinal study with a detailed description of injury, illness and training load (*Study One*); a systematic review identifying the most commonly used AROMs in sport for monitoring training responses together with an assessment of the risk of bias, measurement properties, and level of evidence, based the COSMIN guidelines (*Study Two*); an updated conceptual framework intended to facilitate the validation and interpretation of physical training measures, with particularly reference to subjective measures (*Study Three*); the validation of single items measures fatigue and recovery for monitoring training effects in dancers (*Study Four*).

The findings of this thesis contribute to an improved understanding of health outcomes and measures in professional dance. The results of this thesis suggest professional dance is associated with high injury and illness incidence rates together with high training loads. In addition, dancers continue to train, albeit with modifications, even when affected by medical attention injury or illness. Furthermore, there are very few recovery periods occur within and between macrocycles. Collectively, this is concerning from a health perspective and highlights the need for further studies to understand how to decrease the

risk. The results of *Study One* also highlight that injury and illness in relation to training load will be different dependant on the categorization method, and although a common method of analysis in sport science, confirmed the decision to exclude training load variables from our analysis. Additionally, the decision to exclude 'wellness' variables from *Study One* was confirmed in the results of *Study Two* which showed single item AROMs commonly used in sport have not been validated, suggesting all conclusions based on these AROMs are questionable. In the preparation of *Study Four*, we provided an updated conceptual framework intended to facilitate the validation and interpretation of physical training measures (*Study Three*). *Study Four* results provide preliminary evidence confirming the construct validity, reliability and agreement for the single item of fatigue in dancers. The recovery item, despite acceptable reliability and agreement, was only partially confirmed in terms of construct validity, when using the SRSS recovery items as reference. The results of this study allow for the measurement of fatigue, providing a validated and scientific platform for the future evaluation of health outcomes in dancers. Collectively, the findings of this thesis provide novel information concerning health issues in dancers together with a conceptual framework that may help validate measures of the physical training process, particularly subjective measures in future studies.

Future directions

To expand upon the findings of this thesis and improve health outcomes of professional dancers and the use of subjective measures for monitoring training effects, it is recommended further research investigate:

- The results of *Study Two* highlighted that single items (e.g., wellness) AROMs used in sport science have not been validated. Additionally, multiple item AROMs have inadequate content validity and measurement error. Therefore, further research is warranted to validate and improve the methodological properties of these AROMs. As such, *Study Five* (ongoing) we are examining the relevance of concepts (content validity) included in the wellness items from a final user perspective in dance.
- Professional dancers experience high training loads relative to other athletes and concomitantly high injury and illness incidences and risks. Therefore, the development of preventive interventions and educational initiatives is suggested to improve health outcomes in professional dancers. Additionally, this could also be extended to amateur or pre professional dancers.
- The aim of *Study One* was to provide details on the nature and incidence of injury and preliminary information on illness occurrence. These results together with the training loads experienced by the dancers can be used to develop and hypothesize an etiologic framework and hence a causal structure (e.g., directed acyclic graph). Alternatively, other than a conceptual framework to develop a causal structure, multicentre studies or prediction models for risk stratification (that do not require causation) are also warranted.

- Application of the updated conceptual framework (*Study Three*) to test the hypothetical examples we presented for items measuring subjective training effects (e.g., wellness items) that may be related to performance readiness.
- In *Study Four* we confirmed the construct validity and reliability of the single item fatigue and partially confirmed the single item recovery in terms of construct validity, when using the SRSS recovery items as reference. Therefore, future studies may utilise the single item fatigue in the assessment of acute training affects. However, further research is warranted confirming the construct validity of the recovery item.
- Finally, use of the updated conceptual framework may guide the development, implementation, and evaluation of a more comprehensive and transparent approach to dancer monitoring and validation of measures. Future research in dance could examine the relationships between training effects (acute and chronic), performance outcomes, and individual and contextual factors, together with the interconnections of these concepts within the training process.

Chapter Nine

References

Abbott, W., Brownlee, T. E., Harper, L. D., Naughton, R. J., & Clifford, T. (2018). The independent effects of match location, match result and the quality of opposition on subjective wellbeing in under 23 soccer players: A case study. *Research in Sports Medicine (Print), 26*(3), 262-275. doi:10.1080/15438627.2018.1447476

Ade, J. D., Drust, B., Morgan, O. J., & Bradley, P. S. (2020). Physiological characteristics and acute fatigue associated with position-specific speed endurance soccer drills: Production vs maintenance training. *Science & Medicine in Football*, 1-12. doi:10.1080/24733938.2020.1789202

Ahmun, R., McCaig, S., Tallent, J., Williams, S., & Gabbett, T. (2019). Association of daily workload, wellness, and injury and illness during tours in international cricketers. *International Journal of Sports Physiology & Performance, 14*(3), 369-377.

Akenhead, R., Marques, J. B., & Paul, D. J. (2017). Accelerometer load: A new way to measure fatigue during repeated sprint training? *Science & Medicine in Football, 1*(2), 151-156. doi:10.1080/24733938.2017.1330550

Allen, N., Nevill, A., Brooks, J., Koutedakis, Y., & Wyon, M. (2012). Ballet injuries: Injury incidence and severity over 1 year. *Journal of Orthopaedic and Sports Physical Therapy, 42*(9), 781-790. doi:10.2519/jospt.2012.3893

Almeida, F. R., Silva, C. B., Vendramini, C. M., Campos, C., & Brandão, M. (2017). Psychometric properties of the brazilian version of the brums scale for basketball referees. *Revista de Psicologia del Deporte, 27*, 37-44.

Alonso, J. M., Tscholl, P. M., Engebretsen, L., Mountjoy, M., Dvorak, J., & Junge, A. (2010). Occurrence of injuries and illnesses during the 2009 iaaf world athletics championships. *British Journal of Sports Medicine, 44*(15), 1100-1105. doi:10.1136/bjsm.2010.078030

Altman, D. G. (1991). *Practical statistics for medical research* (6th ed.). London: Chapman and Hall.

American College of Sports Medicine. (2002). *Acsm's guidelines for exercise testing and prescription. 8th ed*. Philadelphia, PA, USA: Lippincott Williams & Wilkins.

American Educational Research Association, A. P. A., National Council on Measurement in Education. (2014). Standards for educational & psychological testing. Washington, DC.

Andrade, E., Arce, C., de Francisco, C., Torrado, J., & Garrido, J. (2013). Versión breve en español del cuestionario poms para deportistas adultos y población general = abbreviated version in spanish of the poms questionnaire for adult athletes and general population. *Revista de Psicología del Deporte, 22*(1), 95-102.

Andrade, E., Arce, C., Torrado, J., Garrido, J., De Francisco, C., & Arce, I. (2010). Factor structure and invariance of the poms mood state questionnaire in spanish. *Spanish Journal of Psychology, 13*(1), 444-452. doi:10.1017/s1138741600003991

Andrade, E., & Rodriguez, D. (2018). Factor structure of mood over time frames and circumstances of measurement: Two studies on the profile of mood states questionnaire. *PLoS One, 13*(10), e0205892. doi:10.1371/journal.pone.0205892

Angioi, M., Metsios, G., Koutedakis, Y., Twitchett, E., & Wyon, M. (2009). Physical fitness and severity of injuries in contemporary dance. *Medical Problems of Performing Artists, 24*(1), 26-29.

Anglem, N., Lucas, S. J., Rose, E. A., & Cotter, J. D. (2008). Mood, illness and injury responses and recovery with adventure racing. *Wilderness & Environmental Medicine, 19*(1), 30-38. doi:10.1580/07-WEME-OR-091.1

Arce, C., De Francisco, C., Andrade, E., Seoane, G., & Raedeke, T. (2012). Adaptation of the athlete burnout questionnaire in a spanish sample of athletes. *Spanish Journal of Psychology, 15*(3), 1529-1536. doi:10.5209/rev_sjop.2012.v15.n3.39437

Arcelus, J., Witcomb, G. L., & Mitchell, A. (2014). Prevalence of eating disorders amongst dancers: A systemic review and meta-analysis. *European Eating Disorders Review, 22*(2), 92-101. doi:10.1002/erv.2271

Atlaoui, D., Duclos, M., Gouarne, C., Lacoste, L., Barale, F., & Chatard, J. C. (2004). The 24-h urinary cortisol/cortisone ratio for monitoring training in elite swimmers. *Medicine &*

Science in Sports & Exercise, 36(2), 218-224. doi:10.1249/01.MSS.0000113481.03944.06
Australian Institute of Sport. (2020). *Prescription of training load in relation to loading and unloading phases of training (2nd ed)*. Retrieved from Bruce, ACT, Australia:
Baar, K., & McGee, S. (2008). Optimizing training adaptations by manipulating glycogen. *European Journal of Sport Science, 8*(2), 97-106. doi:10.1080/17461390801919094
Baird, M. F., Graham, S. M., Baker, J. S., & Bickerstaff, G. F. (2012). Creatine-kinase- and exercise-related muscle damage implications for muscle performance and recovery. *Journal of Nutrition and Metabolism, 2012*, 960363. doi:10.1155/2012/960363
Baklouti, H., Rejeb, N., Aloui, A., Jaafar, H., Ammar, A., Chtourou, H., . . . Souissi, N. (2017). Short versus long small-sided game training during ramadan in soccer players. *Physical Therapy in Sport, 24*, 20-25. doi:10.1016/j.ptsp.2016.10.002
Bartlett, J. D., Hawley, J. A., & Morton, J. P. (2015). Carbohydrate availability and exercise training adaptation: Too much of a good thing? *European Journal of Sport Science., 15*(1), 3-12. doi:10.1080/17461391.2014.920926
Beck, K. L., Mitchell, S., Foskett, A., Conlon, C. A., & von Hurst, P. R. (2015). Dietary intake, anthropometric characteristics, and iron and vitamin d status of female adolescent ballet dancers living in new zealand. *International Journal of Sport Nutrition & Exercise Metabolism, 25*(4), 335-343. doi:10.1123/ijsnem.2014-0089
Beck, K. L., Thomson, J. S., Swift, R. J., & von Hurst, P. R. (2015). Role of nutrition in performance enhancement and postexercise recovery. *Open Access Journal of Sports Medicine, 6*, 259-267. doi:10.2147/OAJSM.S33605
Behm, D. G., Young, J. D., Whitten, J. H. D., Reid, J. C., Quigley, P. J., Low, J., . . . Granacher, U. (2017). Effectiveness of traditional strength vs. Power training on muscle strength, power and speed with youth: A systematic review and meta-analysis. *Frontiers in Physiology, 8*, 423. doi:10.3389/fphys.2017.00423
Bennell, K. L., Malcolm, S. A., Thomas, S. A., Reid, S. J., Brukner, P. D., Ebeling, P. R., & Wark, J. D. (1996). Risk factors for stress fractures in track and field athletes: A twelve-month prospective study. *American Journal of Sports Medicine, 24*(6), 810-818.
Benson, J. E., Geiger, C. J., Eiserman, P. A., & Wardlaw, G. M. (1989). Relationship between nutrient intake, body mass index, menstrual function, and ballet injury. *Journal of the American Dietetic Association 89*(1), 58-63.
Berriel, G. P., Costa, R. R., da Silva, E. S., Schons, P., de Vargas, G. D., Peyré-Tartaruga, L. A., & Kruel, L. F. M. (2020). Stress and recovery perception, creatine kinase levels, and performance parameters of male volleyball athletes in a preseason for a championship. *Sports Medicine, 6*(1), 26-26. doi:10.1186/s40798-020-00255-w
Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices*. South Florida, USA: Global Text Project.
Bize, R., Johnson, J. A., & Plotnikoff, R. C. (2007). Physical activity level and health-related quality of life in the general adult population: A systematic review. *Preventive Medicine, 45*(6), 401-415. doi:10.1016/j.ypmed.2007.07.017
Blanch, P., & Gabbett, T. J. (2016). Has the athlete trained enough to return to play safely? The acute:Chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury. *British Journal of Sports Medicine, 50*(8), 471-475. doi:10.1136/bjsports-2015-095445
Boeding, J. R. E., Visser, E., Meuffels, D. E., & de Vos, R. J. (2019). Is training load associated with symptoms of overuse injury in dancers? A prospective observational study. *Journal of Dance Medicine & Science, 23*(1), 11-16. doi:10.12678/1089-313X.23.1.11
Borg, G., Ljunggren, G., & Ceci, R. (1985). The increase of perceived exertion, aches and pain in the legs, heart rate and blood lactate during exercise on a bicycle ergometer. *European Journal of Applied Physiology and Occupational Physiology, 54*(4), 343-349.
Boukhris, O., Trabelsi, K., Shephard, R. J., Hsouna, H., Abdessalem, R., Chtourou, L., . . . Chtourou, H. (2019). Sleep patterns, alertness, dietary intake, muscle soreness, fatigue,

and mental stress recorded before, during and after ramadan observance. *Sports (Basel), 7*(5). doi:10.3390/sports7050118

Bourgeois, A., LeUnes, A. D., & Meyers, M. (2010). Full-scale and short-form of the profile of mood states: A factor analytic comparision. *Journal of Sport Behavior, 33*(4), 355-376.

Bowen, L., Gross, A. S., Gimpel, M., Bruce-Low, S., & Li, F. X. (2020). Spikes in acute:Chronic workload ratio (acwr) associated with a 5-7 times greater injury rate in english premier league football players: A comprehensive 3-year study. *British Journal of Sports Medicine, 54*(12), 731-738. doi:10.1136/bjsports-2018-099422

Bowen, L., Gross, A. S., Gimpel, M., & Li, F. X. (2017). Accumulated workloads and the acute:Chronic workload ratio relate to injury risk in elite youth football players. *British Journal of Sports Medicine, 51*(5), 452-459. doi:10.1136/bjsports-2015-095820

Brito, J., Hertzog, M., & Nassis, G. P. (2016). Do match-related contextual variables influence training load in highly trained soccer players? *Journal of Strength & Conditioning Research, 30*(2), 393-399. doi:10.1519/JSC.0000000000001113

Bronner, S., McBride, C., & Gill, A. (2018). Musculoskeletal injuries in professional modern dancers: A prospective cohort study of 15 years. *Journal of Sports Sciences, 36*(16), 1880-1888. doi:10.1080/02640414.2018.1423860

Bronner, S., Ojofeitimi, S., & Rose, D. (2003). Injuries in a modern dance company: Effect of comprehensive management on injury incidence and time loss. *American Journal of Sports Medicine, 31*(3), 365-373.

Bronner, S., Ojofeitimi, S., & Spriggs, J. (2003). Occupational musculoskeletal disorders in dancers. *Physical Therapy Reviews, 8*(2), 57-68. doi:10.1179/108331903225002416

Bronner, S., & Wood, L. (2017). Impact of touring, performance schedule, and definitions on 1-year injury rates in a modern dance company. *Journal of Sports Sciences, 35*(21), 2093-2104. doi:10.1080/02640414.2016.1255772

Brooks, J. H., Fuller, C. W., Kemp, S. P., & Reddin, D. B. (2008). An assessment of training volume in professional rugby union and its impact on the incidence, severity, and nature of match and training injuries. *Journal of Sports Sciences, 26*(8), 863-873. doi:10.1080/02640410701832209

Buchheit, M., Cholley, Y., & Lambert, P. (2016). Psychometric and physiological responses to a preseason competitive camp in the heat with a 6-hour time difference in elite soccer players. *International Journal of Sports Physiology & Performance, 11*(2), 176-181.

Buchheit, M., Racinais, S., Bilsborough, J. C., Bourdon, P. C., Voss, S. C., Hocking, J., . . . Coutts, A. J. (2013). Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport 16*(6), 550-555. doi:10.1016/j.jsams.2012.12.003

Buchheit, M., Simpson, B. M., Schmidt, W. F., Aughey, R. J., Soria, R., Hunt, R. A., . . . Bourdon, P. C. (2013). Predicting sickness during a 2-week soccer camp at 3600 m (isa3600). *British Journal of Sports Medicine*, i124-i127. doi:10.1136/bjsports-2013-092757

Bühlmayer, L., Birrer, D., Röthlin, P., Faude, O., & Donath, L. (2017). Effects of mindfulness practice on performance-relevant parameters and performance outcomes in sports: A meta-analytical review. *Sports Medicine, 47*(11), 2309-2321. doi:10.1007/s40279-017-0752-9

Burtscher, M., Niedermeier, M., Burtscher, J., Pesta, D., Suchy, J., & Strasser, B. (2018). Preparation for endurance competitions at altitude: Physiological, psychological, dietary and coaching aspects. A narrative review. *Frontiers in Physiology, 9*, 1504. doi:10.3389/fphys.2018.01504

Cahill, B. R., & Griffith, E. H. (1978). Effect of preseason conditioning on the incidence and severity of high school football knee injuries. *American Journal of Sports Medicine, 6*(4), 180-184. doi:10.1177/036354657800600406

Calvert, T. W., Banister, E. W., Savage, M. V., & Bach, T. (1976). A systems model of the effects of training on physical performance. *IEEE Transactions on Systems, Man & Cybernetics, 6*(2), 94-102. doi:10.1109/TSMC.1976.5409179.

Camp, W. G. (2001). Formulating and evaluating theoretical frameworks for career and technical education research. *Journal of Vocational Education & Training, 26*(1), 27-39.
Campbell, R. S., Lehr, M. E., Livingston, A., McCurdy, M., & Ware, J. K. (2019). Intrinsic modifiable risk factors in ballet dancers: Applying evidence based practice principles to enhance clinical applications. *Physical Therapy in Sport, 38*, 106-114.
Cañadas, E., Monleón, C., Sanchis, C., Fargueta, M., & Blasco, E. (2017). Spanish validation of brums in sporting and non-sporting populations. *European Journal of Human Movement, 38*, 105-117.
Carey, D. L., Blanch, P., Ong, K. L., Crossley, K. M., Crow, J., & Morris, M. E. (2017). Training loads and injury risk in australian football-differing acute: Chronic workload ratios influence match injury risk. *British Journal of Sports Medcine, 51*(16), 1215-1220. doi:10.1136/bjsports-2016-096309
Caris, A. V., & Santos, R. V. T. (2019). Performance and altitude: Ways that nutrition can help. *Nutrition, 60*, 35-40. doi:10.1016/j.nut.2018.09.030
Chamari, K., Haddad, M., Wong, D. P., Dellal, A., & Chaouachi, A. (2012). Injury rates in professional soccer players during ramadan. *Journal of Sports Sciences, 30*(Supp 1), S93-S102. doi:10.1080/02640414.2012.696674
Chatard, J. C., Atlaoui, D., Pichot, V., Gourne, C., Duclos, M., & Guezennec, Y. C. (2003). Suivi de l'entrainement de nageurs de haut niveau par questionnaire de fatigue, dosages hormonaux et variabilite de la frequence cardiaque. / training follow up by questionnaire fatigue, hormones and heart rate variability measurements. *Science & Sports, 18*(6), 302-304.
Cheung, K., Hume, P., & Maxwell, L. (2003). Delayed onset muscle soreness: Treatment strategies and performance factors. *Sports Medicine, 33*, 145-164.
Chiu, Y. H., Lu, F. J., Lin, J. H., Nien, C. L., Hsu, Y. W., & Liu, H. Y. (2016). Psychometric properties of the perceived stress scale (pss): Measurement invariance between athletes and non-athletes and construct validity. *PeerJ, 4*, e2790. doi:10.7717/peerj.2790
Chtourou, H., Hammouda, O., Souissi, H., Chamari, K., Chaouachi, A., & Souissi, N. (2011). The effect of ramadan fasting on physical performances, mood state and perceived exertion in young footballers. *Asian Journal of Sports Medicine, 2*(3), 177-185.
Clarke, D., & Skiba, P. (2013). Rationale and resources for teaching the mathematical modeling of athletic training and performance. *Advances in Physiology Education, 37 2*, 134-152.
Clarke, N., Farthing, J. P., Lanovaz, J. L., & Krentz, J. R. (2015). Direct and indirect measurement of neuromuscular fatigue in canadian football players. *Applied Physiology, Nutrition, and Metabolism, 40*(5), 464-473. doi:10.1139/apnm-2014-0465
Clemente, F. M., Martinho, R., Calvete, F., & Mendes, B. (2019). Training load and well-being status variations of elite futsal players across a full season: Comparisons between normal and congested weeks. *Physiology & Behavior, 201*, 123-129. doi:10.1016/j.physbeh.2019.01.001
Clemente, F. M., Mendes, B., Bredt, S., Praca, G. M., Silverio, A., Carrico, S., & Duarte, E. (2019). Perceived training load, muscle soreness, stress, fatigue, and sleep quality in professional basketball: A full season study. *Journal of Human Kinetics, 67*, 199-207. doi:10.2478/hukin-2019-0002
Clemente, F. M., Mendes, B., Palao, J. M., Silverio, A., Carrico, S., Calvete, F., & Nakamura, F. Y. (2019). Seasonal player wellness and its longitudinal association with internal training load: Study in elite volleyball. *Journal of Sports Medicine and Physical Fitness, 59*(3), 345-351. doi:10.23736/S0022-4707.18.08312-3
Clemente, F. M., Oliveira, H., Vaz, T., Carrico, S., Calvete, F., & Mendes, B. (2019). Variations of perceived load and well-being between normal and congested weeks in elite case study handball team. *Research in Sports Medicine, 27*(3), 412-423. doi:10.1080/15438627.2018.1530998
Coffey, V. G., & Hawley, J. A. (2007). The molecular bases of training adaptation. *Sports Medicine, 37*(9), 737-763.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi:10.1177/001316446002000104
Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220. doi:10.1037/h0026256
Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385-396.
Colby, M. J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., Drew, M. K., . . . Lester, L. (2017). Multivariate modelling of subjective and objective monitoring data improve the detection of non-contact injury risk in elite australian footballers. *Journal of Science & Medicine in Sport, 20*(12), 1068-1074. doi:10.1016/j.jsams.2017.05.010
Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., & Poole, C. (2009). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology, 39*(2), 417-420. doi:10.1093/ije/dyp334
Coles, E., Wells, M., Maxwell, M., Harris, F. M., Anderson, J., Gray, N. M., . . . MacGillivray, S. (2017). The influence of contextual factors on healthcare quality improvement initiatives: What works, for whom and in what setting? Protocol for a realist review. *Systematic Reviews, 6*(1), 168-168. doi:10.1186/s13643-017-0566-8
Cormack, S., & Coutts, A. (2016). Monitoring training load. In D. Joyce & D. Lewindon (Eds.), *Sports injury prevention and rehabilitation: Integrating medicine and science for performance solutions.* Abingdon, Oxon ; New York, NY: Routledge.
Coutts, A. J., & Cormack, S. (2014). *Monitoring the training response* Champaign, Illinois: Human Kinetics
Coutts, A. J., Crowcroft, S., & Kempton, T. (2017). Developing athlete monitoring systems: Theoretical basis and practical applications. In M. Kellmann & J. Beckman (Eds.), *Sport, recovery, and performance: Interdisciplinary insights*: Taylor & Francis.
Coutts, A. J., Slattery, K. M., & Wallace, L. K. (2007). Practical tests for monitoring performance, fatigue and recovery in triathletes. *Journal of Science and Medicine in Sport, 10*(6), 372-381. doi:10.1016/j.jsams.2007.02.007
Cox, A. J., Gleeson, M., Pyne, D. B., Callister, R., Hopkins, W. G., & Fricker, P. A. (2008). Clinical and laboratory evaluation of upper respiratory symptoms in elite athletes. *Clinical Journal of Sport Medicine, 18*(5), 438-445. doi:10.1097/JSM.0b013e318181e501
Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perception of burden. *Social Science Computer Review, 19*, 146-162.
Cresswell, S. L., & Eklund, R. C. (2006). The convergent and discriminant validity of burnout measures in sport: A multi-trait/multi-method analysis. *Journal of Sports Sciences, 24*(2), 209-220. doi:10.1080/02640410500131431
Crookshanks, D. (1999). *Safe dance report iii: The occurrence of injury in the australian professional dance population.* Retrieved from Canberra, Australia:
Crowcroft, S., McCleave, E., Slattery, K., & Coutts, A. J. (2017). Assessing the measurement sensitivity and diagnostic characteristics of athlete-monitoring tools in national swimmers. *International Journal of Sports Physiology and Performance, 12*, S295-S2100. doi:10.1123/ijspp.2016-0406
Cullen, B. D., McCarren, A. L., & Malone, S. (2020). Ecological validity of self-reported wellness measures to assess pre-training and pre-competition preparedness within elite gaelic football. *Journal of Sport Science & Health*. doi:10.1007/s11332-020-00667-x
Cullen, T. O. M., Thomas, A. W., Webb, R., Phillips, T., & Hughes, M. G. (2017). Sil-6r is related to weekly training mileage and psychological well-being in athletes. *Medicine & Science in Sports & Exercise, 49*(6), 1176-1183. doi:10.1249/MSS.0000000000001210
Cummins, C., Welch, M., Inkster, B., Cupples, B., Weaving, D., Jones, B., . . . Murphy, A. (2019). Modelling the relationships between volume, intensity and injury-risk in professional rugby league players. *Journal of Science & Medicine in Sport, 22*(6), 653-660. doi:10.1016/j.jsams.2018.11.028

