Measuring sport fantasy proneness and deflated reality in sport and performance

Zhang, Shuge; Boardley, Ian

DOI:
10.1016/J.PSYCHSPORT.2023.102491

Psychology of Sport & Exercise 69 (2023) 102491

Contents lists available at ScienceDirect

Psychology of Sport & Exercise

journal homepage: www.elsevier.com/locate/psychsport

Measuring sport fantasy proneness and deflated reality in sport and performance: Development and validation of two context-specific instruments

Shuge Zhang [a,*], Ian Boardley [b]

[a] School of Human Sciences, University of Derby, UK
[b] School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, UK

ARTICLE INFO

Keywords:
Sport fantasy
Deflated reality
Scale development
Construct validity
Measurement invariance
Temporal stability

ABSTRACT

Fantasy-prone personalities and generalized states of pessimism and hopelessness have been linked to various behavioural, cognitive, and health related outcomes in the general population. Nevertheless, to date, knowledge of sport-specific fantasy proneness and athletes' perception of deflated reality in sport is scarce, possibly due to a lack of appropriate psychometric instruments to examine these two important context-specific constructs. In this research, we developed the *Sport Fantasy Proneness Scale* (SFPS) and the *Deflated Reality in Sport Scale* (DRSS), first assessing the content validity of items for the instruments following a rigorous process. Through two cross-sectional samples (Study 1: Sample 1 $N = 255$, Sample 2 $N = 260$) and one longitudinal sample (Study 2: Sample 3 $N = 118$) of competitive athletes in the UK, we then examined factorial, convergent, discriminant, and concurrent validity and measurement invariance (Study 1), as well as internal consistency, and test-retest reliability (Study 2) of the two new scales. Results revealed robust construct validity and reliability of scores on the two instruments and suggested very good invariance when implementing the new scales to athletes from different sports and competitive levels and good invariance for cross-gender comparisons. The new scales fill a gap in the sport psychology literature and provide researchers and practitioners with robust psychometric instruments to examine new research questions and tackle issues relevant to athletes' sport-specific fantasy proneness and perceptions of deflated reality.

Fantasy proneness, or a fantasy-prone personality referring to a disposition of frequent and intensive fantasizing or imaginative experience, is first described in Wilson and Barber's (1982) intensive interview study of individuals who reported excellent hypnotic responsiveness. Individuals high in fantasy proneness (described as *fantasizers* by Wilson and Barber) reported conscious but immersive experiences of fantasy engagement, of which the fantasies were real as life and suspended the disbelief in nonreality (e.g., seeing imaginative experience as real despite knowing that the fantasized events are not actually happening; see also Plante et al., 2017). Later studies of high fantasy individuals suggested the role of aversive childhood events or trauma in developing fantasy proneness (Rhue & Lynn, 1987). It has been suggested that fantasy experience is a means to cope with adversity or escape from negative, distressed life experiences (Lawrence et al., 1995). Research has further considered fantasy-prone characteristics as a mental resource contributing to cognitive abilities such as problem-solving (Henderson & Wilson, 1991) and counterfactual thinking (Bacon et al., 2013).

Nevertheless, negative consequences are associated with high fantasy proneness and excessive fantasy engagement. For example, fantasy-prone individuals reported using fantasy to construct and maintain (overly) positive self-concept, which might partially explain the subset of fantasizers who exhibit a significant degree of psychopathology (Rhue & Lynn, 1987). Individuals engaging in excessive fantasies (e.g., addictive daydreaming) demonstrate more obsessive-compulsive behaviour and thoughts, greater attention deficit, and more frequent mental illness symptoms such as depression, anxiety, and hostility (Klinger et al., 2009; Somer et al., 2016). Being a fantasy-prone individual and engaging in day-to-day fantasy experiences, therefore, can lead to both positive (e.g., adaptation) and negative (e.g., escapism) outcomes that influence one's daily functioning (Plante et al., 2017).