Cunniffe, B., Griffiths, H., Proctor, W., Davies, B., Baker, J. S., & Jones, K. P. (2011). Mucosal immunity and illness incidence in elite rugby union players across a season. *Medicine & Science in Sports & Exercise, 43*(3), 388-397. doi:10.1249/MSS.0b013e3181ef9d6b

da Silva, C. C., Goldberg, T. B., Soares-Caldeira, L. F., Dos Santos Oliveira, R., de Paula Ramos, S., & Nakamura, F. Y. (2015). The effects of 17 weeks of ballet training on the autonomic modulation, hormonal and general biochemical profile of female adolescents. *Journal of Human Kinetics, 47*, 61-71. doi:10.1515/hukin-2015-0062

Davis, H. I., Orzeck, T., & Keelan, P. (2007). Psychometric item evaluations of the recovery-stress questionnaire for athletes. *Psychology of Sport and Exercise, 8*(6), 917–938.

de Boer, A. G., van Lanschot, J. J., Stalmeier, P. F., van Sandick, J. W., Hulscher, J. B., de Haes, J. C., & Sprangers, M. A. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research, 13*(2), 311-320. doi:10.1023/B:QURE.0000018499.64574.1f

de Francisco, C., Arce, C., Andrade, E., Arce, I., & Raedeke, T. (2009). Propiedades psicométricas preliminares de la versión española del athlete burnout questionnaire en una muestra de jóvenes futbolistas = preliminary psychometric properties of the spanish version of the athlete burnout questionnaire in a sample of young soccer players. *Cuadernos de Psicología del Deporte, 9*(2), 45-56.

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). *International handbook of survey methodology*. New York: Psychology Press.

de Vet, H. C., Terwee, C. B., & Bouter, L. M. (2003). Current challenges in clinimetrics. *Journal of Clinical Epidemiology, 56*(12), 1137-1141. doi:10.1016/j.jclinepi.2003.08.012

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide.* : Cambridge: Cambridge University Press.

Decroix, L., Lamberts, R. P., & Meeusen, R. (2018). Can the lamberts and lambert submaximal cycle test reflect overreaching in professional cyclists? *International Journal of Sports Physiology and Performance, 13*(1), 23-28. doi:10.1123/ijspp.2016-0685

Decroix, L., Piacentini, M. F., Rietjens, G., & Meeusen, R. (2016). Monitoring physical and cognitive overload during a training camp in professional female cyclists. *International Journal of Sports Physiology & Performance, 11*(7), 933-939.

Delecroix, B., McCall, A., Dawson, B., Berthoin, S., & Dupont, G. (2018). Workload and non-contact injury incidence in elite football players competing in european leagues. *European Journal of Sport Sciences, 18*(9), 1280-1287. doi:10.1080/17461391.2018.1477994

Dergaa, I., Fessi, M. S., Chaabane, M., Souissi, N., & Hammouda, O. (2019). The effects of lunar cycle on the diurnal variations of short-term maximal performance, mood state, and perceived exertion. *Chronobiology International, 36*(9), 1249-1257. doi:10.1080/07420528.2019.1637346

Doeven, S. H., Brink, M. S., Huijgen, B. C. H., de Jong, J., & Lemmink, K. (2019). High match load's relation to decreased well-being during an elite women's rugby sevens tournament. *International Journal of Sports Physiology & Performance, 14*(8), 1036-1042. doi:10.1123/ijspp.2018-0516

Drew, M. K., & Finch, C. F. (2016). The relationship between training load and injury, illness and soreness: A systematic and literature review. *Sports Medicine, 46*(6), 861-883. doi:10.1007/s40279-015-0459-8

Dryden, D. M., Francescutti, L. H., Rowe, B. H., Spence, J. C., & Voaklander, D. C. (2000). Personal risk factors associated with injury among female recreational ice hockey players. *Journal of Science and Medicine in Sport, 3*(2), 140-149.

Dubey-Pathak, M. (2014). The rock art of the bhimbetka area in india. *Adoranten*, 5-22.

Ducher, G., Kukuljan, S., Hill, B., Garnham, A. P., Nowson, C. A., Kimlin, M. G., & Cook, J. (2011). Vitamin d status and musculoskeletal health in adolescent male ballet dancers a pilot study. *Journal of Dance Medicine & Science, 15*(3), 99-107.

Durand-Bush, N., & Salmela, J. H. (2002). The development and maintenance of expert athletic performance: Perceptions of world and olympic champions. *Journal of Applied Sport Psychology, 14*(3), 154-171.

Eckard, T. G., Padua, D. A., Hearn, D. W., Pexa, B. S., & Frank, B. S. (2018). The relationship between training load and injury in athletes: A systematic review. *Sports Medicine, 48*(8), 1929-1961. doi:10.1007/s40279-018-0951-z

Edwards, R. H. T. (1983). *Biochemical basis of fatigue in exercise performance*. Champaign, IL Human Kinetics.

Edwards, W. B. (2018). Modeling overuse injuries in sport as a mechanical fatigue phenomenon. *Exercise and Sport Sciences Reviews, 46*(4), 224-231. doi:10.1249/jes.0000000000000163

Ehrmann, F. E., Duncan, C. S., Sindhusake, D., Franzsen, W. N., & Greene, D. A. (2016). Gps and injury prevention in professional soccer. *Journal of Strength & Conditioning Research, 30*(2), 360-367. doi:10.1519/JSC.0000000000001093

Ekegren, C. L., Quested, R., & Brodrick, A. (2014). Injuries in pre-professional ballet dancers: Incidence, characteristics and consequences. *Journal of Science and Medicine in Sport, 17*(3), 271-275. doi:10.1016/j.jsams.2013.07.013

Emery, C. A. (1999). Does decreased muscle strength cause acute muscle strain injury in sport? A systematic review of the evidence. *Physical Therapy Reviews, 4*, 79-85.

Emery, C. A. (2003). Risk factors for injury in child and adolescent sport: A systematic review of the literature. *Clinical Journal of Sports Medicine, 13*(4), 256-268.

Emery, C. A. (2010). Injury prevention in paediatric sport-related injuries: A scientific approach. *British Journal of Sports Medicine, 44*(1), 64-69. doi:10.1136/bjsm.2009.068353

Emery, C. A., Meeuwisse, W. H., & McAllister, J. R. (2006). Survey of sport participation and sport injury in calgary and area high schools. *Clinical Journal of Sport Medicine, 16*(1), 20-26. doi:10.1097/01.jsm.0000184638.72075.b7

Endler, S., Hoffmann, S., Sterzing, B., Simon, P., & Pfeiffer, M. (2017). The perpot simulated anaerobic threshold - a comparison to typical lactate-based thresholds. *International Journal of Human Movement Science, 5*(1), 9-15. doi:DOI: 10.13189/saj.2017.050102

Engebretsen, L., Soligard, T., Steffen, K., Alonso, J. M., Aubry, M., Budgett, R., . . . Renström, P. A. (2013). Sports injuries and illnesses during the london summer olympic games 2012. *British Journal of Sports Medicine, 47*(7), 407-414. doi:10.1136/bjsports-2013-092380

Engebretsen, L., Steffen, K., Alonso, J. M., Aubry, M., Dvorak, J., Junge, A., . . . Wilkinson, M. (2010). Sports injuries and illnesses during the winter olympic games 2010. *British Journal of Sports Medicine, 44*(11), 772-780. doi:10.1136/bjsm.2010.076992

Engel, F. A., Faude, O., Kolling, S., Kellmann, M., & Donath, L. (2019). Verbal encouragement and between-day reliability during high-intensity functional strength and endurance performance testing. *Frontiers in Physiology, 10*, 460. doi:10.3389/fphys.2019.00460

Enoka, R. M., & Duchateau, J. (2016). Translating fatigue to human performance. *Medicine and Science in Sports and Exercise, 48*(11), 2228-2238.

European Medicines Agency. (2005). *Reflection paper on the regulatory guidance for the use of health related quality of life (hrql) measures in the evaluation of medicinal products*. Retrieved from London, UK:

Eysenbach, G. (2004). Improving the quality of web surveys: The checklist for reporting results of internet e-surveys (cherries). *Journal of Medical Internet Research, 6*(3), e34-e34. doi:10.2196/jmir.6.3.e34

Fanchini, M., Schena, F., Castagna, C., Petruolo, A., Combi, F., McCall, A., & Impellizzeri, M. (2015). External responsiveness of the yo-yo ir test level 1 in high-level male soccer players. *International Journal of Sports Medicine, 36*(9), 735-741. doi:10.1055/s-0035-1547223

Fazackerley, R., Lane, A. M., & Mahoney, C. (2003). Confirmatory factor analysis of the brunel mood scale for use with water-skiing competition. *Perceptual & Motor Skills, 97*(2), 657-661. doi:10.2466/pms.2003.97.2.657

Feinstein, A. R. (1987). Clinimetric perspectives. *Journal of Chronic Diseases, 40*(6), 635-640. doi:10.1016/0021-9681(87)90027-0

Fernández, E. M. A., Fernández, C. A., & Pesqueira, G. S. (2002). Adaptación al español del cuestionario perfil de los estados de ánimos en una muestra de deportistas = adaptation of the profile of mood states into spanish with a sample of athletes. *Psicothema, 14*(4), 708-713.

Fietzer, A. L., Chang, Y., & Kulig, K. (2012). Dancers with patellar tendinopathy exhibit higher vertical and braking ground reaction forces during landing. *Journal of Sports Sciences, 30*(11), 1157-1163. doi:10.1080/02640414.2012.695080

Finch, C. F., & Fortington, L. V. (2018). So you want to understand subsequent injuries better? Start by understanding the minimum data collection and reporting requirements. *British Journal of Sports Medicine, 52*(17), 1077-1078. doi:10.1136/bjsports-2017-098225

Fitzpatrick, J. F., Akenhead, R., Russell, M., Hicks, K. M., & Hayes, P. R. (2019). Sensitivity and reproducibility of a fatigue response in elite youth football players. *Science & Medicine in Football, 3*(3), 214-220. doi:10.1080/24733938.2019.1571685

Fitzpatrick, J. F., Hicks, K. M., Russell, M., & Hayes, P. R. (2019). The reliability of potential fatigue-monitoring measures in elite youth soccer players. *Journal of Strength & Conditioning Research*. doi:10.1519/JSC.0000000000003317

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382. doi:10.1037/h0031619

Flore, P., Sarrazin, P., & Favre-Juvin, A. (2003). Le questionnaire de surentrainement de la societe francaise de medecine du sport: Reproductibilite a court terme. / overtraining questionnaire of the societe francaise de medecine du sport (french society of sport medicine): Short term reproducibility. *Science & Sports, 18*(6), 290-292.

Foster, C. (1998). Monitoring training in athletes with reference to overtraining syndrome. *Medicine & Science in Sports & Exercise, 30*(7), 1164-1168.

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

Fowler, P., Duffield, R., & Vaile, J. (2014). Effects of domestic air travel on technical and tactical performance and recovery in soccer. *International Journal of Sports Physiology and Performance, 9*(3), 378-386. doi:10.1123/IJSPP.2013-0484

Freitas, C. G., Aoki, M. S., Franciscon, C. A., Arruda, A. F., Carling, C., & Moreira, A. (2014). Psychophysiological responses to overloading and tapering phases in elite young soccer players. *Pediatric Exercise Science, 26*(2), 195-202. doi:10.1123/pes.2013-0094

Freslon, M., Soyer, J., Iborra, J. P., Hadjadj, S., Pries, P., & Clarac, J. P. (2004). An unusual cause of a stress femoral fracture in a long-distance runner: A case of bilateral fracture. *Journal of Trauma, 56*(2), 433-436. doi:10.1097/01.ta.0000029369.59608.41

Fry, R. W., Morton, A. R., & Keast, D. (1991). Overtraining in athletes. An update. *Sports Medicine, 12*(1), 32-65.

Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., . . . Meeuwisse, W. H. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scandinavian Journal of Medicine and Science in Sports, 16*(2), 83-92. doi:10.1111/j.1600-0838.2006.00528.x

Gabbett, T. J. (2016). The training—injury prevention paradox: Should athletes be training smarter and harder? *British Journal of Sports Medicine, 50*(5), 273-280. doi:10.1136/bjsports-2015-095788

Galambos, S. A., Terry, P. C., Moyle, G. M., Locke, S. A., & Lane, A. M. (2005). Psychological predictors of injury among elite athletes. *British Journal of Sports Medicine, 39*(6), 351-354; discussion 351-354. doi:10.1136/bjsm.2005.018440

Gamboa, J. M., Roberts, L. A., Maring, J., & Fergus, A. (2008). Injury patterns in elite preprofessional ballet dancers and the utility of screening programs to identify risk characteristics. *Journal of Orthopaedic & Sports Physical Therapy, 38*(3), 126-136. doi:10.2519/jospt.2008.2390

Garrick, J. G., & Lewis, S. L. (2001). Career hazards for the dancer. *Occupational Medicine, 16*(4), 609-618, iv.

Gastin, P. B., Meyer, D., & Robinson, D. (2013). Perceptions of wellness to monitor adaptive responses to training and competition in elite australian football. *Journal of Strength & Conditioning Research, 27*(9), 2518-2526. doi:10.1519/JSC.0b013e31827fd600

Gathercole, R., Sporer, B., & Stellingwerff, T. (2015). Countermovement jump performance with increased training loads in elite female rugby athletes. *International Journal of Sports Medicine, 36*(9), 722-728. doi:10.1055/s-0035-1547262

Gerber, M., Gustafsson, H., Seelig, H., Kellmann, M., Ludyga, S., Colledge, F., . . . Bianchi, R. (2018). Usefulness of the athlete burnout questionnaire (abq) as a screening tool for the detection of clinically relevant burnout symptoms among young elite athletes. *Psychology of Sport and Exercise, 39*, 107-113.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482. doi:10.1146/annurev-psych-120709-145346

Giles, S., Fletcher, D., Arnold, R., Ashfield, A., & Harrison, J. (2020). Measuring well-being in sport performers: Where are we now and how do we progress? *Sports Medicine*. doi:10.1007/s40279-020-01274-z

Gimeno-Santos, E., Frei, A., Dobbels, F., Rüdell, K., Puhan, M. A., Garcia-Aymerich, J., & the, P. c. (2011). Validity of instruments to measure physical activity may be questionable due to a lack of conceptual frameworks: A systematic review. *Health Quality of Life Outcomes, 9*(1), 86. doi:10.1186/1477-7525-9-86

Gleeson, M., Bishop, N., Oliveira, M., McCauley, T., Tauler, P., & Muhamad, A. S. (2012). Respiratory infection risk in athletes: Association with antigen-stimulated il-10 production and salivary iga secretion. *Scandinavian Journal of Medicine and Science in Sports, 22*(3), 410-417. doi:10.1111/j.1600-0838.2010.01272.x

Gleeson, M., Bishop, N., Oliveira, M., & Tauler, P. (2013). Influence of training load on upper respiratory tract infection incidence and antigen-stimulated cytokine production. *Scandinavian Journal of Medicine & Science in Sports, 23*(4), 451-457. doi:10.1111/j.1600-0838.2011.01422.x

Gleeson, M., & Pyne, D. B. (2016). Respiratory inflammation and infections in high-performance athletes. *Immunology and Cell Biology, 94*(2), 124-131. doi:10.1038/icb.2015.100

Gomes, R. V., Moreira, A., Lodo, L., Nosaka, K., Coutts, A. J., & Aoki, M. S. (2013). Monitoring training loads, stress, immune-endocrine responses and performance in tennis players. *Biology of Sport, 30*(3), 173-180. doi:10.5604/20831862.1059169

Goncalves, L., Clemente, F. M., Silva, B., Mendes, B., Lima, R., Bezerra, P., & Camoes, M. (2020). Variations of season workload and well-being status among professional roller-hockey players: Full season analysis. *Physiology & Behavior, 215*, 112785. doi:10.1016/j.physbeh.2019.112785

González-Boto, R., Salguero, A., Tuero, C., Márquez, S., & Kellmann, M. (2008). Spanish adaptation and analysis by structural equation modeling of an instrument for monitoring overtraining: The recovery-stress questionnaire. *Social Behavior and Personality, 36*(5), 635-650. doi:10.2224/sbp.2008.36.5.635

Goode, K. T., & Roth, D. L. (1993). Factor analysis of cognitions during running: Association with mood change. *Journal of Sport & Exercise Psychology, 15*, 375-389.

Gould, D., Damarjian, N., & Greenleaf, C. (2002). Imagery training for peak performance. In J. L. Van Raalte & B. W. Brewer (Eds.), *Exploring sport and exercise psychology* (pp. 49-74): American Psychological Association.

Govus, A. D., Coutts, A., Duffield, R., Murray, A., & Fullagar, H. (2018). Relationship between pretraining subjective wellness measures, player load, and rating-of-perceived-exertion training load in american college football. *International Journal of Sports Physiology and Performance, 13*(1), 95-101. doi:10.1123/ijspp.2016-0714

Greenland, S., & Brumback, B. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology, 31*(5), 1030-1037. doi:10.1093/ije/31.5.1030

Grove, J. R., & Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated profile of mood states. *International Journal of Sport Psychology, 23*(2), 93-109.

Guidetti, L., Gallotta, M. C., Emerenziani, G. P., & Baldari, C. (2007). Exercise intensities during a ballet lesson in female adolescents with different technical ability. *International Journal of Sports Medicine, 28*(9), 736-742. doi:10.1055/s-2007-964909

Halson, S. (2014). Monitoring training load to understand fatigue in athletes. *Sports Medicine, 44*, 139-147. doi:10.1007/s40279-014-0253-z

Hamilton, G. M., Meeuwisse, W. H., Emery, C. A., & Shrier, I. (2012). Examining the effect of the injury definition on risk factor analysis in circus artists. *Scandinavian Journal of Medicine & Science in Sports, 22*(3), 330-334. doi:10.1111/j.1600-0838.2010.01245.x

Hamilton, W. G., Hamilton, L. H., Marshall, P., & Molnar, M. (1992). A profile of the musculoskeletal characteristics of elite professional ballet dancers. *American Journal of Sports Medicine, 20*(3), 267-273. doi:10.1177/036354659202000306

Harel, O., Mitchell, E. M., Perkins, N. J., Cole, S. R., Tchetgen Tchetgen, E. J., Sun, B., & Schisterman, E. F. (2018). Multiple imputation for incomplete data in epidemiologic studies. *American Journal of Epidemiology, 187*(3), 576-584. doi:10.1093/aje/kwx349

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., . . . Duda, S. N. (2019). The redcap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, 103208. doi:10.1016/j.jbi.2019.103208

Harrison, C., Ruddock-hudson, M., Ruddock, S., Mayes, S., O'Halloran, P., & Cook, J. (2019). Wellness monitoring in professional ballet dancers: A pilot study. *Journal of Science & Medicine in Sport, 22*.