Similar findings on the contradictory influences of fantasy-prone

* Corresponding author. School of Human Sciences, University of Derby, Kedleston Road, Derby, DE22 1GB, UK.
E-mail address: s.zhang@derby.ac.uk (S. Zhang).

https://doi.org/10.1016/j.psychsport.2023.102491
Received 20 March 2023; Received in revised form 10 July 2023; Accepted 10 July 2023
Available online 13 July 2023

characteristics exist in the sport and performance arena. Specifically, research on narcissism (i.e., a personality trait that features self-centredness and the tendency to use fantasy for constructing and maintaining a sense of grandiosity; Raskin & Novacek, 1991) in sport has consistently found athletes high in narcissism are more capable of performing under pressure (see Roberts et al., 2018, for review). Meanwhile, these individuals tend to engage poorly in training (Zhang et al., 2021), appear more prone to muscle dysmorphia (Boulter & Sandgren, 2022), demonstrate a greater risk of intentional doping (Zhang & Boardley, 2022), and contribute to intragroup conflict (Boulter et al., 2022). Although these narcissism-associated outcomes in sport are unlikely purely driven by fantasy-prone characteristics, one should not overlook the role of fantasy in sport and performance settings. For example, narcissism-related fantasy may magnify one's willingness to dominate and thus, on the one hand, makes one strive under challenging situations (Zhang et al., 2020), but on the other hand, could increase risk of moral disengagement and increase the likelihood of immoral conduct (e.g., antisocial behaviour) to maintain the fantasized performance advantage (Jones et al., 2017). Despite the potential influences of fantasy-prone characteristics and the use of fantasy in sport and performance settings, to date, no research has directly examined fantasy proneness in the competitive arena, possibly due to the lack of psychometrics for assessing sport-specific fantasies. As such, there is a need to develop a valid and reliable measure for assessing fantasy proneness in sport.

Meanwhile, the most widely used assessment for fantasy proneness (i.e., the *Creative Experience Questionnaire - CEQ*; Merckelbach et al., 2001) is a general measure based on Wilson and Barber's (1982) original conceptualization (see also Somer et al., 2016). Despite its strong psychometric properties, the CEQ has yet to receive any attention or application in sport research due to its heavy weight (i.e., 9 out of 25 items) on childhood imaginative experience to reflect early year experience as indicators of fantasy-prone characteristics (e.g., "*As a child, I have my own make-believe friend or animal*") and its lack of account for sporting scenarios due to its generalized conceptualization of fantasy experience (e.g., "Many of my fantasies are often just as lively as a good movie"). Therefore, one would not expect the CEQ to capture sport-specific fantasy proneness precisely. Informed by Wilson and Barber's (1982) conceptualization of fantasy proneness and applying it to the sport context, we defined sport fantasy proneness as "*an individual's disposition to engage extensively and deeply in fantasizing about oneself being an exceptional performer and receiving glory that is considered unrealistic*". Based on this definition, we aimed to develop a valid, reliable, and sport-specific measure for fantasy proneness, namely the *Sport Fantasy Proneness Scale* (SFPS).

Whilst addressing the need for a sport-specific measure for fantasy proneness, we also acknowledge the literature suggestion that there is a closely relevant psychological factor to one's fantasy prone characteristics; that is, an individual's perception of the self when facing harsh, deflated reality (see Twenge & Campbell, 2009). Indeed, research has consistently demonstrated that individuals holding an inflated self-view (e.g., via fantasy prone characteristics) tend to report a mismatched life reality (e.g., poor employment status, high debt loads) (Twenge, Campbell et al., 2012; Twenge, Konrath, Foster, Campbell, & Bushman, 2008). Individuals who perceive more threats regarding and hold more negative view towards a reality are more likely to discredit the reality and withdraw from challenges associated with the reality in favour of easier ways for self-enhancement, such as engaging in fantasy or activities that promote fantasy experience (McCain et al., 2015; Wallace, Ready, & Weitenhagen, 2009). In other words, deflated reality may augment fantasy engagement, making it difficult to determine if one is truly fantasy prone or simply using fantasy as a way to cope with deflated reality. As such, it is vital to account for one's perception of deflated reality when examining the effect of fantasy proneness.