Hartwig, T. B., Gabbett, T. J., Naughton, G., Duncan, C., Harries, S., & Perry, N. (2019). Training and match volume and injury in adolescents playing multiple contact team sports: A prospective cohort study. *Scandinavian Journal of Medicine & Science in Sports, 29*(3), 469-475. doi:10.1111/sms.13343

Heath, G. W., Ford, E. S., Craven, T. E., Macera, C. A., Jackson, K. L., & Pate, R. R. (1991). Exercise and the incidence of upper respiratory tract infections. *Medicine & Science in Sports & Exercise, 23*(2), 152-157.

Hellard, P., Avalos, M., Guimaraes, F., Toussaint, E.-F., & Pyne, D. B. (2015). Training-related risk of common illnesses in elite swimmers over a 4-yr period. *Medicine & Science in Sports & Exercise, 47*(4), 698-707. doi:10.1249/MSS.0000000000000461

Hellard, P., Avalos, M., Guimaraes, F., Toussaint, J. F., & Pyne, D. B. (2015). Training-related risk of common illnesses in elite swimmers over a 4-yr period. *Medicine & Science in Sports & Exercise, 47*(4), 698-707. doi:10.1249/MSS.0000000000000461

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology., 15*(5), 615-625. doi:10.1097/01.ede.0000135174.63482.43

Herold, F., Torpel, A., Hamacher, D., Budde, H., & Gronwald, T. (2020). A discussion on different approaches for prescribing physical interventions - four roads lead to rome, but which one should we choose? *Journal of Personalized Medicine, 10*(3), 55. doi:10.3390/jpm10030055

Hills, S. P., & Rogerson, D. J. (2018). Associatons between self-reported well-being and neuromuscular performance during a professional rugby union season. *Journal of Strength & Conditioning Research, 32*(9), 2498-2509. doi:10.1519/jsc.0000000000002531

Hincapie, C. A., Morton, E. J., & Cassidy, J. D. (2008). Musculoskeletal injuries and pain in dancers: A systematic review. *Archives of Physical Medicine and Rehabilitation, 89*(9), 1819-1829. doi:10.1016/j.apmr.2008.02.020

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. doi:10.1026/1612-5010/a000150

Hitzschke, B., Wiewelhove, T., Raeder, C., Ferrauti, A., Meyer, T., Pfeiffer, M., . . . Kölling, S. (2017). Evaluation of psychological measures for the assessment of recovery and stress during a shock-microcycle in strength and high-intensity interval training. *Performance Enhancement & Health, 5*(4), 147-157.

Hodgson, L., Gissane, C., Gabbett, T., & King, D. (2007). For debate: Consensus injury definitions in team sports should focus on encompassing all injuries. *Clinical Journal of Sports Medicine, 17*, 188-191. doi:10.1097/JSM.0b013e3180547513

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305-312. doi:10.1016/j.jsat.2011.04.005

Hooper, S. L., & Mackinnon, L. T. (1995). Monitoring overtraining in athletes. Recommendations. *Sports Medicine, 20*(5), 321-327.

Hooper, S. L., MacKinnon, L. T., & Hanrahan, S. (1997). Mood states as an indication of staleness and recovery. *International Journal of Sport & Exercise Psychology, 28*(1), 1-12.

Hooper, S. L., Mackinnon, L. T., Howard, A., Gordon, R. D., & Bachmann, A. W. (1995). Markers for monitoring overtraining and recovery. *Medicine & Science in Sports & Exercise, 27*(1), 106-112.

Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine, 48*(8), 708-712. doi:10.1136/bjsports-2013-092524

Hulin, B. T., Gabbett, T. J., Caputi, P., Lawson, D. W., & Sampson, J. A. (2016). Low chronic workload and the acute:Chronic workload ratio are more predictive of injury than between-match recovery time: A two-season prospective cohort study in elite rugby league players. *British Journal of Sports Medicine, 50*(16), 1008-1012. doi:10.1136/bjsports-2015-095364

Hulin, B. T., Gabbett, T. J., Lawson, D. W., Caputi, P., & Sampson, J. A. (2016). The acute:Chronic workload ratio predicts injury: High chronic workload may decrease injury risk in elite rugby league players. *British Journal of Sports Medicine, 50*(4), 231-236. doi:10.1136/bjsports-2015-094817

Hurdiel, R., Pezé, T., Daugherty, J., Girard, J., Poussel, M., Poletti, L., . . . Theunynck, D. (2015). Combined effects of sleep deprivation and strenuous exercise on cognitive performances during the north face® ultra trail du mont blanc® (utmb®). *Journal of Sports Sciences, 33*(7), 670-674. doi:10.1080/02640414.2014.960883

Impellizzeri, F., & Marcora, S. (2009). Test validation in sport physiology: Lessons learned from clinimetrics. *International Journal of Sports Physiology & Performance, 4*, 269-277. doi:10.1123/ijspp.4.2.269

Impellizzeri, F. M., Jones, D. M., Griffin, D., Harris-Hayes, M., Thorborg, K., Crossley, K. M., . . . Kemp, J. (2020). Patient-reported outcome measures for hip-related pain: A review of the available evidence and a consensus statement from the international hip-related pain research network, zurich 2018. *British Journal of Sports Medicine, 54*(14), 848-857. doi:10.1136/bjsports-2019-101456

Impellizzeri, F. M., & Maffiuletti, N. A. (2007). Convergent evidence for construct validity of a 7-point likert scale of lower limb muscle soreness. *Clinical Journal of Sport Medicine, 17*(6), 494-496. doi:10.1097/JSM.0b013e31815aed57

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 years on. *International Journal of Sports Physiology & Performance, 14*(2), 270-273. doi:10.1123/ijspp.2018-0935

Impellizzeri, F. M., McCall, A., Ward, P., Bornn, L., & Coutts, A. J. (2020). Training load and its role in injury prevention, part 2: Conceptual and methodologic pitfalls. *Journal of Athletic Training, 55*(9), 893-901. doi:10.4085/1062-6050-501-19

Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training load and its role in injury prevention, part i: Back to the future. *Journal of Athletic Training, 55*(9), 885-892. doi:10.4085/1062-6050-500-19

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. *Journal of Sports Sciences, 23*(6), 583-592. doi:10.1080/02640410400021278

Isoard-Gautheur, S., Oger, M., Guillet, E., & Martin-Krumm, C. (2010). Validation of a french version of the athlete burnout questionnaire (abq). *European Journal of Psychological Assessment, 26*(3), 203-211.

Jacobs, C. L., Cassidy, J. D., Cote, P., Boyle, E., Ramel, E., Ammendolia, C., . . . Schwartz, I. (2017). Musculoskeletal injury in professional dancers: Prevalence and associated factors: An international cross-sectional study. *Clinical Journal of Sports Medicine, 27*(2), 153-160. doi:10.1097/JSM.0000000000000314

Jacobs, C. L., Hincapie, C. A., & Cassidy, J. D. (2012). Musculoskeletal injuries and pain in dancers: A systematic review update. *Journal of Dance Medicine & Science, 16*(2), 74-84.

Jaspers, A., Kuyvenhoven, J. P., Staes, F., Frencken, W. G. P., Helsen, W. F., & Brink, M. S. (2018). Examination of the external and internal load indicators' association with overuse injuries in professional soccer players. *Journal of Science & Medicine in Sport, 21*(6), 579-585. doi:10.1016/j.jsams.2017.10.005

Jeffries, A., Marcora, S., Coutts, A. J., Wallace, L., McCall, A., & Impellizzeri, F. M. (2020). Development of a revised conceptual framework of physical training for measurement validation and other applications. doi:https://doi.org/10.31236/osf.io/wpvek

Jeffries, A. C., Wallace, L., & Coutts, A. J. (2017). Quantifying training loads in contemporary dance. *International Journal of Sports Physiology & Performance, 12*(6), 796-802. doi:10.1123/ijspp.2016-0159

Jeffries, A. C., Wallace, L., Coutts, A. J., Cohen, A. M., McCall, A., & Impellizzeri, F. M. (2020). Injury, illness, and training load in a professional contemporary dance company: A prospective study. *Journal of Athletic Training, 55*(9), 967-976. doi:10.4085/1062-6050-477-19

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 & Performance, 15*(9), 1-13. doi:10.1123/ijspp.2020-0386

Joffe, M., Gambhir, M., Chadeau-Hyam, M., & Vineis, P. (2012). Causal diagrams in systems epidemiology. *Emerging Themes in Epidemiology, 9*(1), 1. doi:10.1186/1742-7622-9-1

Johnson, U., & Ivarsson, A. (2011). Psychological predictors of sport injuries among junior soccer players. *Scandinavian Journal of Medicine & Science in Sports, 21*(1), 129-136. doi:10.1111/j.1600-0838.2009.01057.x

Johnston, R., Cahalan, R., Bonnett, L., Maguire, M., Nevill, A., Glasgow, P., . . . Comyns, T. (2019). Training load and baseline characteristics associated with new injury/pain within an endurance sporting population: A prospective study. *International Journal of Sports Physiology & Performance, 14*(5), 590-597. doi:10.1123/ijspp.2018-0644

Jones, A. M., & Carter, H. (2000). The effect of endurance training on parameters of aerobic fitness. *Sports Medicine, 29*(6), 373-386. doi:10.2165/00007256-200029060-00001

Jones, B. H., Bovee, M. W., Harris, J. M., 3rd, & Cowan, D. N. (1993). Intrinsic risk factors for exercise-related injuries among male and female army trainees. *American Journal of Sports Medicine, 21*(5), 705-710. doi:10.1177/036354659302100512

Joyner, M. J., & Coyle, E. F. (2008). Endurance exercise performance: The physiology of champions. *Journal of Physiology, 586*(1), 35-44. doi:10.1113/jphysiol.2007.143834

Judge, L. W., Urbina, L. J., Hoover, D. L., Craig, B. W., Judge, L. M., Leitzelar, B. M., . . . Bellar, D. M. (2016). The impact of competitive trait anxiety on collegiate powerlifting performance. *Journal of Strength and Conditioning Research, 30*(9), 2399-2405. doi:10.1519/JSC.0000000000001363

Julian, R., Meyer, T., Fullagar, H. H., Skorski, S., Pfeiffer, M., Kellmann, M., . . . Hecksteden, A. (2017). Individual patterns in blood-borne indicators of fatigue-trait or chance. *Journal of Strength & Conditioning Research, 31*(3), 608-619. doi:10.1519/JSC.0000000000001390

Kallus, K. W. (1995). *Der erholungs-belastungs-fragebogen [the recovery stress questionnaire].* Frankfurt: Swets & Zeitlinger.

Kallus, K. W., & Kellmann, M. (2016). *The recovery-stress questionnaires: User manual.* Frankfurt am Main: Pearson Assessment & Information GmbH.

Kellmann, M., & Kallus, K. W. (2001). *Recovery-stress questionnaire for athletes: User manual*. Champaign, IL: Human Kinetics.

Kellmann, M., Kolling, S., & Hitzschke, B. (2016). *Das akutmab 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].* Hellenthal: Sportverlag Straub.

Kenny, S. J., Palacios-Derflingher, L., Whittaker, J. L., & Emery, C. A. (2018). The influence of injury definition on injury burden in preprofessional ballet and contemporary dancers. *Journal of Orthopaedic and Sports Physical Therapy, 48*(3), 185-193. doi:10.2519/jospt.2018.7542

Kenny, S. J., Whittaker, J. L., & Emery, C. A. (2016). Risk factors for musculoskeletal injury in preprofessional dancers: A systematic review. *British Journal of Sports Medicine, 50*(16), 997-1003. doi:10.1136/bjsports-2015-095121

Kenttä, G., & Hassmén, P. (1998). Overtraining and recovery. A conceptual model. *Sports Medicine, 26*(1), 1-16. doi:10.2165/00007256-199826010-00001

Khan, K., Brown, J., Way, S., Vass, N., Crichton, K., Alexander, R., . . . Wark, J. (1995). Overuse injuries in classical ballet. *Sports Medicine, 19*(5), 341-357.

Kibler, B. W., & Chandler, T. J. (1998). Musculoskeletal and orthopedic considerations. . In R. B. Kreider, A. C. Fry, & M. K. O'Toole (Eds.), *Overtraining in sport* (pp. 169-190). Champaign, IL: Human Kinetics.

Kibler, W. B., Chandler, T. J., & Stracener, E. S. (1992). Musculoskeletal adaptations and injuries due to overtraining. *Exercise and Sport Sciences Reviews, 20*, 99-126.

Kiely, J. (2012). Periodization paradigms in the 21st century: Evidence-led or tradition-driven? *International Journal of Sports Physiology & Performance, 7*(3), 242. doi:10.1123/ijspp.7.3.242

Kjaer, M., & Larsson, B. (1992). Physiological profile and incidence of injuries among elite figure skaters. *Journal of Sports Sciences, 10*(1), 29-36. doi:10.1080/02640419208729904

Knuiman, P., Hopman, M. T. E., & Mensink, M. (2015). Glycogen availability and skeletal muscle adaptations with endurance and resistance exercise. *Nutrition & Metabolism, 12*(1), 59. doi:10.1186/s12986-015-0055-9

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*(2 & 3), 529-542.

Kolling, S., Schaffran, P., Bibbey, A., Drew, M., Raysmith, B., Nassi, A., & Kellmann, M. (2019). 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*, 1-10. doi:10.1080/02640414.2019.1684790

Kolling, S., Steinacker, J. M., Endler, S., Ferrauti, A., Meyer, T., & Kellmann, M. (2016). The longer the better: Sleep-wake patterns during preparation of the world rowing junior championships. *Chronobiology International, 33*(1), 73-84. doi:10.3109/07420528.2015.1118384

Kolt, G. S., & Kirkby, R. J. (1999). Epidemiology of injury in elite and subelite female gymnasts: A comparison of retrospective and prospective findings. *British Journal of Sports Medicine, 33*(5), 312-318.

König, D., Grathwohl, D., Weinstock, C., Northoff, H., & Berg, A. (2000). Upper respiratory tract infection in athletes: Influence of lifestyle, type of sport, training effort, and immunostimulant intake. *Exercise Immunology Review, 6*, 102-120.

Kuipers, H., & Keizer, H. A. (1988). Overtraining in elite athletes. Review and directions for the future. *Sports Medicine, 6*(2), 79-92. doi:10.2165/00007256-198806020-00003

Kuno, M., Fukunaga, T., Hirano, Y., & Miyashita, M. (1996). Anthropometric variables and muscle properties of japanese female ballet dancers. *International Journal of Sports Medicine., 17*(2), 100-105. doi:10.1055/s-2007-972815

Lacome, M., Carling, C., Hager, J.-P., Dine, G., & Piscione, J. (2018). Workload, fatigue, and muscle damage in an under-20 rugby union team over an intensified international tournament. *International Journal of Sports Physiology & Performance, 13*(8), 1059-1066.

Lakens, D. (2020). Pandemic researchers - recruit your own best critics. *Nature, 581*(7807), 121. doi:10.1038/d41586-020-01392-8

Lan, M. F., Lane, A. M., Roy, J., & Hanin, N. A. (2012). Validity of the brunel mood scale for use with malaysian athletes. *Journal of Sports Science and Medicine, 11*(1), 131-135.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lane, K., Wolery, M., Reichow, B., & Rogers, L. (2007). Describing baseline conditions: Suggestions for study reports. *Journal of Behavioral Education, 16*(3), 224-234. doi:10.1007/s10864-006-9036-4

Laurent, C. M., Green, J. M., Bishop, P. A., Sjokvist, J., Schumacker, R. E., Richardson, M. T., & Curtner-Smith, M. (2011). A practical approach to monitoring recovery: Development of a perceived recovery status scale. *Journal of Strength & Conditioning Research, 25*(3), 620-628. doi:10.1519/JSC.0b013e3181c69ec6

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods* (Vol. Vols. 1-0). Thousand Oaks, CA: Sage Publications.

Legros, P. (1993). Le surentraînement: Diagnostic des manifestations psychocomportementales précoces. *Science & Sports, 8.*, 71-74.

Lehmann, M., Foster, C., & Keul, J. (1993). Overtraining in endurance athletes: A brief review. *Medicine & Science in Sports & Exercise, 25*(7), 854-862.

Lewis, G., & Wessely, S. (1992). The epidemiology of fatigue: More questions than answers. *Journal of Epidemiology & Community Health, 46*(2), 92-97. doi:10.1136/jech.46.2.92

Liederbach, M., Gleim, G. W., & Nicholas, J. A. (1992). Monitoring training status in professional ballet dancers. *Journal of Sports Medicine and Physical Fitness, 32*(2), 187-195.

Liederbach, M., Hagins, M., Gamboa, J. M., & Welsh, T. M. (2012). Assessing and reporting dancer capacities, risk factors, and injuries: Recommendations from the iadms standard measures consensus initiative. *Journal of Dance Medicine & Science, 16*(4), 139-153.

Liederbach, M., Schanfein, L., & Kremenic, I. J. (2013). What is known about the effect of fatigue on injury occurrence among dancers? *Journal of Dance Medicine and Science, 17*(3), 101-108.

Liu, C. Y., Tseng, M. M., Chang, C. H., Fang, D., & Lee, M. B. (2016). Comorbid psychiatric diagnosis and psychological correlates of eating disorders in dance students. *Journal of the Formosan Medical Association, 115*(2), 113-120.

Luke, A., Kinney, S., & d'Hemecourt, P. A. (2002). Determinants of injuries in young dancers. *Medical Problems of Performing Artists, 17*, 105-112.

Lysenko, E. A., Popov, D. V., Vepkhvadze, T. F., Sharova, A. P., & Vinogradova, O. L. (2019). Signaling responses to high and moderate load strength exercise in trained muscle. *Physiological Reports, 7*(9), e14100. doi:10.14814/phy2.14100

Maffey, L., & Emery, C. (2007). What are the risk factors for groin strain injury in sport? A systematic review of the literature. *Sports Medicine, 37*(10), 881-894.

Main, L., & Grove, R. (2009). A multi-component assessment model for monitoring training distress among athletes. *European Journal of Sport Science, 9*(4), 195-202. doi:10.1080/17461390902818260

Malm, C. (2006). Susceptibility to infections in elite athletes: The s-curve. *Scandinavian Journal of Medicine & Science in Sports, 16*(1), 4-6. doi:10.1111/j.1600-0838.2005.00499.x

Malone, S., Owen, A., Newton, M., Mendes, B., Collins, K. D., & Gabbett, T. J. (2017). The acute:Chonic workload ratio in relation to injury risk in professional soccer. *Journal of Science & Medicine in Sport, 20*(6), 561-565. doi:10.1016/j.jsams.2016.10.014

Malone, S., Owen, A., Newton, M., Mendes, B., Tiernan, L., Hughes, B., & Collins, K. (2018). Wellbeing perception and the impact on external training output among elite soccer players. *Journal of Science and Medicine in Sport, 21*(1), 29-34. doi:10.1016/j.jsams.2017.03.019

Marin, D. P., Bolin, A. P., Campoio, T. R., Guerra, B. A., & Otton, R. (2013). Oxidative stress and antioxidant status response of handball athletes: Implications for sport training monitoring. *International Immunopharmacology, 17*(2), 462-470. doi:10.1016/j.intimp.2013.07.009

Marrier, B., Le Meur, Y., Leduc, C., Piscione, J., Lacome, M., Igarza, G., . . . Robineau, J. (2018). Training periodization over an elite rugby sevens season: From theory to practice. *International Journal of Sports Physiology & Performance*, 1-24. doi:10.1123/ijspp.2017-0839

Mason, B., McKune, A., Pumpa, K., & Ball, N. (2020). The use of acute exercise interventions as game day priming strategies to improve physical performance and athlete readiness in team-sport athletes: A systematic review. *Sports Medicine, 50*(11), 1943-1962. doi:10.1007/s40279-020-01329-1

Mastro, J. V., French, R., & Hall, M. M. (1987). Test-retest reliability of the profile of mood states using visually impaired athletes. *Perceptual and Motor Skills, 65*(2), 593-594. doi:10.2466/pms.1987.65.2.593

Matos, S., Clemente, F. M., Brandao, A., Pereira, J., Rosemann, T., Nikolaidis, P. T., & Knechtle, B. (2019). Training load, aerobic capacity and their relationship with wellness status in recreational trail runners. *Frontiers in Physiology, 10*, 1189. doi:10.3389/fphys.2019.01189

Matveyev, L. (1977). *Fundamentals of sports training*. Moscow. Russia: Fizkultura i Sport Publ.

Maxwell, J. (2005). Conceptual framework: What do you think is going on? In qualitative research design: An interactive approach (2nd ed.). Thousand Oaks: SAGE Publications.

McGahan, J., Burns, C., Lacey, S., Gabbett, T., & O'Neill, C. (2018). Relationship between load and readiness to train in a gaelic football pre-competition training camp. *Journal of Australian Stength & Conditioning, 27*(1), 28-35.

McGahan, J., Burns, C., Lacey, S., Gabbett, T., & O'Neill, C. (2020). Variation in training load and markers of wellness. *Journal of Australian Stength & Conditioning, 27*(3), 6-14.