In sport, while there are various opportunities for personal glory (e.g., achieving excellent performance), one must face the crucial reality that thousands of hours of deliberate practice do not guarantee an individual to develop a desirable sport career (e.g., medal count, winning, securing a professional contract). Individuals who are unable to receive admiration and glory to which they feel entitled may turn to engage in excessive fantasy (e.g., daydreaming about being an exceptional player) to satisfy their desperate need for self-enhancement, which can lead to increased incongruence between an inflated sense of self and the deflated reality (see also McCain et al., 2015). Such disparities are relevant to the negative influences of one's pessimism in sport (e.g., poor goal attainment, emotional instability, and avoidance coping; Gaudreau & Blondin, 2004; Nicholls et al., 2008) and may explain this challenging situation and the increasing tendency for mental health issues among elite sport players (Reardon et al., 2019). We, therefore, applied Marshall et al.'s (1992) theorizing on pessimistic orientation (i.e., a trait characterized by generalized pessimistic expectancy) to sport contexts and defined deflated reality in sport as "*an individual's pessimistic self-perception of his/her sport career involving both training/practice and performance/competition*".

However, to date, there is no validated psychometrics assessing one's pessimistic, hopeless feelings in face of sport-specific harsh reality (hereafter "deflated reality in sport"). Although a generalized measure of pessimistic orientation and hopeless feelings towards the reality one is facing exists (e.g., the *Hopelessness Scale - HS*; Beck et al., 1974), the HS was originally developed to capture individuals' disposition towards negative feelings and interpretations of their life experience and thus cannot precisely capture a sport-specific sense of deflated reality (e.g., ones holding negative views on their sport training does not necessarily appear pessimistic towards general life events beyond sport). As such, we also aimed to develop a valid and reliable measure to assess athletes' perceptions of deflated reality in sport based on our context-specific definition of the construct.

In sum, sport fantasy proneness and deflated reality in sport are two relevant constructs that have potential implications for athletes' training, performance, and health. Given the lack of measures that can precisely capture the two constructs in sport settings, we aimed to develop and validate two psychometric scales, including the *Sport Fantasy Proneness Scale* (SFPS) and the *Deflated Reality in Sport Scale* (DRSS). We present two studies, detailing how we developed the scale items, conducted a content validity assessment, followed by examinations of construct (i.e., factorial, convergent, discriminant, concurrent) validity, measurement invariance, internal consistency, as well as test-retest reliability and temporal stability of the scales.

1. Study 1: Methods

1.1. Study 1: Scale development

We developed a pool of items for the SFPS and the DRSS by rewriting appropriate items from the *Creative Experience Questionnaire* (CEQ; Merckelbach et al., 2001) and the *Hopelessness Scale* (HS; Beck et al., 1974), respectively, to a sport-specific context based on our definitions for sport fantasy proneness and deflated reality in sport. The CEQ was developed from Wilson and Barber's (1982) original conceptualization of fantasy proneness and is considered the most widely used assessment for the construct (Somer et al., 2016). The HS was developed to capture the generalized feelings of hopelessness and has been validated in varied, large populations (Fraser et al., 2014). Since both the CEQ (Merckelbach et al., 2001) and HS (Beck et al., 1974) are unidimensional scales, we expected the new scales (i.e., SFPS, DRSS) to be unidimensional, subject to confirmation via factor analyses.