McGuinness, A., McMahon, G., Malone, S., Kenna, D., Passmore, D., & Collins, K. (2018). Monitoring wellness, training load, and running performance during a major international female field hockey tournament. *Journal of Strength & Conditioning Research*. doi:10.1519/jsc.0000000000002835

McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., & Weston, M. (2018). The relationships between internal and external measures of training load and intensity in team sports: A meta-analysis. *Sports Medicine, 48*(3), 641-658. doi:10.1007/s40279-017-0830-z

McLean, B. D., Coutts, A. J., Kelly, V., McGuigan, M. R., & Cormack, S. J. (2010). Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *International Journal of Sports Physiology and Performance, 5*(3), 367-383.

McNair, D. M., Lorr, M., & Droppleman, L. (1971). *Manual for the profile of mood state.* San Diego, CA: Educational and Industrial Testing Service.

Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., . . . Urhausen, A. (2013). Prevention, diagnosis, and treatment of the overtraining syndrome: Joint consensus statement of the european college of sport science and the american college of sports medicine. *Medicine & Science in Sports Exercise, 45*(1), 186-205. doi:10.1249/MSS.0b013e318279a10a

Meeuwisse, W. (1994). Assessing causation in sport injury: A multifactorial model. *Clinical Journal of Sports Medicine, 4*, 166-170.

Meeuwisse, W. H. (1991). Predictability of sports injuries. What is the epidemiological evidence? *Sports Medicine, 12*(1), 8-15.

Meeuwisse, W. H., Tyreman, H., Hagel, B., & Emery, C. (2007). A dynamic model of etiology in sport injury: The recursive nature of risk and causation. *Clinical Journal of Sports Medicine, 17*(3), 215-219. doi:10.1097/JSM.0b013e3180592a48

Melvin, J. W. (1993). Fracture mechanics of bone. *Journal of Biomechanical Engineering, 115*(4b), 549-554. doi:10.1115/1.2895538

Mendes, B., Palao, J. M., Silvério, A., Owen, A., Carriço, S., Calvete, F., & Clemente, F. M. (2018). Daily and weekly training load and wellness status in preparatory, regular and congested weeks: A season-long study in elite volleyball players. *Research in Sports Medicine, 26*(4), 462-473. doi:10.1080/15438627.2018.1492393

Mester, J., & Perl, J. (2000). Limits of human adaptability and performance from a systemic point of view: Time series analyzes and an it metamodel for investigating physiological adaptation processes. *Competition & Sports, 30*(1), 43-51.

Michielsen, H. J., De Vries, J., Heck, G. L., Van de Vijver, F., & Sijtsma, K. (2004). Examination of the dimensionality of fatigue: The construction of the fatigue assessment scale (fas). *European Journal of Psychological Assessment, 20*, 39. doi:10.1027//1015-5759.20.1.39

Milton, K., Bull, F. C., & Bauman, A. (2011). Reliability and validity testing of a single-item physical activity measure. *British Journal of Sports Medicine, 45*(3), 203-208. doi:10.1136/bjsm.2009.068395

Miranda Rohlfs, I., Rotta, T., Bernardi Luft, C., Andrade, A., Krebs, R., & Carvalho, T. (2008). A escala de humor de brunel (brums): Instrumento para detecção precoce da síndrome do excesso de treinamento. *Revista Brasileira de Medicina do Esporte, 14*(3), 176-181.

Miranda Rohlfs, I., Terry, P., Rotta, T., Luft, C., Andrade, A., Krebs, R., . . . Iizuka, C. (2008). *Development and initial validation of the brazil mood scale.* Paper presented at the Proceedings of the 43rd Australian Psychological Society Annual Conference: Psychology Leading Change, Australia.

Mitchell, A. J. (2007). Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders. *Journal of Clinical Oncology, 25*(29), 4670-4681. doi:10.1200/jco.2006.10.0438

Moalla, W., Fessi, M. S., Farhat, F., Nouira, S., Wong, D. P., & Dupont, G. (2016). Relationship between daily training load and psychometric status of professional soccer players. *Research in Sports Medicine, 24*(4), 387-394. doi:10.1080/15438627.2016.1239579

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *Journal of Clinical Epidemiology, 62*(10), 1006-1012. doi:10.1016/j.jclinepi.2009.06.005

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . De Vet, H. (2012). The cosmin checklist manual. Retrieved from http://www.cosmin.nl////

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). The cosmin checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. *Quality of Life, 19*(4), 539-549. doi:10.1007/s11136-010-9606-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). The cosmin study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745. doi:10.1016/j.jclinepi.2010.02.006

Moller, M., Nielsen, R. O., Attermann, J., Wedderkopp, N., Lind, M., Sorensen, H., & Myklebust, G. (2017). Handball load and shoulder injury rate: A 31-week cohort study of 679 elite youth handball players. *British Journal of Sports Medicine, 51*(4), 231-237. doi:10.1136/bjsports-2016-096927

Moran, J., Paxton, K., Jones, B., Granacher, U., Sandercock, G. R., Hope, E., & Ramirez-Campillo, R. (2020). Variable long-term developmental trajectories of short sprint speed and jumping height in english premier league academy soccer players: An applied case study. *Journal of Sports Sciences*, 1-7. doi:10.1080/02640414.2020.1792689

Moreira, A., Arsati, F., De Oliveira Lima-Arsati, Y. B., Simões, A. C., & De Araújo, V. C. (2011). Monitoring stress tolerance and occurrences of upper respiratory illness in basketball players by means of psychometric tools and salivary biomarkers. *Stress and Health, 27*(3), e166-e172. doi:10.1002/smi.1354

Morgan, W. P., Brown, D. R., Raglin, J. S., O'Connor, P. J., & Ellickson, K. A. (1987). Psychological monitoring of overtraining and staleness. *British Journal of Sports Medicine, 21*(3), 107-114.

Morton, R. H., Fitz-Clarke, J. R., & Banister, E. W. (1990). Modeling human performance in running. *Journal of Applied Physiology, 69*(3), 1171-1177. doi:10.1152/jappl.1990.69.3.1171

Motta-Valencia, K. (2006). Dance-related injury. *Physical Medicine and Rehabilitation Clinics of North America, 17*(3), 697-723. doi:10.1016/j.pmr.2006.06.001

Mountjoy, M., Junge, A., Alonso, J. M., Engebretsen, L., Dragan, I., Gerrard, D., . . . Dvorak, J. (2010). Sports injuries and illnesses in the 2009 fina world championships (aquatics). *British Journal of Sports Medicine, 44*(7), 522-527. doi:10.1136/bjsm.2010.071720

Myers, N. L., Aguilar, K. V., Mexicano, G., Farnsworth, J. L., 2nd, Knudson, D., & Kibler, W. B. (2020). The acute: Chronic workload ratio is associated with injury in junior tennis players. *Medicine & Science in Sports & Exercise, 52*(5), 1196-1200. doi:10.1249/mss.0000000000002215

Nascimento, A. L., Luna, J. V., & Fontenelle, L. F. (2012). Body dysmorphic disorder and eating disorders in elite professional female ballet dancers. *Annals of Clinical Psychiatry, 24*(3), 191-194.

Nassi, 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. doi:10.1080/17461391.2017.1318180

Nédélec, M., Halson, S., Delecroix, B., Abaidia, A. E., Ahmaidi, S., & Dupont, G. (2015). Sleep hygiene and recovery strategies in elite soccer players. *Sports Medicine, 45*(11), 1547-1559. doi:10.1007/s40279-015-0377-9

Nederhof, E., Brink, M. S., & Lemmink, K. (2008). Reliability and validity of the dutch recovery stress questionnaire for athletes. *International Journal of Sport Psychology 39*(4), 301-311.

Nicholls, A. R., Backhouse, S. H., Polman, R. C. J., & McKenna, J. (2009). Stressors and affective states among professional rugby union players. *Scandinavian Journal of Medicine and Science in Sports, 19*(1), 121-128. doi:10.1111/j.1600-0838.2007.00757.x

Nicolas, M., Vacher, P., Martinent, G., & Mourot, L. (2019). Monitoring stress and recovery states: Structural and external stages of the short version of the restq sport in elite swimmers before championships. *Journal of Sport & Health Science, 8*(1), 77-88. doi:10.1016/j.jshs.2016.03.007

Nilsson, C., Leanderson, J., Wykman, A., & Strender, L. E. (2001). The injury panorama in a swedish professional ballet company. *Knee Surgery, Sports Traumatology, Arthroscopy, 9*(4), 242-246.

Noh, Y. E., Morris, T., & Andersen, M. (2011). Psychosocial factors and ballet injuries. *International Journal of Sport & Exercise Psychology, 3*(1), 79-90. doi:10.1080/1612197X.2005.9671759

Noon, M. R., James, R. S., Clarke, N. D., Akubat, I., & Thake, C. D. (2015). Perceptions of well-being and physical performance in english elite youth footballers across a season. *Journal of Sports Sciences, 33*(20), 2106-2115. doi:10.1080/02640414.2015.1081393

Nosek, P., Brownlee, T. E., Drust, B., & Andrew, M. (2020). Feedback of gps training data within professional english soccer: A comparison of decision making and perceptions between coaches, players and performance staff. *Science & Medicine in Football*, 1-13. doi:10.1080/24733938.2020.1770320

Novas, A. M. P., Rowbottom, D. G., & Jenkins, D. G. (2003). Tennis, incidence of urti and salivary iga. *International Journal of Sports Medicine, 24*(3), 223-229. doi:10.1055/s-2003-39096

O'Neill, J. R., Pate, R. R., & Liese, A. D. (2011). Descriptive epidemiology of dance participation in adolescents. *Research Quarterly for Exercise and Sport, 82*(3), 373-380. doi:10.1080/02701367.2011.10599769

Oliveira, R., Brito, J. P., Martins, A., Mendes, B., Marinho, D. A., Ferraz, R., & Marques, M. C. (2019). In-season internal and external training load quantification of an elite european soccer team. *PLoS One, 14*(4), e0209393. doi:10.1371/journal.pone.0209393

Oliver, J. L., Lloyd, R. S., & Whitney, A. (2015). Monitoring of in-season neuromuscular and perceptual fatigue in youth rugby players. *European Journal of Sport Science, 15*(6), 514-522.

Olson, K. (2007). A new way of thinking about fatigue: A reconceptualization. *Oncology Nursing Forum, 34*(1), 93-99. doi:10.1188/07.onf.93-99

Orchard, J., & Hoskins, W. (2007). For debate: Consensus injury definitions in team sports should focus on missed playing time. *Clinical Journal of Sports Medicine, 17*(3), 192-196. doi:10.1097/JSM.0b013e3180547527

Osiecki, R., Rubio, T., Coelho, R., Novack, L., Conde, J., Alves, C., & Malfatti, C. (2015). The total quality recovery scale (tqr) as a proxy for determining athletes' recovery state after a professional soccer match. *Journal of Exercise Physiology Online, 18*, 27-32.

Ostenberg, A., & Roos, H. (2000). Injury risk factors in female european football. A prospective study of 123 players during one season. *Scandinavian Journal of Medicine & Science in Sports, 10*(5), 279-285. doi:10.1034/j.1600-0838.2000.010005279.x

Patrician, P. A. (2004). Single-item graphic representational scales. *Nursing Research, 53*(5), 347-352. doi:10.1097/00006199-200409000-00011

Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (pro) instruments for medical product evaluation: Ispor pro good research practices task force report: Part 1--eliciting concepts for a new pro instrument. *Value Health, 14*(8), 967-977. doi:10.1016/j.jval.2011.06.014

Pattyn, N., Van Cutsem, J., Dessy, E., & Mairesse, O. (2018). Bridging exercise science, cognitive psychology, and medical practice: Is "cognitive fatigue" a remake of "the emperor's new clothes"? *Frontiers in Psychology, 9*, 1246-1246. doi:10.3389/fpsyg.2018.01246

Peake, J. M., Neubauer, O., Della Gatta, P. A., & Nosaka, K. (2017). Muscle damage and inflammation during recovery from exercise. *Journal of Applied Physiology, 122*(3), 559-570. doi:10.1152/japplphysiol.00971.2016

Pedrosa, I., & García-Cueto, E. (2012). Adaptación al español del athlete burnout questionnaire (abq) en una muestra española de deportistas - spanish adaptation of the athlete burnout questionnaire (abq) in a spanish athlete sample.. *Ansiedad y estrés, 18*, 155-166.

Pelka, M., Ferrauti, A., Meyer, T., Pfeiffer, M., & Kellmann, M. (2017). How does a short, interrupted recovery break affect performance and how is it assessed? A study on acute effects. *International Journal of Sports Physiology & Performance, 12*(Suppl 2), S2114-S2121. doi:10.1123/ijspp.2016-0399

Pelka, M., Kolling, 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*(3), 216-223. doi:10.1080/02640414.2016.1161208

Perl, J. (2001). Perpot: A metamodel for simulation of load performance interaction. *European Journal of Sport Science, 1*(2), 1-13. doi:10.1080/17461390100071202

Perl, J. (2004). Perpot – a meta-model and software tool for analysis and optimisation of load-performance-interaction. *International Journal of Performance Analysis in Sport, 4*(2), 61-73. doi:10.1080/24748668.2004.11868305

Pinto, M., Kuhn, J. E., Greenfield, M. L., & Hawkins, R. J. (1999). Prospective analysis of ice hockey injuries at the junior a level over the course of one season. *Clinical Journal of Sports Medicine, 9*(2), 70-74.

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. *Sports Medicine, 43*(9), 773-781. doi:10.1007/s40279-013-0071-8

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). Cosmin guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. doi:10.1007/s11136-018-1798-3

Puta, C., Steidten, T., Baumbach, P., Wohrl, T., May, R., Kellmann, M., . . . Gabriel, H. H. W. (2018). Standardized assessment of resistance training-induced subjective symptoms and objective signs of immunological stress responses in young athletes. *Frontiers in Physiology, 9*, 698. doi:10.3389/fphys.2018.00698

Putlur, P., Foster, C., Miskowski, J. A., Kane, M. K., Burton, S. E., Scheett, T. P., & McGuigan, M. R. (2004). Alteration of immune function in women collegiate soccer players and college students. *Journal of Sports Science & Medicine, 3*(4), 234-243.

R Core Team. (2019). R: A language and environment for statistical computing. R foundation for statistical computing. v3.6.2. Retrieved from https://www.R-project.org

Rabbani, A., Baseri, M. K., Reisi, J., Clemente, F. M., & Kargarfard, M. (2018). Monitoring collegiate soccer players during a congested match schedule: Heart rate variability versus subjective wellness measures. *Physiology and Behavior, 194*, 527-531. doi:10.1016/j.physbeh.2018.07.001

Rabbani, A., Clemente, F. M., Kargarfard, M., & Chamari, K. (2019). Match fatigue time-course assessment over four days: Usefulness of the hooper index and heart rate variability in professional soccer players. *Frontiers in Physiology, 10*, 109. doi:10.3389/fphys.2019.00109

Raedeke, T., Arce, C., De Francisco, C., Seoane, G., & Ferraces, M. (2013). The construct validity of the spanish version of the abq using a multi-trait/multi-method approach. *Anales de Psicología, 29*, 693-700.

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

Raeder, C., Wiewelhove, T., Simola, R. Á., Kellmann, M., Meyer, T., Pfeiffer, M., & Ferrauti, A. (2016). Assessment of fatigue and recovery in male and female athletes after 6 days of intensified strength training. *Journal of Strength and Conditioning Research, 30*(12), 3412-3427.

Rafferty, S. (2010). Considerations for integrating fitness into dance training. *Journal of Dance Medicine and Science, 14*(2), 45-49.

Raglin, J. S., & Morgan, W. P. (1989). Development of a scale to measure training-induced distress. *Medicine and Science in Sports and Exercise, 21*(Supplemental), s60.

Ramel, E., & Moritz, U. (1998). Psychosocial factors at work and their association with professional ballet dancers' musculoskeletal disorders. *Medical Problems of Performing Artists, 13*, 66-74.

Reid, D. C., Burnham, R. S., Saboe, L. A., & Kushner, S. F. (1987). Lower extremity flexibility patterns in classical ballet dancers and their correlation to lateral hip and knee injuries. *American Journal of Sports Medicine, 15*, 347–352.

Reid, V. L., Gleeson, M., Williams, N., & Clancy, R. L. (2004). Clinical investigation of athletes with persistent fatigue and/or recurrent infections. *British Journal of Sports Medicine, 38*(1), 42-45. doi:10.1136/bjsm.2002.002634

Reissmann, D. R., John, M. T., Kriston, L., & Schierz, O. (2013). Insufficient diagnostic accuracy of a single-item questionnaire to detect psychosocial distress in temporomandibular disorder patients. *Clinical Oral Investigations, 17*(8), 1937-1945. doi:10.1007/s00784-012-0892-0

Ringham, R., Klump, K., Kaye, W., Stone, D., Libman, S., Stowe, S., & Marcus, M. (2006). Eating disorder symptomatology among ballet dancers. *International Journal of Eating Disorders, 39*(6), 503-508. doi:10.1002/eat.20299

Robey, E., Dawson, B., Halson, S., Gregson, W., Goodman, C., & Eastwood, P. (2014). Sleep quantity and quality in elite youth soccer players: A pilot study. *European Journal of Sport Science, 14*(5), 410-417. doi:10.1080/17461391.2013.843024

Roe, G., Darrall-Jones, J., Till, K., Phibbs, P., Read, D., Weakley, J., & Jones, B. (2016). Between-days reliability and sensitivity of common fatigue measures in rugby players. *International Journal of Sports Physiology and Performance, 11*(5), 581-586. doi:10.1123/ijspp.2015-0413

Rogalski, B., Dawson, B., Heasman, J., & Gabbett, T. J. (2013). Training and game loads and injury risk in elite australian footballers. *Journal of Science and Medicine in Sport., 16*(6), 499-503. doi:10.1016/j.jsams.2012.12.004

Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods, 10*, 195–209.

Romani, W. A., Gieck, J. H., Perrin, D. H., Saliba, E. N., & Kahler, D. M. (2002). Mechanisms and management of stress fractures in physically active persons. *Journal of Athletic Training, 37*(3), 306-314.

Romdhani, M., Hammouda, O., Chaabouni, Y., Mahdouani, K., Driss, T., Chamari, K., & Souissi, N. (2019). Sleep deprivation affects post-lunch dip performances, biomarkers of muscle damage and antioxidant status. *Biology of Sport 36*(1), 55-65. doi:10.5114/biolsport.2018.78907

Rothman, M. L., Beltran, P., Cappelleri, J. C., Lipscomb, J., & Teschendorf, B. (2007). Patient-reported outcomes: Conceptual issues. *Value in Health, 10 Suppl 2*, S66-75. doi:10.1111/j.1524-4733.2007.00269.x

Rowan, F. E., Benjamin-Laing, H., Kennedy, A., De Medici, A., Beasley, I., & Haddad, F. S. (2019). Self-directed oral vitamin d supplementation in professional ballet dancers: A randomized controlled trial pilot study. *Journal of Dance Medicine in Science, 23*(3), 91-96. doi:10.12678/1089-313x.23.3.91

Rushall, B. S. (1990). A tool for measuring stress tolerance in elite athletes. *Journal of Applied Sport Psychology, 2*(1), 51-66. doi:10.1080/10413209008406420

Ryan, S., Coutts, A. J., Hocking, J., Dillon, P. A., Whitty, A., & Kempton, T. (2018). Physical preparation factors that influence technical and physical match performance in professional australian football. *International Journal of Sports Physiology and Performance, 13*(8), 1021-1027. doi:10.1123/ijspp.2017-0640

Ryan, S., Kempton, T., Impellizzeri, F., & Coutts, A. (2019). Training monitoring in professional australian football: Theoretical basis and recommendations for coaches and scientists. *Science & Medicine in Football*, 1-7. doi:10.1080/24733938.2019.1641212

Sargent, C., Lastella, M., Halson, S. L., & Roach, G. D. (2014). The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiology International, 31*(10), 1160-1168. doi:10.3109/07420528.2014.957306

Sassi, A. (1997). *Allenamento e sovrallenamento*. MIlano, Italy: EdiErmes.

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 & Performance, 12*(Suppl 2), S2127-s2135. doi:10.1123/ijspp.2016-0395

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine, 50*(5), 281-291. doi:10.1136/bjsports-2015-094758

Sawczuk, T., Jones, B., Scantlebury, S., & Till, K. (2018). The influence of training load, exposure to match play and sleep duration on daily wellbeing measures in youth athletes. *Journal of Sports Sciences, 36*(21), 2431-2437. doi:10.1080/02640414.2018.1461337

Schafle, M. D. (1990). The child dancer. Medical considerations. *Pediatric Clinics of North America, 37*(5), 1211-1221. doi:10.1016/s0031-3955(16)36984-x

Schantz, P. G., & Astrand, P. O. (1984). Physiological characteristics of classical ballet. *Medicine and Science in Sports and Exercise, 16*(5), 472-476.

Schimpchen, J., Wagner, M., Ferrauti, A., Kellmann, M., Pfeiffer, M., & Meyer, T. (2017). Can cold water immersion enhance recovery in elite olympic weightlifters? An individualized perspective. *Journal of Strength and Conditioning Research, 31*(6), 1569-1576. doi:10.1519/JSC.0000000000001591

Schnelle, J. F., Buchowski, M. S., Ikizler, T. A., Durkin, D. W., Beuscher, L., & Simmons, S. F. (2012). Evaluation of two fatigability severity measures in elderly adults. *Journal of the American Geriatrics Society, 60*(8), 1527-1533. doi:10.1111/j.1532-5415.2012.04062.x

Schwellnus, M., Derman, W., Lambert, M., Redhead, C., & Page, T. (2011). Epidemiology of illness during the super 14 rugby tournament – a prospective cohort study. *British Journal of Sports Medicine, 45*(4), 316-317. doi:10.1136/bjsm.2011.084038.20

Schwellnus, M., Soligard, T., Alonso, J. M., Bahr, R., Clarsen, B., Dijkstra, H. P., . . . Engebretsen, L. (2016). How much is too much? (part 2) international olympic committee consensus statement on load in sport and risk of illness. *British Journal of Sports Medicine, 50*(17), 1043-1052. doi:10.1136/bjsports-2016-096572

Schwellnus, M. P., Derman, W. E., Jordaan, E., Page, T., Lambert, M. I., Readhead, C., . . . Webb, S. (2012). Elite athletes travelling to international destinations >5 time zone differences from their home country have a 2-3-fold increased risk of illness. *British Journal of Sports Medicine, 46*(11), 816-821. doi:10.1136/bjsports-2012-091395

Selmi, O., Goncalves, B., Ouergui, I., Levitt, D. E., Sampaio, J., & Bouassida, A. (2019). Influence of well-being indices and recovery state on the technical and physiological aspects of play during small-sided games. *Journal of Strength & Conditioning Research*. doi:10.1519/JSC.0000000000003228

Service/EDITS, E. a. I. T. (1999). *Edits research and developments*. San Diego, CA: Educational and Industrial Testing Service.

Shacham, S. (1983). A shortened version of the profile of mood states. *Journal of Personality Assessment, 47*(3), 305-306. doi:10.1207/s15327752jpa4703_14

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *British Medical Journal, 350*, g7647. doi:10.1136/bmj.g7647

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Šišková, M., Hudáková, M., & Sollár, T. (2017). *Psychometric properties of the slovak version of the athlete burnout questionnaire (abq)*. Paper presented at the Czech-Slovak Psychological Conference (not only) for PhD and PhD students, Olomouc; Czech Republic.

Sloan, J. A., Aaronson, N., Cappelleri, J. C., Fairclough, D. L., Varricchio, C., & Clinical Significance Consensus Meeting, G. (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proceedings, 77*(5), 479-487.

Smoliga, J. M., & Zavorsky, G. S. (2017). Team logo predicts concussion risk: Lessons in protecting a vulnerable sports community from misconceived, but highly publicized epidemiologic research. *Epidemiology., 28*(5), 753-757. doi:10.1097/ede.0000000000000694

Soligard, T., Schwellnus, M., Alonso, J. M., Bahr, R., Clarsen, B., Dijkstra, H. P., . . . Engebretsen, L. (2016). How much is too much? (part 1) international olympic committee consensus statement on load in sport and risk of injury. *British Journal of Sports Medicine, 50*(17), 1030-1041. doi:10.1136/bjsports-2016-096581

Soligard, T., Steffen, K., Palmer-Green, D., Aubry, M., Grant, M. E., Meeuwisse, W., . . . Engebretsen, L. (2015). Sports injuries and illnesses in the sochi 2014 olympic winter games. *British Journal of Sports Medicine, 49*(7), 441-447. doi:10.1136/bjsports-2014-094538

Soos, I., Lane, A., Leibinger, E., Istvan, K., & Hamar, P. (2005). *Validity of the profile of mood states (sf) among hungarian students*. Paper presented at the ISSP.

Sparkes, W., Turner, A. N., Weston, M., Russell, M., Johnston, M. J., & Kilduff, L. P. (2020). The effect of training order on neuromuscular, endocrine and mood response to small-sided games and resistance training sessions over a 24-h period. *Journal of Science & Medicine in Sport, 23*(9), 866-871.

Speirs, D. E., Bennett, M. A., Finn, C. V., & Turner, A. P. (2016). Unilateral vs. Bilateral squat training for strength, sprints, and agility in academy rugby players. *Journal of Strength & Conditioning Research, 30*(2), 386-392. doi:10.1519/jsc.0000000000001096

Spence, L., Brown, W. J., Pyne, D. B., Nissen, M. D., Sloots, T. P., McCormack, J. G., . . . Fricker, P. A. (2007). Incidence, etiology, and symptomatology of upper respiratory illness in elite athletes. *Medicine & Science in Sports & Exercise, 39*(4), 577-586.

Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: Implications for training prescription. *Sports Medicine., 43*(12), 1259-1277. doi:10.1007/s40279-013-0083-4

Stares, J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., Drew, M., . . . Lester, L. (2018). Identifying high risk loading conditions for in-season injury in elite australian football players. *Journal of Science & Medicine in Sport, 21*(1), 46-51. doi:10.1016/j.jsams.2017.05.012

Steinberg, N., Siev-Ner, I., Peleg, S., Dar, G., Masharawi, Y., Zeev, A., & Hershkovitz, I. (2012). Extrinsic and intrinsic risk factors associated with injuries in young dancers aged 8-16 years. *Journal of Sports Sciences, 30*(5), 485-495. doi:10.1080/02640414.2011.647705

Stephan, A., Mainzer, J., Kummel, D., & Impellizzeri, F. M. (2019). Measurement properties of promis short forms for pain and function in orthopedic foot and ankle surgery patients. *Quality of Life Research, 28*(10), 2821-2829. doi:10.1007/s11136-019-02221-w

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press.

Svendsen, I. S., Taylor, I. M., Tonnessen, E., Bahr, R., & Gleeson, M. (2016). Training-related and competition-related risk factors for respiratory tract and gastrointestinal infections in elite cross-country skiers. *British Journal of Sports Medicine, 50*(13), 809-815. doi:10.1136/bjsports-2015-095398

Tanner, R. K., Gore, C. J., & Australian Institute of Sport. (2013). *Physiological tests for elite athletes*. Champaign, IL: Human Kinetics.
Terry, P. C., & Lane, A. M. (2000). Normative values for the profile of mood states for use with athletic samples. *Journal of Applied Sport Psychology, 12*, 93-109.
Terry, P. C., Lane, A. M., & Fogarty, G. J. (2003). Construct validity of the poms-a for use with adults. *Psychology of Sport and Exercise, 4*(2), 125-139.
Terry, P. C., Lane, A. M., Lane, H. J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. *Journal of Sports Sciences, 17*(11), 861-872. doi:10.1080/026404199365425
Terry, P. C., & Youngs, E. L. (1996). Discriminant effectivness of psychological state measures in predicting selection during field hockey trials. *Perceptual and Motor Skills, 82*(2), 371-377. doi:10.2466/pms.1996.82.2.371
Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012
Terwee, C. B., Jansma, E. P., Riphagen, II, & de Vet, H. C. (2009). Development of a methodological pubmed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115-1123. doi:10.1007/s11136-009-9528-5
Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the cosmin checklist. *Quality of Life Research, 21*(4), 651-657. doi:10.1007/s11136-011-9960-1
Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., . . . Mokkink, L. B. (2018). Cosmin methodology for evaluating the content validity of patient-reported outcome measures: A delphi study. *Quality of Life Research, 27*(5), 1159-1170. doi:10.1007/s11136-018-1829-0
Thomas, J. J., Keel, P. K., & Heatherton, T. F. (2005). Disordered eating attitudes and behaviors in ballet students: Examination of environmental and individual risk factors. *International Journal of Eating Disorders, 38*(3), 263-268. doi:10.1002/eat.20185
Thomas, K., Brownstein, C. G., Dent, J., Parker, P., Goodall, S., & Howatson, G. (2018). Neuromuscular fatigue and recovery after heavy resistance, jump, and sprint training. *Medicine & Science in Sports & Exercise, 50*(12), 2526-2535. doi:10.1249/mss.0000000000001733
Thornton, H. R., Delaney, J. A., Duthie, G. M., & Dascombe, B. J. (2017). Importance of various training-load measures in injury incidence of professional rugby league athletes. *International Journal of Sports Physiology & Performance, 12*(6), 819-824. doi:10.1123/ijspp.2016-0326
Thornton, H. R., Delaney, J. A., Duthie, G. M., Scott, B. R., Chivers, W. J., Sanctuary, C. E., & Dascombe, B. J. (2016). Predicting self-reported illness for professional team-sport athletes. *International Journal of Sports Physiology & Performance, 11*(4), 543-550.
Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2016). Tracking morning fatigue status across in-season training weeks in elite soccer players. *International Journal of Sports Physiology & Performance, 11*(7), 947-952.
Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2017). The influence of changes in acute training load on daily sensitivity of morning-measured fatigue variables in elite soccer players. *International Journal of Sports Physiology & Performance, 12*(Suppl 2), S2107-S2113. doi:10.1123/ijspp.2016-0433
Tibana, R. A., Sousa, N. M. F., Prestes, J., Feito, Y., Ferreira, C. E., & Voltarelli, F. A. (2019). Monitoring training load, well-being, heart rate variability, and competitive performance of a functional-fitness female athlete: A case study. *Sports (Basel), 7*(2). doi:10.3390/sports7020035

Timoteo, T. F., Debien, P. B., Miloski, B., Werneck, F. Z., Gabbett, T., & Bara Filho, M. G. (2018). Influence of workload and recovery on injuries in elite male volleyball players. *Journal of Strength & Conditioning Research*. doi:10.1519/JSC.0000000000002754

Timpka, T., Alonso, J. M., Jacobsson, J., Junge, A., Branco, P., Clarsen, B., . . . Edouard, P. (2014). Injury and illness definitions and data collection procedures for use in epidemiological studies in athletics (track and field): Consensus statement. *British Journal of Sports Medicine, 48*(7), 483-490. doi:10.1136/bjsports-2013-093241

Tofari, P. J., Kemp, J. G., & Cormack, S. J. (2020). Measuring the response to simulated fixture congestion in soccer. *Science & Medicine in Football*, 1-12. doi:10.1080/24733938.2020.1746824

Tolfrey, K., Hansen, S. A., Dutton, K., McKee, T., & Jones, A. M. (2009). Physiological correlates of 2-mile run performance as determined using a novel on-demand treadmill. *Applied Physiology, Nutrition, and Metabolism, 34*(4), 763-772. doi:10.1139/h09-069

Tropp, H., Ekstrand, J., & Gillquist, J. (1984). Stabilometry in functional instability of the ankle and its value in predicting injury. *Medicine & Science in Sports & Exercise, 16*(1), 64-66.

Twitchett, E., Angioi, M., Koutedakis, Y., & Wyon, M. (2010). The demands of a working day among female professional ballet dancers. *Journal of Dance Medicine and Science, 14*(4), 127-132.

Twitchett, E., Brodrick, A., Nevill, A., Koutedakis, Y., Angioi, M., & Wyon, M. (2010). Does physical fitness affect injury occurrence and time loss due to injury in elite vocational ballet students? *Journal of Dance Medicine & Science, 14*, 26-31.

Twitchett, E. A., Koutedakis, Y., & Wyon, M. A. (2009). Physiological fitness and professional classical ballet performance: A brief review. *Journal of Strength and Conditioning Research, 23*(9), 2732-2740. doi:10.1519/JSC.0b013e3181bc1749

U.S.Department of Health and Human ServicesFood and Drug Administration (FDA), C. f. D. E. a. R. C., Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). (2009). *Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims*. Retrieved from Rockville, MD:

Urhausen, A., Gabriel, H. H. W., Weiler, B., & Kindermann, W. (1998). Ergometric and psychological findings during overtraining: A long-term follow-up study in endurance athletes. *International Journal of Sports Medicine, 19*(2), 114-120. doi:10.1055/s-2007-971892

van Hooff, M. L., Geurts, S. A., Kompier, M. A., & Taris, T. W. (2007). "How fatigued do you currently feel?" Convergent and discriminant validity of a single-item fatigue measure. *Journal of Occupational Health, 49*(3), 224-234. doi:10.1539/joh.49.224

van Smeden, M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology, 16*(1), 163. doi:10.1186/s12874-016-0267-3

VanderWeele, T. J., & Robins, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology, 18*(5), 561-568. doi:10.1097/EDE.0b013e318127181b

Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A., & Drust, B. (2017). Training load monitoring in team sports: A novel framework separating physiological and biomechanical load-adaptation pathways. *Sports Medicine, 47*(11), 2135-2142. doi:10.1007/s40279-017-0714-2

Verschueren, J., Tassignon, B., De Pauw, K., Proost, M., Teugels, A., Van Cutsem, J., . . . Meeusen, R. (2020). Does acute fatigue negatively affect intrinsic risk factors of the lower extremity injury risk profile? A systematic and critical review. *Sports Medicine, 50*(4), 767-784. doi:10.1007/s40279-019-01235-1

Veugelers, K. R., Young, W. B., Fahrner, B., & Harvey, J. T. (2016). Different methods of training load quantification and their relationship to injury and illness in elite australian

football. *Journal of Science & Medicine in Sport, 19*(1), 24-28. doi:10.1016/j.jsams.2015.01.001
Viana, M. F., de Almeida, P. L., & Santos, R. C. (2001). Adaptação portuguesa da versão reduzida do perfil de estados de humor–poms = portuguese adaptation for the reduced version of the profile of mood states (poms). *Análise Psicológica, 19*(1), 77-92.
Victora, C. G., Huttly, S. R., Fuchs, S. C., & Olinto, M. T. (1997). The role of conceptual frameworks in epidemiological analysis: A hierarchical approach. *International Journal of Epidemiology, 26*(1), 224-227. doi:10.1093/ije/26.1.224
Viru, A., & Viru, M. (2000). *Nature of training effects*. Philadelphia: Lippincott Williams and Wilkins.
Vitale, J. A., Banfi, G., Galbiati, A., Ferini-Strambi, L., & La Torre, A. (2019). Effect of a night game on actigraphy-based sleep quality and perceived recovery in top-level volleyball athletes. *International Journal of Sports Physiology & Performance, 14*(2), 265-269. doi:10.1123/ijspp.2018-0194
von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotzsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2007). The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies. *PLoS Medicine, 4*(10), e296. doi:10.1371/journal.pmed.0040296
Walsh, N. P., Gleeson, M., Pyne, D. B., Nieman, D. C., Dhabhar, F. S., Shephard, R. J., . . . Kajeniene, A. (2011). Position statement. Part two: Maintaining immune health. *Exercise Immunolgy Review, 17*, 64-103.
Walsh, N. P., Gleeson, M., Shephard, R. J., Gleeson, M., Woods, J. A., Bishop, N. C., . . . Simon, P. (2011). Position statement. Part one: Immune function and exercise. *Exercise Immunoloy Review, 17*, 6-63.
Wiesler, E. R., Hunter, D. M., Martin, D. F., Curl, W. W., & Hoen, H. (1996). Ankle flexibility and injury patterns in dancers. *American Journal of Sports Medicine, 24*(6), 754-757.
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 & Performance, 11*(8), 1060-1066. doi:10.1123/ijspp.2015-0494
Wiewelhove, T., Schneider, C., Doweling, A., Hanakam, F., Rasche, C., Meyer, T., . . . Ferrauti, A. (2018). Effects of different recovery strategies following a half-marathon on fatigue markers in recreational runners. *PLoS One, 13*(11), e0207313. doi:10.1371/journal.pone.0207313
Wilder, R. P., Greene, J. A., Winters, K. L., Long, W. B., 3rd, Gubler, K., & Edlich, R. F. (2006). Physical fitness assessment: An update. *Journal of Long-Term Effects of Medical Implants, 16*(2), 193-204. doi:10.1615/jlongtermeffmedimplants.v16.i2.90
Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Journal of the American Medical Association, 273*(1), 59-65.
Wolman, R., Wyon, M. A., Koutedakis, Y., Nevill, A. M., Eastell, R., & Allen, N. (2013). Vitamin d status in professional ballet dancers: Winter vs. Summer. *Journal of Science & Medicine in Sport, 16*(5), 388-391. doi:10.1016/j.jsams.2012.12.010
Wyon, M. (2010). Preparing to perform: Periodization and dance. *Journal of Dance Medicine and Science, 14*(2), 67-72.
Wyon, M. A., Koutedakis, Y., Wolman, R., Nevill, A. M., & Allen, N. (2014). The influence of winter vitamin d supplementation on muscle function and injury occurrence in elite ballet dancers: A controlled study. *Journal of Science & Medicine in Sport, 17*(1), 8-12. doi:10.1016/j.jsams.2013.03.007
Zhang, C. Q., Si, G., Chung, P. K., Du, M., & Terry, P. C. (2014). Psychometric properties of the brunel mood scale in chinese adolescents and adults. *Journal of Sports Sciences, 32*(15), 1465-1476. doi:10.1080/02640414.2014.898184
Zinner, C., Pelka, M., Ferrauti, A., Meyer, T., Pfeiffer, M., & Sperlich, B. (2017). Responses of low and high compression during recovery after repeated sprint training in well-trained

handball players. *European Journal of Sport Science, 17*(10), 1304-1310. doi:10.1080/17461391.2017.1380707

Zorrilla, E. P., Luborsky, L., McKay, J. R., Rosenthal, R., Houldin, A., Tax, A., . . . Schmidt, K. (2001). The relationship of depression and stressors to immunological assays: A meta-analytic review. *Brain, Behavior & Immunity, 15*(3), 199-226. doi:10.1006/brbi.2000.0597

Chapter Ten

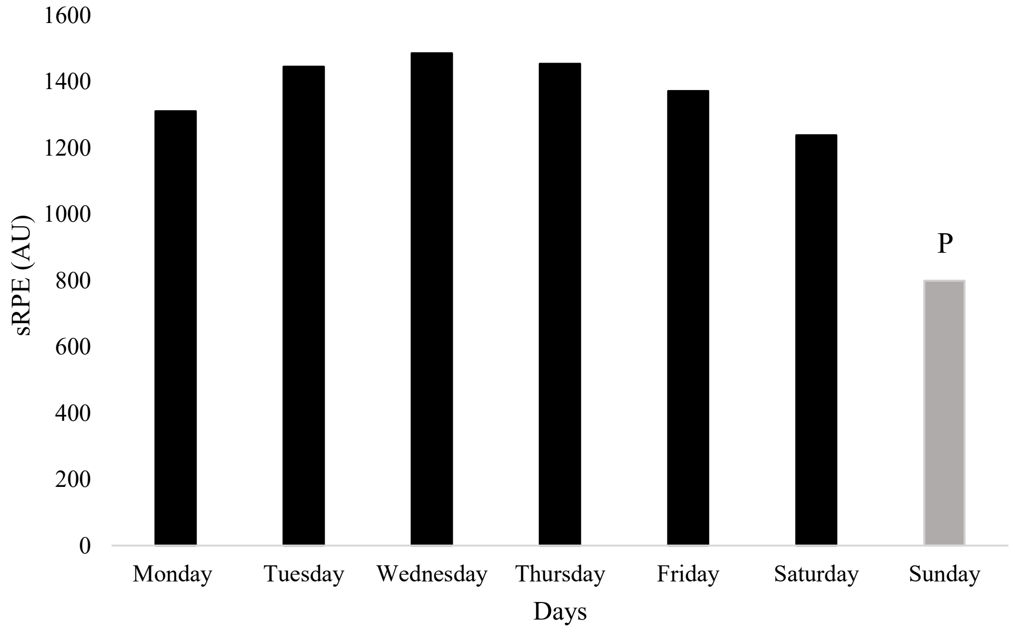
Appendices

Appendix A: Mean session RPE and duration

Session type	Mean ± SD	
	Rating of Perceived exertion[a]	Duration, min (Mean ± SD)
Ballet	5.5 ± 1.4	73.6 ± 6.5
Contemporary	5.0 ± 1.4	73.4 ± 8.3
Rehearsal	5.2 ± 1.9	109.6 ± 34.1
Performance	8.0 ± 1.4	84.9 ± 34.1

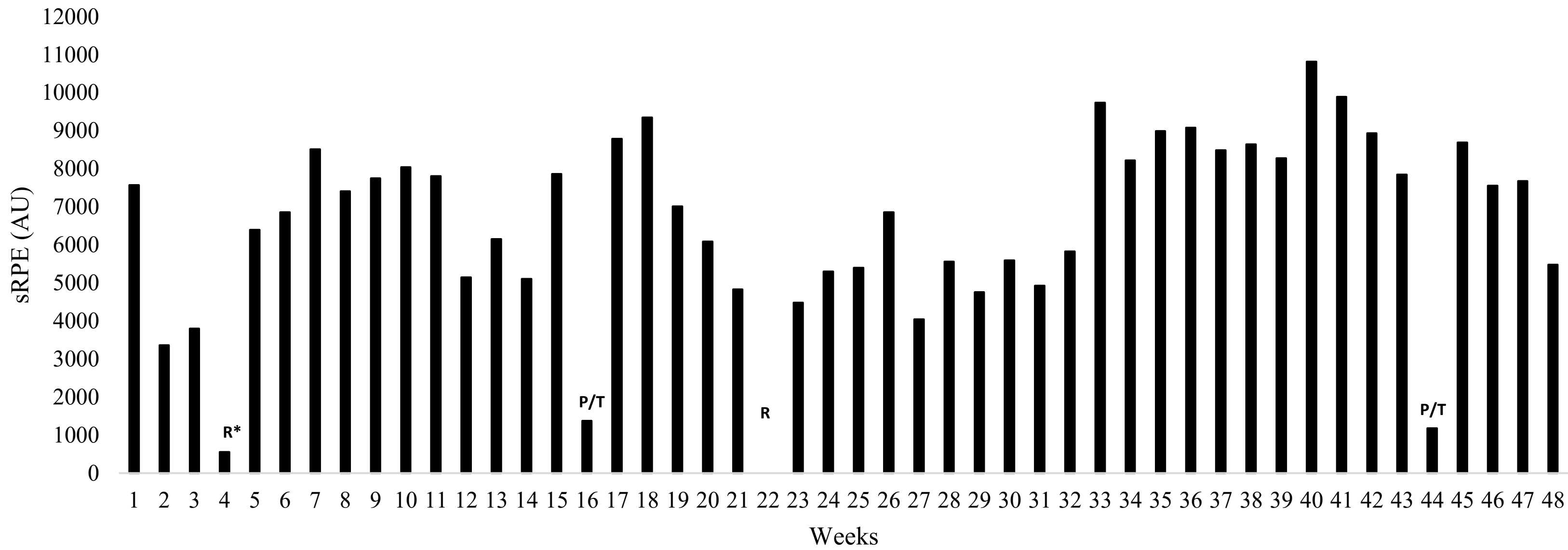
[a]Borg category ratio 10 scale (0= no exertion at all, >10 = absolute maximal)

Appendix B: Mean daily group session RPE



Appendix B: Mean daily group session ratings of perceived exertion (sRPE) based training load. AU: arbitrary units. P: performance only.