The two initial item pools for the new scales each contained 12 items, which were subjected to content validity assessment. Following guidance (Almanasreh et al., 2019), we examined content validity through expert evaluation and opinion of item relevance and response format appropriateness. Specifically, one Professor in social psychology, two Professors in sport psychology, and four Associate Professors/Senior

Lecturers in sport psychology, who were known for expertise in psychometrics, individual differences, and personality, completed an independent content validity assessment for the initial SFPS and DRSS items. These experts rated each item using a Likert scale ranging from −3 (*not at all representative*) to 3 (*very representative*), commented on the relevance of each item, and the appropriateness of construct definition to the sporting context. We also asked the expert panel to provide suggestions on modifications, deletions, and additions to the initial SFPS and DRSS items if appropriate. Based on experts' evaluation and comments, we deleted one item (i.e., "*People close to me do not know I have such detailed fantasies about myself in sport*") and modified three items from the initial SFPS. We also modified four items from the initial DRSS following suggestions from the reviewing experts. All retained items (i. e., 11 for SFPS, 12 for SFPS) achieved an average of greater than 2 in representativeness ratings after removing any significantly deviated low rating from the second lowest rating (e.g., removing scoring 0 when the next lowest was 2). We used the post-review version of the SFPS (e.g., "*I have my own make-believe sporting abilities or skills*"; rated from "*1 – not at all*" to "*4 – somewhat true*" to "*7 – very much so*") and DRSS (e.g., "*Not seeing the bright side of my training*"; rated from "*1 – never*" to "*4 – sometimes*" to "*7 – very often*") for further validation.

1.2. Study 1: Participants

We recruited two samples of UK athletes to assess the factorial, convergent, discriminant, and concurrent validity of the SFPS and the DRSS. We also used the two samples to test measurement invariance across gender (i.e., male/female), sport type (i.e., team/individual), competitive level (i.e., regional/national/international), and the internal consistency of the new instruments. *Sample 1* ($N = 255$) participants were competitive athletes who averaged 24 years old (SD = 11.8) and 14 years of sport training (SD = 11.05). These participants (152 male) were from team (n = 205; e.g., football, netball, hockey) or individual (n = 50; e.g., tennis, swimming, badminton) sports, competing at regional (n = 208) or national/international (n = 47) level at the time of participation. *Sample 2* ($N = 260$) participants were competitive athletes who averaged 22 years old (SD = 5.03) and nine years of sport training (SD = 5.93). These participants (143 male) were from different team (n = 201; e.g., football, volleyball, basketball) or individual (n = 59; e.g., badminton, karate, climbing) sports, competing at regional (n = 181) or national/international (n = 78) at the time of participation. Both samples fulfilled Mundfrom et al.'s (2005) rule of thumb for sample size requirement when conducting factor analysis (i.e., 20 times the number of scale items).

1.3. Study 1: Measures

We employed the below measures to assess the convergent (i.e., the extent to which the measure score correlates to a conceptually similar construct), discriminant (i.e., the extent to which the measure score is distinguishable from that of a closely related construct), and concurrent validity (i.e., the extent to which the measure score is related to an external, relevant construct when collecting data at the same time) of scores generated with two new scales (i.e., SFPS, DRSS) based on Vaughn and Daniel's (2012) conceptualization of these validities. All measures were assessed in both samples.

Grandiose fantasy. We used the 7-item grandiose fantasy subscale of the *Pathological Narcissism Inventory* (PNI; Pincus et al., 2009) to assess the convergent validity of the SFPS. Grandiose fantasy assessed by Pincus et al.'s inventory can indicate the extent to which an individual fantasizes about oneself being admired and rewarded for exceptional accomplishments in general life domains, which is conceptually close to our theorizing on sport fantasy (i.e., an athlete fantasizing about oneself being an exceptional performer in his/her sport and receiving glory that is considered unrealistic). As such, we anticipated athletes high in grandiose fantasy in general life domains also tend to demonstrate high sport-specific fantasy. The grandiose fantasy items from the PNI ask participants to what extent they agree on a series of grandiose fantasy dispositions describing themselves (e.g., "I often fantasize about performing heroic deeds"), rated from 0 (*not at all like me*) to 5 (*very much like me*). We generated mean scores for further analysis. Cronbach's alpha ranged from 0.93 to 0.94.