Appendix C: Mean weekly group sRPE based training load



Appendix C: Mean weekly group sRPE based training load. R: rest. R*: week 4, one dancer continued rehabilitation & training. P/T: performance/travel

ABQ 12 (Isoard-Gautheur et al., 2010)		DALDA (Rushall, 1990)		DALDA- modified (Nicholls et al., 2009)		PSS (Cohen et al., 1983)		TQR (Merge Kenttä & Hassmén, 1998)		PRSS (Laurent et al., 2011)	
Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
+	High					-	High				
+	High					+	High				
						+	High				
						-	Low				
+	High					+	High				
		+	Low	-	Low			+	Moderate		

POMS 36 (Viana et al., 2001)		POMS 30 (Bourgeois et al., 2010)		BRUMS (Terry et al., 1999)		OFSM (Legros, 1993)		ABQ 15 (Raedeke & Smith, 2001)		ABQ 18 (Arce et al., 2012)	
Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Very low	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
-	Moderate	?	Moderate	-	High			+	High	-	High
+	High	+	High	-	Moderate			+	High	+	High
						-	Very low	?	High	+	Moderate
+	Moderate			+	High			+	High		
				?	Very low						

Appendix D: Quality of the evidence for measurement properties of AROMs

	POMS 65 (McNair et al., 1971)		POMS 63 (Andrade et al., 2010)		POMS 48 (Fernández et al., 2002)		POMS 44 (Andrade et al., 2013)		POMS 40 (Grove & Prapavessis, 1992)	
	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence	Overall rating	Quality of evidence
	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low	+/ - / ?	High, moderate, low, very low
Content Validity	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
Relevance	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
Comprehensiveness	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
Comprehensibility	-	Very low	-	Very low	-	Very low	-	Very low	-	Very low
Structural validity			+	Moderate	+	High	+	High	-	High
Internal consistency			+	High					+	Low
Cross-cultural validity							+	Moderate		
Measurement invariance										
Reliability	-	Very low								
Measurement error										
Construct validity	?	Moderate								
Responsiveness	?	Moderate							?	Very low

“ +” = sufficient, “ - “ = insufficient, “ ±” = inconsistent,” ?” = indeterminate. POMS: Profile of Mood States; BRUMS: Brunel Mood Scale; DALDA: Daily Analysis of Life Demands for Athletes; ABQ: Athlete Burnout Questionnaire; OFSM: Overtraining questionnaire of the French Society for Sports Medicine; PSS: Perceived Stress Scale; PRSS: Perceived Recovery Stress Scale; TQR: Total Quality Recovery Scale.

Appendix E: Results of the Athlete Reported Outcome Measures

To reduce the length of this review we have presented the results for each single AROM in this appendix.

Recovery-Stress Questionnaire Athletes-Sport 76

The content validity was graded inadequate due to the poor description of concept elicitation and lack of clear description of any cognitive interview process. Structural validity was assessed with confirmatory factor analyses in two studies (Davis, Orzeck, & Keelan, 2007; González-Boto, Salguero, Tuero, Márquez, & Kellmann, 2008) and overall rated positive with a high grade of evidence. Additionally, internal consistency was rated positive with a high grade of evidence with most scales (α=>0.70) (Davis et al., 2007; González-Boto et al., 2008; Nederhof, Brink, & Lemmink, 2008). Reliability was examined in one study (Nederhof et al., 2008) and graded very low quality of evidence, as the stability of participants not clearly outlined, conditions of the re-test administration not clearly described and insufficient results for the interclass correlation coefficient (ICC) (ICC=0.34-0.67). Despite no hypothesis reported, there was moderate evidence of construct validity, as demonstrated by very low to moderate correlations (r=0.00-0.60) between subscales of the sport specific RESTQ with relevant subscales of the POMS (Nederhof et al., 2008).

Recovery-Stress Questionnaire Athletes-Sport 36

Content validity was rated inadequate due to the poor description of concept elicitation and lack of clear description of any cognitive interview process. There was high level of evidence and positive rating for structural validity. Internal consistency was downgraded to moderate level of evidence with an indeterminate rating due to no Cronbach alpha reported and only inter item correlations (r=0.21-0.60). Cross cultural validity was rated positive, as no important differences were found between group factors (gender, level and sport type), but downgraded to a moderate evidence, due to lack of clarity if samples were similar for relevant characteristics except the group variable. Evidence for responsiveness was graded moderate due to the suboptimal sample size (<100) according to the COSMIN guidelines.

Acute Recovery Stress Scale

Whilst the content of the ARSS was assessed based on expert survey, the content validity was rated inadequate, as the cognitive interview process were not clearly described and in line with the COSMIN standards. The measurement properties of the ARSS have been evaluated in eight studies (Hitzschke et al., 2017; Julian et al., 2017; Kölling et al., 2015; Kolling et al., 2019; Nassi, Ferrauti, Meyer, Pfeiffer, & Kellmann, 2017; Puta et al., 2018; Schimpchen et al., 2017; Zinner et al., 2017). There was a high level of evidence and overall positive rating for structural validity, internal consistency and cross-cultural validity. Construct validity was assessed in two studies by correlating the ARSS with subscales of sport specific stress and recovery from the RESTQ-Sport. Despite no formulation of specific hypotheses, moderate evidence was found for positive construct validity with one study reporting very low to moderate correlations between overall stress (ARSS) and subscales of sport specific stress (RESTQ-Sport) (0.33-0.46) and overall recovery (ARSS) and sport specific recovery (RESTQ-Sport) (0.08-0.44). Quality of evidence was therefore downgraded and rated indeterminate. Responsiveness was reported in six studies (Hitzschke et al., 2017; Julian et al., 2017; Kölling et al., 2015; Puta et al., 2018; Schimpchen et al., 2017; Zinner et al., 2017) and whilst most studies lacked specific hypotheses, overall the rating was positive with a high level of evidence.

Short Recovery Stress Scale

Content validity was rated inadequate as the cognitive interview process are not clearly described and in line with the COSMIN standards. The measurement properties of the SRSS have been examined in eleven studies (Engel, Faude, Kolling, Kellmann, & Donath, 2019; Hitzschke et al., 2015; Hitzschke et al., 2017; Kolling et al., 2019; Kolling et al., 2016; Nassi et al., 2017; Pelka, Ferrauti, Meyer, Pfeiffer, & Kellmann, 2017; Pelka et al., 2017; Raeder et al., 2016; Wiewelhove et al., 2016; Wiewelhove et al., 2018). Overall there was a high level of evidence and positive rating for structural validity, internal consistency and cross-cultural validity. Additionally, although not all studies formulated a hypothesis, there was a high level of evidence and positive rating for both construct validity and responsiveness.

Profile of Mood States-65

The content validity of the POMS-65 was rated inadequate since the original population involved in the development process were not the population of interest of this systematic review. Measurement properties have been evaluated in four studies (Goode & Roth, 1993; Mastro, French, & Hall, 1987; Terry & Lane, 2000; Terry & Youngs, 1996). Evidence for reliability was graded very low with an overall insufficient rating due to poor methodology with no clear outline of participant stability in the interim period;

suboptimal sample size (n=15) and no ICC reported. Although not all studies reported a hypothesis, there was moderate evidence for construct validity with an indeterminate rating due to methodological quality and inconsistency of results. Responsiveness was downgraded to moderate evidence with an indeterminate rating due to questionable methodology; no clear description provided of the intervention.

Profile of Mood States-63

Content validity was rated inadequate as it is not clear whether participants were used in the development process. Structural validity was downgraded to moderate evidence as there was suboptimal sample size. Internal consistency was rated high level of evidence with Cronbach alpha (α= 0.77-0.92). No other measurement properties have been reported.

Profile of Mood States-48

As it is not clear whether participants were used in the development process, content validity was therefore rated inadequate. Evidence for structural validity was very low and rated insufficient due to the suboptimal sample size. No other measurement properties have been reported.

Profile of Mood States-44

Content validity was rated inadequate as it is not clear whether participants were used in the development process. Confirmatory factor analysis for structural validity provided high evidence and rated positive. Cross cultural validity was graded moderate due to lack of clarity if participants were similar for relevant characteristics except for the group variable.

Profile of Mood States-40

Whilst the POMS-40 was administered in a sport setting there is no evidence of concept elicitation or any cognitive interview process and thus content validity was rated inadequate. The POMS-40 has been examined in two studies (Grove & Prapavessis, 1992; Soos, Lane, Leibinger, Istvan, & Hamar, 2005). Structural validity was graded high for quality of evidence; however, the overall rating was insufficient due to the poor factor analysis results (comparative fit index (CFI) = 0.41-0.92). Internal consistency was low quality of evidence due to the suboptimal sample size ($n = 45$); however, the overall rating was positive with Cronbach alpha for scales (α= 0.66-0.954). Responsiveness was graded very low evidence as there was only one study with suboptimal sample size ($n=45$), thus the overall rating was indeterminate.

Profile of Mood States-36

Content validity was rated inadequate due to lack of evidence of concept elicitation or cognitive interview process. Evidence for structural validity was poor as only exploratory factor analysis was performed. Internal consistency was high quality of evidence and positive for overall rating. Construct validity was determined by correlating POMS scales with Training Distress Scales (Raglin & Morgan, 1989). There was limited evidence of construct validity as no specific hypotheses were formulated and the psychometrics of the comparator instruments were not adequately described.

Profile of Mood States-30

Measurement properties of the POMS-30 have been evaluated in two studies (Andrade & Rodriguez, 2018; Bourgeois et al., 2010). Content validity was rated inadequate due to the lack of detail regarding concept elicitation or any cognitive interview process. There was moderate evidence with indeterminate overall rating for structural validity due to inconsistency of factor analysis results. Internal consistency was graded high for evidence with a positive overall rating.

Brunel Mood Scale (Profile of Mood States-24)

Content validity was rated inadequate as concept elicitation and the cognitive interview process involved school children (14-16 years) and schoolteachers. Measurement properties have been evaluated in nine studies (Almeida, Silva, Vendramini, Campos, & Brandão, 2017; Cañadas, Monleón, Sanchis, Fargueta, & Blasco, 2017; Fazackerley, Lane, & Mahoney, 2003; Lan, Lane, Roy, & Hanin, 2012; Miranda Rohlfs et al., 2008; Miranda Rohlfs et al., 2008; Terry, Lane, & Fogarty, 2003; Terry et al., 1999; Zhang, Si, Chung, Du, & Terry, 2014). There was a high level of evidence for structural validity, however the overall rating was insufficient due to the < 75% of the confirmatory factor analyses reporting the same factor structure. Internal consistency was downgraded to moderate evidence due to inconsistency of results. Construct validity was graded high level of evidence with a positive overall rating, despite most studies with no hypothesis formulated. Evidence for responsiveness was very low due to poor methodology and suboptimal sample size (n=34).

Daily Analysis of Life Demands for Athletes

Content validity was rated inadequate since it was unclear whether athletes and experts in the field were involved in the development. Additionally, it is not clear if the process of concept elicitation and cognitive interviews were conducted. Structural validity and internal consistency for the DALDA seem not to have been assessed. Responsiveness has been examined in three studies (Coutts et al., 2007; Gomes et al., 2013; Moreira et al., 2011) and rated overall positive with low grade evidence due to suboptimal of sample size (n=41).

Daily Analysis of Life Demands for Athletes-modified

Content validity was inadequate due to no concept elicitation or cognitive interview process. Only responsiveness was examined and rated insufficient with low evidence due to suboptimal sample size (n=16).

Athlete Burnout Questionnaire

Content validity was rated inadequate due to lack of clarity regarding concept elicitation based on the COSMIN standards. Four measurement properties were examined in seven studies (Cresswell & Eklund, 2006; de Francisco, Arce, Andrade, Arce, & Raedeke, 2009; Gerber et al., 2018; Pedrosa & García-Cueto, 2012; Raedeke, Arce, De Francisco, Seoane, & Ferraces, 2013; Raedeke & Smith, 2001; Šišková, Hudáková, & Sollár, 2017). Structural validity and internal consistency were graded high quality of evidence with an overall positive rating. Reliability was also graded high quality of evidence, however with indeterminate overall rating due to <75% of results with ICC ≥ 0.70. Construct validity was excellent with a high level of evidence and overall positive rating, and half of the studies formulated a hypothesis.

Athlete Burnout Questionnaire-18

Content validity was inadequate due to no clear concept elicitation process or cognitive interview. Quality of evidence was high for structural validity but rated insufficient due to poor confirmatory factor index result (<0.95). There was high level of evidence for internal consistency. Reliability was downgraded to moderate evidence as there were methodological concerns regarding conditions and stability for the re-test sample.

Athlete Burnout Questionnaire-12

Content validity was rated inadequate since there was no cognitive interview process. Structural validity and internal consistency were rated positive with a high quality of evidence. There was a high level of evidence for construct validity, despite no hypothesis formulation.

Perceived Stress Scale

Content validity was rated inadequate since the original population involved in the development process were not the population of interest of this systematic review. One study has extensively examined the measurement properties of the PSS (Chiu et al., 2016). Structural validity was rated insufficient due to poor confirmatory fit index (<0.95); however, the quality of evidence was high. There was a high level of evidence with positive rating for both internal consistency and measurement invariance. Reliability was rated insufficient due to correlations <0.80 and level of evidence was graded low due to suboptimal sample size (n=39). Despite no formulation of hypothesis, construct validity was rated positive with a high level of evidence.

Total Quality Recovery Scale

The content validity was rated inadequate as this appears to be more based on a theoretical framework with no concept elicitation or cognitive interview process involved. Three studies were available all examining responsiveness. There was moderate level of evidence due to suboptimal sample (n=50) and only one study formulated a hypothesis.

Overtraining questionnaire of the Societe Francaise de Medecine du Sport

Content validity was graded inadequate due to the lack of clear description of any cognitive interview process and poor description of concept elicitation. Reliability has been examined in one study with very limited evidence, as the test conditions and stability of the re-test sample not clearly outlined, and no ICC reported.

Multi-Component Training Distress Scale

Content validity was graded inadequate due to the lack of clear description of any cognitive interview process and poor description of concept elicitation. Evidence for structural validity was moderate with insufficient rating as there was only one study of adequate quality. Both internal consistency and construct validity had a high level of evidence.

Wellness and single items

There was one study that examined the construct validity in single item measures (Impellizzeri & Maffiuletti, 2007). The convergent validity of the seven-point Likert scale of muscle soreness was assessed with the Visual Analogue Scale (VAS). There was an overall positive rating with only low grade of evidence due to the suboptimal sample size (n= 26). Reliability, measurement error and responsiveness have been examined for 'wellness' items in two studies (Crowcroft et al., 2017; Fitzpatrick et al., 2019). Overall, there was a low level of evidence for all measurement properties due to suboptimal sample size and inconsistent overall rating for reliability and measurement error due to indeterminate results.

Other instruments

There were no studies examining measurement properties found in the literature for the POMS-33, fatigue questionnaires, Perceived recovery stress scale or other single item measures.

Appendix F: Quality of studies on measurement properties

AROM	Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Construct validity		Responsiveness	
						Convergent validity	Known groups validity	Comparison Between subgroups	Comparison Before and after intervention
RESTQ (Davis et al., 2007)	V	V							
RESTQ (Nederhof et al., 2008)		V		D		A			
RESTQ (González-Boto et al., 2008)	I	V							
RESTQ (Nicolas et al., 2019)	V	A	D						V
ARSS (Kolling et al., 2019)	V	V	V			V			
ARSS (Kölling et al., 2015)									D
ARSS (Nassi et al., 2017)	V	V				A			
ARSS (Schimpchen et al., 2017)									V
ARSS (Hitzschke et al., 2017)									V
ARSS (Julian et al., 2017)									V
ARSS (Zinner et al., 2017)								V	V
ARSS (Puta et al., 2018)									V
SRSS (Kolling et al., 2019)	V	V	V			V			
SRSS (Hitzschke et al., 2015)						D			D
SRSS (Nassi et al., 2017)	V	V				A			
SRSS (Pelka et al., 2017)									V
SRSS (Raeder et al., 2016)									V
SRSS (Wiewelhove et al., 2016)								V	V
SRSS (Hitzschke et al., 2017)									V
SRSS(Pelka et al., 2017)								V	V
SRSS (Kolling et al., 2016)									V
SRSS (Wiewelhove et al., 2018)								V	V
SRSS (Engel et al., 2019)									V
POMS (Mastro et al., 1987)				D					
POMS (Goode & Roth, 1993)						A			D

POMS (Terry & Youngs, 1996)						A		
POMS (Terry & Lane, 2000)						V		A
POMS (Andrade et al., 2010)	A	V						
POMS (Fernández et al., 2002)	I	V						
POMS (Andrade et al., 2013)	V		A					
POMS(Grove & Prapavessis, 1992)		V					A	
POMS (Soos et al., 2005)	V							
POMS (Viana et al., 2001)	A	V			A	A		
POMS (Andrade & Rodriguez, 2018)	V							
POMS (Bourgeois et al., 2010)	V	V						
BRUMS (Miranda Rohlfs et al., 2008)								D
BRUMS (Cañadas et al., 2017)	V	V				A		
BRUMS (Zhang et al., 2014)	V	V			V			
BRUMS (Fazackerley et al., 2003)	V	V						
BRUMS (Almeida et al., 2017)	A	I						
BRUMS (Lan et al., 2012)	V	V						
BRUMS (Miranda Rohlfs et al., 2008)	V	V						
BRUMS (Terry et al., 2003)	V	D			V	V		
BRUMS (Terry et al., 1999)	V	V			V	A		
OFSM (Flore, Sarrazin, & Favre-Juvin, 2003)				I				
ABQ (Raedeke & Smith, 2001)	V	V		A				
ABQ (de Francisco et al., 2009)	V	V		D				
ABQ (Šišková et al., 2017)	V	V						
ABQ (Pedrosa & García-Cueto, 2012)	V	V			V			
ABQ (Raedeke et al., 2013)	V				V			
ABQ (Cresswell & Eklund, 2006)		V			V			
ABQ (Gerber et al., 2018)	V	V		A	V			
ABQ (Arce et al., 2012)	V	V		D				
ABQ (Isoard-Gautheur et al., 2010)	V	V			V			
DALDA (Coutts et al., 2007)							V	V

DALDA (Gomes et al., 2013)								V
DALDA (Moreira et al., 2011)								V
DALDA (Nicholls et al., 2009)								V
TQR (Vitale, Banfi, Galbiati, Ferini-Strambi, & La Torre, 2019)								V
TQR (Freitas et al., 2014)								V
TQR (Osiecki et al., 2015)								D
Muscle soreness (Impellizzeri & Maffiuletti, 2007)						V		
PSS (Chiu et al., 2016)	V	V	V	D		V		
MTDS (Main & Grove, 2009)	A	V						
Wellness (Crowcroft et al., 2017)				D	D		V	
Wellness (Fitzpatrick et al., 2019)				V	A			

V= very good; A=adequate; D=doubtful; I=inadequate. RESTQ: Recovery-Stress Questionnaire Athletes-Sport; ARSS: Acute Recovery and Stress Scale; SRSS: Short Recovery and Stress Scale; POMS: Profile of Mood States; BRUMS: Brunel Mood Scale; DALDA: Daily Analysis of Life Demands for Athletes; ABQ: Athlete Burnout Questionnaire; OFSM: Overtraining questionnaire of the French Society for Sports Medicine; MTDS: Multi-Component Training Distress Scale; PSS: Perceived Stress Scale; TQR: Total Quality Recovery Scale. RTT-Q: Readiness to train questionnaire.

Appendix G: Measurement properties according to COSMIN

Content validity

Content validity is considered the most important measurement property of a PROM. It is the degree in which the content of an instrument is an adequate reflection of the construct measured (Mokkink et al., 2010). Content validity refers to the relevance, comprehensiveness and comprehensibility of the PROM, for the construct, population and context of interest (Terwee et al., 2018). For example, all items in an AROM should be appropriate for the construct of interest (within the specified population and context of use), the AROM should be comprehensive regarding participant concerns (American Educational Research Association, 2014) and understood by participants as proposed. Additionally, the importance of content validity is highlighted by several medical and health agencies (European Medicines Agency, 2005; U.S.Department of Health and Human ServicesFood and Drug Administration (FDA), 2009).

Absence of content validity can influence all other measurement properties. For example, irrelevant items may reduce structural validity, internal consistency and interpretability of the AROM. Furthermore, an acceptable Cronbach's alpha does not ensure the construct of interest is being measured (Sijtsma, 2009) and responsiveness does not suggest that all items are relevant. Furthermore, missing concepts may decrease responsiveness and validity. Finally, irrelevant or missing items may potentially lead to participant frustration and response bias (de Leeuw, Hox, & Dillman, 2008; Streiner & Norman, 2008).

In the context of our review, content validity should have been assessed by asking athletes and the appropriate professionals (e.g. coaches, trainers and sports scientists) about the relevance, comprehensiveness and comprehensibility of the items, response options and instructions of the AROM (Streiner & Norman, 2008). In assessing the methodological quality of the content validity, our results show that all AROMs were inadequate with very low level of evidence. The majority of studies identified in part two did not report any component of the content validity process. Whilst, several studies reported an element of the content validity process (e.g. relevance or comprehension) it was either in the wrong population of interest (e.g. school children) or the methods were poorly described and not in line with the COSMIN methodology standards. Therefore, on this basis, the content validity for all AROMs rated inadequate. If there is a high level of evidence that the content validity of the AROM is inadequate, COSMIN guidelines suggest no further evaluation of other measurement properties. However, despite the content validity for all AROMs being rated inadequate we choose to continue evaluating the rest of the measurement properties.

Structural validity

COSMIN defines structural validity as the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured (Mokkink et al., 2010). To determine the structure of an instrument, a factor analysis is the preferred statistic when using classical test theory (CTT) and the criteria for a sufficient rating either >0.95 for comparative fit index (CFI), or <0.06 Root Mean Square Error of Approximation (RMSEA) or < 0.08[2] Standardized Root Mean Residuals (SRMR). In our results, we were liberal with our rating, if CFI or RMSEA fell just short of the COSMIN criteria we still rated positive.

Internal consistency

Internal consistency is the degree of the interrelatedness among items (Mokkink et al., 2010). The criteria for a sufficient rating is at least low evidence (as defined by the GRADE approach) for sufficient structural validity and Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale. We extended the criteria for a sufficient rating even if the Cronbach’s alpha fell slightly below the COSMIN standard and rated methodological quality adequate despite no Cronbach’s alpha reported.

Cross-cultural validity/Measurement invariance

Cross-cultural validity refers to the degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument (Mokkink et al., 2010). The interpretation of ‘culturally different population’ was broad and did not only include different ethnicity or language groups, but also other groups such as different gender or age groups, or different patient populations. The criteria for a positive rating was no important differences found between group factors in multiple group factor analysis OR no important differential item functioning for group factors (McFadden's $R^2 < 0.02$).

Reliability

COSMIN define reliability as the degree to which the measurement is free from measurement error. This can be extended to the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same PROM (internal consistency); over time (test-retest); by different persons on the same occasion (interrater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater) (Mokkink et al., 2010). The criteria for positive rating is ICC or weighted kappa ≥ 0.70. If only Pearson correlation or spearman Rho reported, we rated insufficient.

Construct validity

Construct validity is defined as the degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured. To prevent report bias, the COSMIN checklist recommends authors formulate specific hypothesis before data collection (Mokkink et al., 2010) and if no hypothesis reported this will be an indeterminate rating. However, the majority of studies we assessed did not report a specific hypothesis or specify the expected magnitude of correlations. Therefore, we were liberal in our criteria and rated positive despite no hypothesis reported, as long as correlations were consistent.

Responsiveness

Responsiveness is defined as the ability of a PROM to detect change over time in the construct to be measured (Mokkink et al., 2010). The COSMIN criteria for a positive rating is that the result is in accordance with the hypothesis OR area under the curve (AUC) ≥ 0.70. Similar to construct validity, we extended the criteria for a positive rating to include studies that didn't report a hypothesis, as long as correlations were consistent.

Appendix H: Discussion Athlete Reported Outcome Measures

Recovery-Stress Questionnaire Athletes (RESTQ-Sport)

The RESTQ-Sport (Kallus & Kellmann) contains 77 items (19 scales with four items each plus one warm-up item) and is a sport specific extension of the basic RESTQ by Kallus (1995). It includes the scales: general stress and recovery and sport specific stress and recovery. The RESTQ-Sport was developed to measure the frequency of current stress together with the frequency of recovery-associated activities. Whilst, the RESTQ-Sport has a strong theoretical framework, concept elicitation and the cognitive interview process was not clearly described and in line with the COSMIN standards. Overall, there was good evidence for most measurement properties except reliability, for which there was very low evidence. This was due to only one study of poor quality showing methodological issues. Additionally, the RESTQ-Sport has several abbreviated versions, one of which met the criteria for this review, the RESTQ-Sport 36, an abbreviated French version (Nicolas et al., 2019). There was poor evidence for cross cultural validity due to methodological issues in which no group differences were discussed. Whilst the RESTQ-Sport is one of the most frequently used AROMs, further research is required concerning clarity of content validity and reliability.

Acute Recovery and Stress Scale (ARSS)

The ARSS is a psychometric scale used to assess acute recovery and stress in athletes. Originally developed by Kellmann et al. (2016) in German and later translated to English. Containing 32 adjectives the scale consists of: physical performance capability, mental performance capability, emotional balance, and overall recovery in the dimension of recovery; and muscular stress, lack of activation, negative emotional state and overall stress in the dimension of stress. In order to describe the recovery and stress of an athlete the ARSS is based on a multidimensional approach. Despite a strong theoretical framework and both qualitative and quantitative expert interviews conducted, content validity was inadequate according to the COSMIN guidelines. It is unclear the exact process of concept elicitation and the cognitive interview study. For example, whether there was data coding, group meetings with recording or transcription verbatim, if athletes were asked for comprehension, or comprehensibility. Therefore, elaboration of the content validity process or further research is suggested to be line with the COSMIN methodology standards for PROMs. Generally, there was high level of evidence across all measurement properties, except construct validity, with only moderate evidence. Even though there was no predefined hypothesis or specification of the expected magnitude of correlations, we still rated the ARSS positive, but downgraded the level of evidence due to low correlations between measures.

Short Recovery and Stress Scale (SRSS)

The SRSS was derived from the eight scales of the ARSS and grouped into the Short Recovery Scale and the Short Stress Scale consisting of four items each (Hitzschke et al., 2015). As the SRSS was derived from the ARSS, content validity was rated inadequate due to the same reasons specified above. The SRSS performed the best of all AROMs with all measurement properties graded high quality of evidence with positive rating. However, we were unable to find studies reporting measurement error. Therefore, further research extending the content validity process together with research examining measurement error is warranted.

Profile of Mood States (POMS)

The original POMS was developed in 1971 by McNair et al. (1971) and is a 65 five-point adjective rating scale that assesses six dimensions of the mood construct: tension-anxiety; depression-dejection; anger-hostility; vigour-activity; fatigue-inertia; and confusion-bewilderment. The original POMS was recommended as a measure of mood states and changes in mood states in psychiatric outpatients. The basis of the original data came from three reported sources: experimental clinical research, clinic assessment from university medical psychiatric clinics and selected samples from normal college students. Consequently, content validity was inadequate, as the foundation paper was never developed for the population of interest in this systematic review (athletes). Despite this, the POMS-65 is one of the most widely used AROMs in athletic monitoring. Surprisingly, we did not find evidence of studies examining structural validity and internal consistency within the sporting literature. Additionally, since reproducible and accurate measurements are pre-requisites for an acceptable instrument, adequate reliability is essential. The POMS-65 demonstrated very low evidence for reliability. The inadequate sample size resulted in a serious risk of bias. We believe this finding, together with poor content validity indicates a striking need for re-evaluation of these measurement properties in future research. Over the years there have been many modifications and revisions of the original POMS-65, together with name changes. Altogether seven versions were included in this review and will be discussed below.

There have been several translations and abbreviations of the POMS-65, including the Spanish POMS-63 (Andrade et al., 2010), POMS-48 (Fernández et al., 2002), POMS-44 (Andrade et al., 2013) and the Portuguese POMS-36 (Viana et al., 2001). Collectively, the content validity for all versions was poor. In the Spanish versions, only structural validity, internal consistency and cross-cultural validity have been evaluated. The POMS-

48 had very low evidence for structural validity due to the serious risk of bias with suboptimal sample size. Additionally, there were methodological issues with the POMS-36, with only exploratory factor analysis performed and, on this basis, results rated insufficient. Therefore, the use of the POMS-63, POMS-48, POMS-44 and POMS-36 in athletic populations should be recommended cautiously due to poor content validity, insufficient evidence, and several measurement properties that are yet to be examined.

The POMS-40 is a modified form of (Shacham, 1983) short POMS, which was designed for use with hospital patients. Content validity was poor as the original validation study was in a clinical setting. A concern with the structural validity are the insufficient results, which fall below the COSMIN standards recommended criteria for CFI (comparative fit index). Additionally, there was a serious risk of bias for internal consistency due to the suboptimal sample size. Similarly, there was a serious risk of bias for responsiveness also due to an inadequate sample size. Collectively, the evidence suggests that studies of higher methodological quality are needed for proper interpretation of these measurement properties.

Another abbreviation of the original POMS-65 is the POMS-30, also named the EDITS short form of the POMS (EPOMS) and was developed by the original publisher of the full scale version of the POMS (Service/EDITS, 1999). Content validity was poor due to the original POMS designed in a clinical setting. A major concern with the POMS-30 is many measurement properties haven't yet been investigated, with only evidence for structural validity and internal consistency. Furthermore, one study had insufficient factor analysis results according to the COSMIN standards which resulted in an indeterminate rating overall. Given these issues and the lack of research regarding other measurement properties, together with poor content validity, caution should be taken when applying this AROM in the athletic setting.

The POMS-24, originally titled the Profile of Mood States-Adolescents (POMS-A) and later renamed to the Brunel Mood Scale (BRUMS), is a shortened version of the POMS-65. Developed by Terry et al. (1999), the POMS-24 was designed to assess mood in younger populations and consists of 24 items. The original population of interest was school children and based on this we rated content validity poor. Many studies have examined the BRUMS reporting examination of structural validity, internal consistency, construct validity and responsiveness. Responsiveness was rated very low evidence with a serious risk of bias due to methodological issues and suboptimal sample size. Despite good evidence for other measurement properties, further research is warranted regarding content validity, responsiveness and evaluation of other measurement properties (e.g., measurement error) when considering the BRUMS for use in athletic populations.

Daily Analysis of Life Demands for Athletes (DALDA)

The DALDA was developed by Rushall (1990) as a tool for measuring stress tolerance in elite athletes. It is a two-part inventory; Part A describes general stress sources (nine items) and Part B determines stress reaction symptoms (25 items). The questionnaire requires subjects to rate each variable as being ‘worse than normal’, ‘normal’, or ‘better than normal’. Due to the lack of clarity regarding the involvement of athletes and experts in the process of concept elicitation and whether cognitive interviews were conducted, the content validity was rated very low evidence. Despite the DALDA being one of the most commonly used AROMs, there was a distinct lack of research of the measurement properties, with only responsiveness examined. Our results indicate low evidence with methodological issues regarding inadequate sample size, based on the COSMIN guidelines. Apart from the previously described DALDA version, one study adapted the response scale to include a horizontal visual analogue scale (Nicholls et al., 2009). Based on the COSMIN guidelines, a modification of the response option warrants a separate version of the DALDA. Similar to the above results, only responsiveness was examined and rated poor due to inadequate sample size. Therefore, there is an urgent need of further high-quality methodological studies to properly assess and strengthen the measurement properties of the DALDA.

Athlete Burnout Questionnaire (ABQ)

The ABQ developed by (Raedeke & Smith, 2001) measures three burnout-related components: physical/emotional exhaustion, sport devaluation and reduced sense of personal accomplishment, with each component consisting of five items, 15 items in total. Both athletes and experts were involved in pilot testing regarding readability and comprehension, however the exact process of concept elicitation (relevance and comprehensiveness) and cognitive interview process was not clearly described and in line with the COSMIN standards. Therefore, content validity was rated inadequate. The ABQ was rated high quality of evidence for all measurement properties examined. Additionally, the ABQ had mostly positive ratings, except for reliability which received an overall indeterminate rating. This was due to the risk of bias in one study which reported Pearson correlations but no ICC calculations. Furthermore, the ABQ has undergone two modifications with an 18 item Spanish version (Arce et al., 2012) and a 12 item French version (Isoard-Gautheur et al., 2010), both with moderate to high quality of evidence. Collectively, the ABQ has strong quality of evidence, however further research is warranted in strengthening the content validity.

Perceived Stress Scale (PSS)

The PSS is a 14-item instrument designed to measure the degree to which situations in one's life are perceived as stressful (Cohen et al., 1983). The PSS was initially recommended for the investigation of non-specific appraised stress in the aetiology of disease and behavioural disorders, and as an outcome measure of experienced levels of stress. The original validation data was collected from college students and participants enrolled in a smoking cessation program. Based on this, the content validity was rated inadequate as the population of interest was not athletic. One study has examined the measurement properties of the PSS (Chiu et al., 2016). Whilst factor analysis was used to determine structural validity, the CFI results fell below the COSMIN standards and therefore was rated insufficient. Additionally, evidence for reliability was low due to serious risk of bias, with doubtful methodological quality and imprecision of sample size. As such, further research regarding content validity is recommended for use of the PSS in athletic populations, together with better quality studies on reliability and structural validity.

Total Quality Recovery Scale (TQR)

The TQR scale was developed by (Kenttä & Hassmén, 1998) for the measurement of psychophysiological recovery in athletes. We included this scale in the current review even if it may be considered a psychophysics scale. However, it was not developed following psychophysics methods so it would also be rated inadequate from this perspective. We therefore examined TQR as an AROM. The TQR is divided into two subscales; TQR perceived (subjective assessment of an athlete's recovery rated from 6 to 20) and the TQR action (objective scale grading actions of athlete's recovery process based on a point system). The majority of the psychometric properties were not examined with the exception of responsiveness, that showed moderate evidence due to suboptimal sample size. This is a concern as the TQR is one of the most frequently used single item measures in athlete monitoring.

Overtraining questionnaire of the Societe Francaise de Medecine du Sport (OFSM)

The OFSM was developed by Legros (1993) for the identification of clinical behavioural symptoms that accompany the onset of overtraining syndrome. The OFSM was based on existing behavioural scales; quality of life research in endocrinology and the Nottingham Health Profile (a general PROM for measurement of subjective health status). The questionnaire contains 54 items with a binary response option. Content validity was poor due lack of information regarding concept elicitation and any form of cognitive interview process. Furthermore, we were only able to identify one study examining reliability (Flore et al., 2003), in which there was very low evidence due to a very serious risk of bias based on methodological issues and imprecision regarding sample size. This is problematic because accurate and reproducible measurements are prerequisite for a measurement instrument, and OFSM is one of the more commonly used AROMs.

Multi-Component Training Distress Scale (MTDS)

The MTDS combines measures of mood disturbance, with measures of perceived stress and symptom intensity (Main & Grove, 2009). It comprises of six training distress factors: depressed mood, perceived vigour, physical symptoms, sleep disturbance, perceived stress, and general fatigue. Content validity was graded inadequate due to the lack of clear description of any cognitive interview process and poor description of concept elicitation. Structural validity was downgraded to moderate with insufficient rating as there was only one study conducted using exploratory factor analysis.

Appendix I: Correlation matrix (Spearman's rho) SRSS and single items

	SRSS recovery				**SRSS stress**				**Single items**			
	Phys perf	Ment perf	Emot bal	Overall recov	Musc stress	Lack activat	Neg emot	Overall stress	Recovery REL	Recovery ABS	Fatigue REL	Fatigue ABS
Phys perf	1	0.55**	0.51**	0.69**	-0.23**	-0.38**	-0.37**	-0.39**	0.45**	0.41**	0.32**	-0.46**
95% CI		0.42 to 0.66	0.40 to 0.63	0.60 to 0.77	-0.36 to -0.09	-0.49 to -0.27	-0.49 to -0.25	-0.50 to -0.29	0.33 to 0.57	0.28 to 0.540	0.19 to 0.46	-0.57 to -0.34
Mental perf	0.55**	1	0.67**	0.46**	-0.13*	-0.43**	-0.53**	-0.40**	0.35**	0.40**	0.38**	-0.44**
95% CI	0.42 to 0.66		0.58 to 0.76	0.33 to 0.58	-0.26 to -0.01	-0.54 to -0.31	-0.64 to -0.44	-0.51 to -0.29	0.23 to 0.47	0.28 to 0.53	0.25 to 0.52	-0.55 to -0.33
Emot bal	0.51**	0.67**	1	0.54**	-0.21**	-0.44**	-0.69**	-0.51**	0.35**	0.39**	0.32**	-0.50**
95% CI	0.40 to 0.63	0.58 to 0.76		0.43 to 0.65	-0.33 to -0.08	-0.56 to -0.33	-0.77 to -0.61	-0.61 to -0.42	0.22 to 0.47	0.28 to 0.52	0.19 to 0.44	-0.60 to -0.40
Overall recov	0.69**	0.46**	0.54**	1	-0.44**	-0.34**	-0.45**	-0.46**	0.45**	0.43**	0.31**	-0.52**
95% CI	0.60 to 0.77	0.33 to 0.58	0.43 to 0.65	.	-0.55 to -0.32	-0.47 to -0.23	-0.57 to -0.33	-0.57 to -0.35	0.34 to 0.56	0.31 to 0.57	0.19 to 0.45	-0.63 to -0.41
Musc stress	-0.22**	-0.13*	-0.21**	-0.44**	1	0.28**	0.38**	0.54**	-0.31**	-0.29**	-0.30**	0.44**
95% CI	-0.36 to -0.09	-0.26 to -0.01	-0.33 to -0.08	-0.55 to -0.32	.	0.14 to 0.42	0.26 to 0.50	0.43 to 0.65	-0.43 to -0.19	-0.42 to -0.16	-0.43 to -0.18	0.31 to 0.56
Lack activat	-0.37**	-0.42**	-0.44**	-0.34**	0.28**	1	0.60**	0.53**	-0.27**	-0.26**	-0.32**	0.37**
95% CI	-0.49 to -0.27	-0.54 to -0.31	-0.56 to -0.33	-0.47 to -0.23	0.14 to 0.42	.	0.50 to 0.70	0.43 to 0.64	-0.40 to -0.15	-0.40 to -0.15	-0.44 to -0.21	0.24 to 0.51
Neg emot	-0.37**	-0.53**	-0.69**	-0.45**	0.38**	0.60**	1	0.64**	-0.30**	-0.29**	-0.28**	0.39**
95% CI	-0.49 to -0.25	-0.64 to -0.44	-0.77 to -0.61	-0.57 to -0.33	0.26 to 0.50	0.50 to 0.70	.	0.55 to 0.73	-0.42 to -0.18	-0.42 to -0.17	-0.40 to -0.16	0.27 to 0.51
Overall stress	-0.39**	-0.40**	-0.51**	-0.46**	0.54**	0.53**	0.64**	1	-0.36**	-0.34**	-0.40**	0.51**
95% CI	-0.50 to -0.29	-0.51 to -0.29	-0.61 to -0.42	-0.57 to -0.35	0.43 to 0.65	0.43 to 0.64	0.55 to 0.73	.	-0.47 to -0.26	-0.46 to -0.23	-0.51 to -0.27	0.39 to 0.62
Recovery REL	0.45**	0.35**	0.35**	0.45**	-0.31**	-0.27**	-0.30**	-0.36**	1	0.66**	0.47**	-0.51**
95% CI	0.33 to 0.57	0.23 to 0.47	0.22 to 0.47	0.34 to 0.56	-0.43 to -0.19	-0.40 to -0.15	-0.42 to -0.18	-0.47 to -0.26	.	0.58 to 0.74	0.36 to 0.60	-0.61 to -0.41
Recovery ABS	0.41**	0.40**	0.39**	0.43**	-0.29**	-0.26**	-0.29**	-0.34**	0.66**	1	0.34**	-0.48**
95% CI	0.28 to 0.54	0.28 to 0.53	0.28 to 0.52	0.31 to 0.57	-0.42 to -0.16	-0.40 to -0.15	-0.42 to -0.17	-0.46 to -0.23	0.58 to 0.74	.	0.22 to 0.48	-0.59 to -0.37
Fatigue REL	0.32**	0.38**	0.32**	0.31**	-0.30**	-0.32**	-0.28**	-0.40**	0.47**	0.35**	1	-0.56**
95% CI	0.19 to 0.46	0.25 to 0.52	0.19 to 0.44	0.19 to 0.45	-0.43 to -0.18	-0.44 to -0.21	-0.40 to -0.16	-0.51 to -0.27	0.36 to 0.60	0.22 to 0.48	.	-0.66 to -0.46
Fatigue ABS	-0.46**	-0.44**	-0.50**	-0.52**	0.44**	0.37**	0.39**	0.51**	-0.51**	-0.49**	-0.56**	1
95% CI	-0.57 to -0.34	-0.55 to -0.33	-0.60 to -0.40	-0.63 to -0.41	0.31 to 0.56	0.24 to 0.51	0.27 to 0.51	0.39 to 0.62	-0.61 to -0.41	-0.59 to -0.37	-0.66 to -0.46	.

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed). SRSS: Short Stress and Recovery Scale; REL: relative/unipolar scale; ABS: absolute/bipolar scale; CI: confidence interval; Phys perf: physical performance; Mental perf: mental performance; Emot bal: emotional balance; Overall recov: overall recovery; Musc stress: muscular stress; Lack activat: lack activation; Neg emot: negative emotional.

Appendix J: Correlation matrix (Spearman's rho) BRUMS with single items

	BRUMS					Single items			
	Tired	Sleepy	Worn out	Exhausted	BRUMS score	Recovery REL	Recovery ABS	Fatigue REL	Fatigue ABS
Tired	1	0.51**	0.73**	0.64**	0.86**	-0.34**	-0.32**	-0.42**	0.64**
95% CI	.	0.41 to 0.61	0.65 to 0.81	0.54 to 0.73	0.82 to 0.90	-0.46 to -0.23	-0.43 to -0.20	-0.52 to -0.31	0.55 to 0.73
Sleepy	0.51**	1	0.52**	0.42**	0.72**	-0.07	-0.09	-0.20**	0.28**
95% CI	0.41 to 0.61	.	0.41 to 0.62	0.30 to 0.54	0.64 to 0.78	-0.21 to 0.06	-0.22 to 0.03	-0.34 to -0.08	0.15 to 0.41
Worn out	0.73**	0.52**	1	0.78**	0.90**	-0.33**	-0.36**	-0.41**	0.63**
95% CI	0.65 to 0.81	0.41 to 0.62	.	0.72 to 0.84	0.88 to 0.93	-0.45 to -0.22	-0.48 to -0.23	-0.52 to -0.31	0.54 to 0.72
Exhausted	0.64**	0.42**	0.78**	1	0.85**	-0.39**	-0.37**	-0.41**	0.64**
95% CI	0.54 to 0.73	0.30 to 0.54	0.72 to 0.84	.	0.81 to 0.89	-0.50 to -0.29	-0.49 to -0.25	-0.51 to -0.32	0.55 to 0.72
BRUMS score	0.86**	0.72**	0.90**	0.85**	1	-0.34**	-0.34**	-0.43**	0.66**
95% CI	0.82 to 0.90	0.64 to 0.78	0.88 to 0.93	0.81 to 0.89	.	-0.46 to -0.23	-0.47 to -0.22	-0.53 to -0.34	0.56 to 0.74
Recovery REL	-0.34**	-0.07	-0.33**	-0.39**	-0.34**	1	0.66**	0.47**	-0.51**
95% CI	-0.46 to -0.23	-0.21 to 0.06	-0.45 to -0.22	-0.50 to -0.29	-0.46 to -0.23	.	0.58 to 0.74	0.36 to 0.60	-0.61 to -0.41
Recovery ABS	-0.32**	-0.097	-0.36**	-0.37**	-0.34**	0.66**	1	0.34**	-0.48**
95% CI	-0.43 to -0.20	-0.22 to 0.03	-0.48 to -0.23	-0.49 to -0.25	-0.47 to -0.22	0.58 to 0.74	.	0.22 to 0.48	-0.59 to -0.37
Fatigue REL	-0.42**	-0.21**	-0.42**	-0.41**	-0.43**	0.47**	0.35**	1	-0.56**
95% CI	-0.52 to -0.31	-0.34 to -0.08	-0.52 to -0.31	-0.51 to -0.32	-0.53 to -0.34	0.36 to 0.60	0.22 to 0.48	.	-0.66 to -0.46
Fatigue ABS	0.65**	0.28**	0.63**	0.64**	0.66**	-0.51**	-0.49**	-0.56**	1
95% CI	0.55 to 0.73	0.15 to 0.41	0.54 to 0.72	0.55 to 0.72	0.56 to 0.74	-0.61 to -0.41	-0.59 to -0.37	-0.66 to -0.46	.

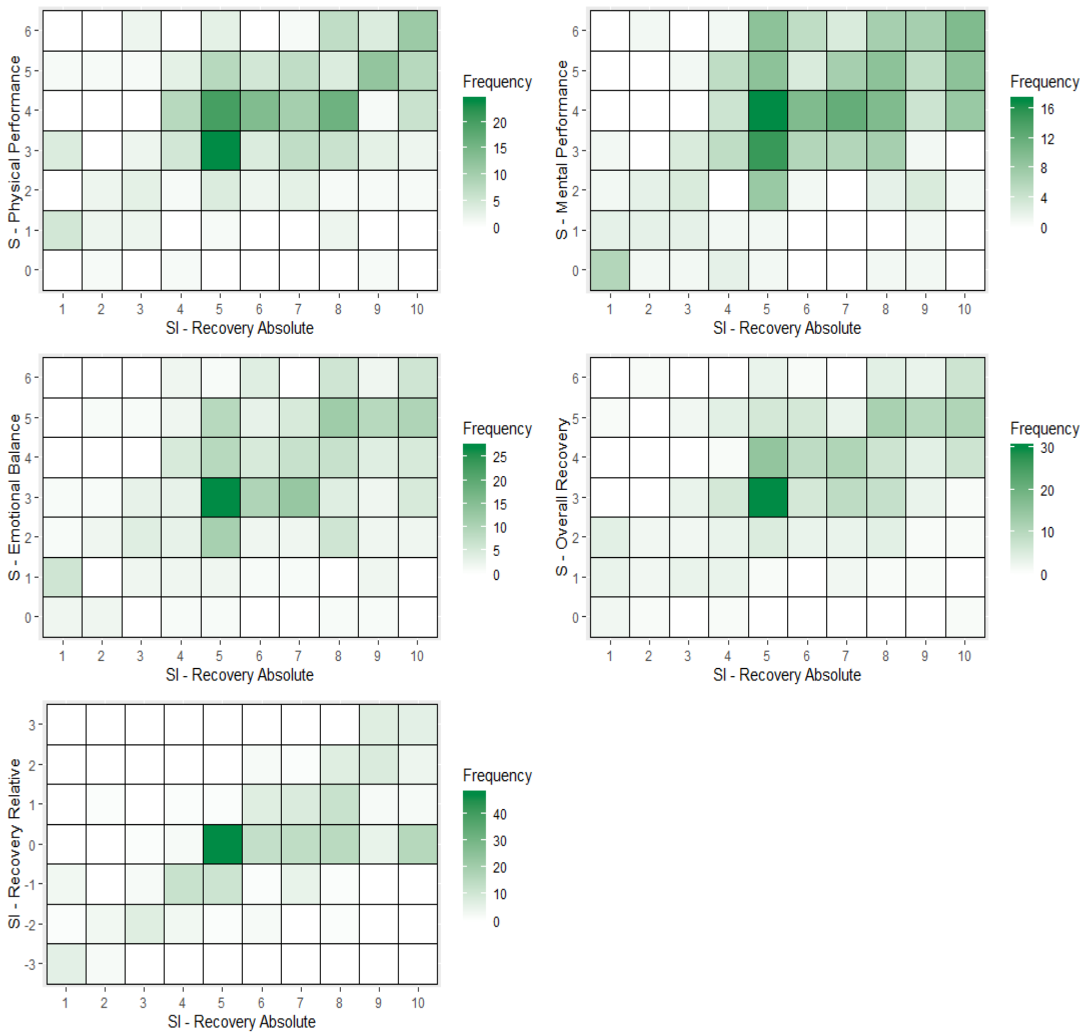
** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed). BRUMS: Brunel Mood Scale; REL: relative/unipolar scale; ABS: absolute/bipolar scale; CI: confidence interval.

Appendix K: Correlation matrix (Spearman's rho) SRSS and BRUMS

	SRSS recovery				SRSS stress				BRUMS				
	Phys perf	Ment perf	Emot bal	Overall recov	Musc stress	Lack activat	Neg emot	Overall stress	Tired	Sleepy	Worn out	Exhausted	BRUMS score
Phys perf	1	0.55**	0.51**	0.69**	-0.23**	-0.38**	-0.37**	-0.39**	-0.31**	-0.13*	-0.34**	-0.38**	-0.34**
95% CI		0.42 to 0.66	0.40 to 0.63	0.60 to 0.77	-0.36 to -0.09	-0.49 to -0.27	-0.49 to -0.25	-0.50 to -0.29	-0.43 to -0.19	-0.25 to 0.01	-0.45 to -0.22	--0.49 to -0.26	-0.46 to -0.22
Mental perf	0.55**	1	0.67**	0.46**	-0.13*	-0.43**	-0.53**	-0.40**	-0.24**	-0.22**	-0.33**	-0.39**	-0.35**
95% CI	0.42 to 0.66		0.58 to 0.76	0.33 to 0.58	-0.26 to -0.01	-0.54 to -0.31	-0.64 to -0.44	-0.51 to -0.29	-0.37 to -0.11	-0.34 to -0.10	-0.46 to -0.21	-0.51 to -0.27	-0.47 to -0.23
Emot bal	0.51**	0.67**	1	0.54**	-0.21**	-0.44**	-0.69**	-0.51**	-0.31**	-0.17**	-0.39**	-0.46**	-0.40**
95% CI	0.40 to 0.63	0.58 to 0.76		0.43 to 0.65	-0.33 to -0.08	-0.56 to - 0.33	-0.77 to -0.61	-0.61 to -0.42	-0.42 to -0.20	-0.30 to -0.04	-0.50 to -0.27	-0.56 to -0.34	-0.51 to -0.28
Overall recov	0.69**	0.46**	0.54**	1	-0.44**	-0.34**	-0.45**	-0.46**	-0.40**	-0.20**	-0.41**	-0.47**	-0.44**
95% CI	0.60 to 0.77	0.33 to 0.58	0.43 to 0.65	.	-0.55 to -0.32	-0.47 to -0.23	-0.57 to -0.33	-0.57 to -0.35	-0.52 to -0.28	-0.33 to -0.08	-0.52 to -0.29	-0.58 to -0.36	-0.54 to -0.32
Musc stress	-0.22**	-0.13*	-0.21**	-0.44**	1	0.28**	0.38**	0.54**	0.36**	0.16**	0.31**	0.34**	0.36**
95% CI	-0.36 to -0.09	-0.26 to -0.01	-0.33 to -0.08	-0.55 to -0.32	.	0.14 to 0.42	0.26 to 0.50	0.43 to 0.65	0.23 to 0.48	0.04 to 0.30	0.17 to 0.44	0.22 to 0.46	0.23 to 0.48
Lack activat	-0.37**	-0.42**	-0.44**	-0.34**	0.28**	1	0.60**	0.53**	0.27**	0.20**	0.30**	0.25**	0.31**
95% CI	-0.49 to -0.27	-0.54 to -0.31	-0.56 to -0.33	-0.47 to -0.23	0.14 to 0.42	.	0.50 to 0.70	0.43 to 0.64	0.15 to 0.41	0.08 to 0.33	0.18 to 0.43	0.14 to 0.39	0.19 to 0.44
Neg emot	-0.37**	-0.53**	-0.69**	-0.45**	0.38**	0.60**	1	0.64**	0.27**	0.24**	0.31**	0.39**	0.37**
95% CI	-0.49 to -0.25	-0.64 to -0.44	-0.77 to -0.61	-0.57 to -0.33	0.26 to 0.50	0.50 to 0.70	.	0.55 to 0.73	0.15 to 0.40	0.14 to 0.37	0.18 to 0.43	0.28 to 0.52	0.25 to 0.49
Overall stress	-0.39**	-0.40**	-0.51**	-0.46**	0.54**	0.53**	0.64**	1	0.37**	0.22**	0.44**	0.41**	0.44**
95% CI	-0.50 to -0.29	-0.51 to -0.29	-0.61 to -0.42	-0.57 to -0.35	0.43 to 0.65	0.43 to 0.64	0.55 to 0.73	.	0.24 to 0.49	0.11 to 0.34	0.33 to 0.56	0.30 to 0.54	0.33 to 0.55
Tired	-0.31**	-0.24**	-0.31**	-0.40**	0.36**	0.27**	0.27**	0.37**	1	0.51**	0.73**	0.64**	0.86**
95% CI	-0.43 to -0.19	-0.37 to -0.11	-0.42 to -0.20	-0.52 to -0.28	0.23 to 0.48	0.15 to 0.41	0.15 to 0.40	0.24 to 0.49	.	0.41 to 0.61	0.65 to 0.81	0.54 to 0.73	0.82 to 0.90
Sleepy	-0.13*	-0.22**	-0.17**	-0.20**	0.16**	0.20**	0.24**	0.22**	0.51**	1	0.52**	0.42**	0.72**
95% CI	-0.25 to 0.01	-0.34 to -0.10	-0.30 to -0.04	-0.33 to -0.08	0.04 to 0.30	0.08 to 0.33	0.14 to 0.37	0.11 to 0.34	0.41 to 0.61	.	0.41 to 0.62	0.30 to 0.54	0.64 to 0.78
Worn out	-0.33**	-0.33**	-0.39**	-0.41**	0.31**	0.30**	0.31**	0.44**	0.73**	0.52**	1	0.78**	0.90**
95% CI	-0.45 to -0.22	-0.46 to -0.21	-0.50 to -0.27	-0.52 to -0.29	0.17 to 0.44	0.18 to 0.43	0.18 to 0.43	0.33 to 0.56	0.65 to 0.81	0.41 to 0.62	.	0.72 to 0.84	0.88 to 0.93
Exhausted	-0.37**	-0.39**	-0.46**	-0.47**	0.34**	0.25**	0.39**	0.41**	0.64**	0.42**	0.78**	1	0.85**
95% CI	-0.49 to -0.26	-0.51 to -0.27	-0.56 to -0.34	-0.58 to -0.36	0.22 to 0.46	0.14 to 0.39	0.28 to 0.52	0.30 to 0.54	0.54 to 0.73	0.30 to 0.54	0.72 to 0.84	.	0.81 to 0.89
BRUMS score	-0.34**	-0.35**	-0.40**	-0.43**	0.36**	0.31**	0.37**	0.44**	0.86**	0.72**	0.90**	0.85**	1
95% CI	-0.46 to -0.22	-0.47 to -0.23	-0.51 to -0.28	-0.54 to -0.32	0.23 to 0.48	0.19 to 0.44	0.25 to 0.49	0.33 to 0.55	0.82 to 0.90	0.64 to 0.78	0.88 to 0.93	0.81 to 0.89	.

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed). SRSS: Short Stress and Recovery Scale; REL: relative/unipolar scale; ABS: absolute/bipolar scale; CI: confidence interval; Phys perf: physical performance; Mental perf: mental performance; Emot bal: emotional balance; Overall recov: overall recovery; Musc stress: muscular stress; Lack activat: lack activation; Neg emot: negative emotional; BRUMS: Brunel Mood Scale

Appendix L: Combined correlogram for the single item recovery absolute and relative (SI) with the Short Recovery Stress Scale items of recovery (S)



Appendix Ma: Concept elicitation saturation grid for fatigue

Concept	Total number in study (n=46)			
	Group 1 N=12	Group 2 N=12	Group 3 N=12	Group 4 N=10
Energy	x			
Tiredness	x			
Feeling normal	x			
Muscle soreness	x			
Rested	x			
Sleep duration	x			
Heavy leg	x			
Muscular fatigue		x		
Mental fatigue		x		
Strength		x		
Physical hindrances			x	
Harder to wake up			x	
Burnout			x	
Sleep quality			x	
Sluggish			x	
Rested			x	
Mood			x	
Stress			x	
Drained				x
Lazy				x

Appendix Mb: Concept elicitation saturation grid for recovery

Concept	Total number in study (n=37)			
	Group 1 N=9	Group 2 N=9	Group 3 N=9	Group 4 N=10
Fatigue	x			
Tiredness	x			
Muscle soreness	x			
Feeling good	x			
Rested	x			
Feeling normal	x			
Heavy leg	x			
Sleepy	x			
Sleep quality		x		
Energy		x		
Exhaustion		x		
Weakness			x	
Stiffness			x	
Physical hindrances			x	
Recover (wish)			x	
Stress				x
Anxious				x

Appendix N: University ethics approval dance study

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Human Research Ethics Committee
Ethics Secretariat C/O Research Office
15 Broadway, Ultimo NSW 2007

T: +61 2 9514 9681
Research.Ethics@uts.edu.au

PO Box 123
Broadway
NSW 2007 Australia www.uts.edu.au

UTS CRICOS PROVIDER CODE 00099F

28 June 2017

Distinguished Professor
Aaron Coutts
Faculty of Health
UNIVERSITY OF TECHNOLOGY SYDNEY

Dear Aaron,

UTS HREC ETH16-1082 – COUTTS (for JEFFRIES) – "Training load and injury in contemporary dance"

Thank you for your response to the Committee's comments for your project titled, "Extending the applicability of the ecological dynamics framework to measuring and understanding team performance in a military setting". The Committee agreed that this application now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application.

Your approval number is UTS HREC REF NO. ETH16-1082.

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.

The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements. Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat (Research.Ethics@uts.edu.au).
- The Principal Investigator will notify the UTS HREC of any event that requires a modification to the protocol or other project documents and submit any required amendments prior to implementation.
- The Principal Investigator will promptly report adverse events to the Ethics Secretariat (Research.Ethics@uts.edu.au). An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.

- The Principal Investigator will report to the UTS HREC annually and notify the HREC when the project is completed at all sites. The Principal Investigator will notify the UTS HREC of any plan to extend the duration of the project past the approval period listed above through the progress report.

- The Principal Investigator will obtain any additional approvals or authorisations as required (e.g. from other ethics committees, collaborating institutions, supporting organisations).

- The Principal Investigator will notify the UTS HREC of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

I also refer you to the AVCC guidelines relating to the storage of data, which require that data be kept for a minimum of 5 years after publication of research. However, in NSW, longer retention requirements are required for research on human subjects with potential long-term effects, research with long-term environmental effects, or research considered of national or international significance, importance, or controversy. If the data from this research project falls into one of these categories, contact University Records for advice on long-term retention.

If you have any queries about your ethics approval, or require any amendments to your research in the future, please do not hesitate to contact Research.Ethics@uts.edu.au.

Yours sincerely

Production Note:
Signature removed prior to publication.

Associate Professor
Beata Bajorek
Chairperson
UTS Human Research Ethics Committee

Appendix O: University ethics approval questionnaire study

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Human Research Ethics Committee
Ethics Secretariat C/O Research Office
15 Broadway, Ultimo NSW 2007

T: +61 2 9514 9681
Research.Ethics@uts.edu.au

PO Box 123
Broadway
NSW 2007
Australia
www.uts.edu.au

01 May 2020

Franco Impellizzeri
Faculty of Health
UNIVERSITY OF TECHNOLOGY SYDNEY

Dear Franco,

UTS HREC ETH20-4675 – FRANCO (for JEFFRIES) – “Validation of athletic monitoring tools”

Your local research office has reviewed your application and agreed that it now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application, subject to any conditions detailed in this document.

You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all UTS policies and guidelines including the Research Management Policy.

Your approval number is UTS HREC REF NO. ETH20-4675.

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.
The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements.
- Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat (Research.Ethics@uts.edu.au).

- The Principal Investigator will notify the UTS HREC of any event that requires a modification to the protocol or other project documents, and submit any required amendments prior to implementation. Instructions on how to submit an amendment application can be found here.

- The Principal Investigator will promptly report adverse events to the Ethics Secretariat. An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.

- The Principal Investigator will report to the UTS HREC annually and notify the HREC when the project is completed at all sites.

- The Principal Investigator will notify the UTS HREC of any plan to extend the duration of the project past the approval period listed above through the progress report.

- The Principal Investigator will obtain any additional approvals or authorisations as required (e.g., from other ethics committees, collaborating institutions, supporting organisations).

- The Principal Investigator will notify the UTS HREC of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

This research must be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

You should consider this your official letter of approval.

If you have any queries about this approval, or require any amendments to your approval in future, please do not hesitate to contact your local research office or the Ethics Secretariat.

Yours sincerely

Production Note:
Signature removed prior to publication.

Racheal Laugery
Acting Research Ethics Manager
UTS Human Research Ethics Committee

Appendix P: Informed consent form dance study

CONSENT FORM

INJURY IN PROFESSIONAL DANCERS: IDENTIFICATION OF RISK FACTORS AND PREVENTION STRATEGIES
UTS HREC REF NO. ETH16-1082

I ____________________ *[participant's name]* agree to participate in the research project "Injury in professional dancers: Identification of risk factors and prevention strategies" *(UTS HREC REF NO. ETH16-1082)* being conducted by Annie Jeffries, 15 Broadway, Ultimo, NSW, 2007, (phone). I understand that funding for this research has been provided by a UTS Doctoral scholarship.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to keep confidential all information including all conversations and discussions, materials and methods provided to me by the UTS research team.

I agree that the research data gathered from this project may be published in a form that:
☐ Does not identify me in any way
☐ May be used for future research purposes

I am aware that I can contact the researcher Annie Jeffries if I have any concerns about the research.

____ ____/____/____
Name and Signature [participant] Date

____ ____/____/____
Name and Signature [researcher or delegate] Date

NOTE:
This study has been approved by the University of Technology Sydney Human Research Ethics Committee [UTS HREC]. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au, and quote the UTS HREC reference number. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

Appendix Q: Informed consent form questionnaire study

INFORMATION SHEET AND CONSENT FORM FOR ONLINE SURVEYS

UTS ETH20-4675 - Validation of athletic monitoring tools

What is the research study about?
The purpose of this research/online survey is to validate several commonly used questionnaires for athlete monitoring.
You have been invited to participate because you are an elite athlete.

Who is conducting this research?
My name is Annie Jeffries and I am an PhD student at UTS. My supervisor is Professor Franco Impellizzeri (Franco.Impellizzeri@uts.edu.au). Additionally, the research team includes Professor Aaron Coutts, Dr Lee Wallace and Dr Alan McCall.

Inclusion/Exclusion Criteria
Before you decide to participate in this research study, we need to ensure that it is ok for you to take part.
Inclusion criteria: You are an elite, high level or professional athlete.
Exclusion criteria: You are less than 18 years of age.

Do I have to take part in this research study?
Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.
If you decide to participate, I will invite you to
- Read the information carefully (ask questions if necessary)
- Complete an online anonymous survey that takes approximately 5 minutes to complete.

You can change your mind at any time and stop completing the surveys without consequences.

Are there any risks/inconvenience?
We don't expect this questionnaire to cause any harm or discomfort.

What will happen to information about me?
Access to the online questionnaire is via the RedCap link. Submission of the online questionnaire/s is an indication of your consent. By clicking the RedCap link you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. Data will be stored in a password protected computer with only the research student having access. Your information will only be used for the purpose of this research project. We plan to publish the results in an academic journal.

What if I have concerns or a complaint?
If you have concerns about the research that you think I can help you with, please feel free to contact me *(*Annie.Jeffries@uts.edu.au)

If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number UTS 20 4675.