Imagination. We used the 10-item imagination subscale of the *International Personality Item Pool* (IPIP; Goldberg, 1999) to assess the discriminant validity of the SFPS. IPIP imagination refers to an individual's dispositional tendency to use fantasy and imaginative minds as a way of creating a richer and more interesting world, which is an indicator of one's openness to experience (i.e., one of the big-five traits; Goldberg, 1999). Imagination is a healthy, general use of fantasy experience (see also Costa & McCrae, 1992) and, therefore, should be distinguishable from sport fantasy. As such, we anticipated athletes high in imagination would score relatively high in SFPS, but the correlation of scores from these two constructs should be weaker than that between grandiose fantasy and SFPS (the latter should be more conceptually close to each other). The imagination subscale of the IPIP contains five reversed items (e.g., "Seldom get lost in thought") and five non-reverse items (e.g., Like to get lost in thought), rated from 1 (*very inaccurate*) to 5 (*very accurate*). We generated item means with higher scores reflecting greater imagination. Cronbach's alpha ranged from 0.78 to 0.82.

Maladaptive daydreaming. We used the 16-item *Maladaptive Daydreaming Scale* (MDS; Somer et al., 2016) to assess the concurrent validity of the SFPS. Maladaptive daydreaming reflects excessive daydreams and fantasies that interfere with one's daily life functioning (Somer et al., 2016). This is similar to excessive sport fantasy that would have an impact on an athlete's training and performance, possibly due to increased incongruence between fantasized/inflated self and the tedious/mundane routine environments in sport (Zhang et al., 2021). As such, we expected a positive correlation between maladaptive daydreaming and SFPS scores. The MDS asks participants to recall certain maladaptive daydreaming experiences (e.g., "How much does your daydreaming interfere with your ability to get basic chores accomplished?" rated from "*0 – no interference at all*" to "*10 – extreme interference*"). We generated mean MDS scores for further analysis. Cronbach's alpha ranged from 0.94 to 0.95.

Athlete psychological strain. We used the 10-item Athlete Psychological Strain Questionnaire (APSQ; Rice et al., 2019) to assess the convergent validity of the DRSS. Psychological strain in sporting contexts reflects athletes' difficulties in adaptation or adjustment to a change in circumstances and impaired social or athletic functioning due to concerns about performance, self-regulatory issues, and externalized/maladaptive coping (Rice et al., 2019). This definition is conceptually close to our operational definition on deflated reality in sport that reflects an athlete's pessimistic, hopeless perception of one's training and performance. As such, we predicted a significant and positive correlation between scores on the APSQ and the DRSS. Each APSQ item asks participants to rate the frequency of certain sporting situations (e.g., "*I could not stop worrying about injury or my performance*"), rated from 1 (*none of the time*) to 5 (*all of the time*). We generated APSQ mean scores for further analysis. Cronbach's alpha was .87 and .82 for Samples 1 and 2, respectively.

Psychological distress. We used the 6-item *Kessler Psychological Distress Scale* (K6; Kessler et al., 2002) to assess the discriminant validity of the DRSS. Psychological distress refers to non-specific symptoms of mental health issues such as stress, anxiety, and depression (Kessler et al., 2002). One would expect an athlete high in psychological distress is likely to experience a greater sense of deflated reality in his/her sport career. However, an athlete's perception of deflated reality in sport should not be purely driven by a psychologically distressed state. In other words, scores of the K6 and the DRSS should be correlated but not too strongly. We, therefore, hypothesized a positive correlation between scores of the K6 and the DRSS, whilst such correlation should be weaker than that of the APSQ and DRSS. This is because the APSQ assesses context-specific distress in athletes, but the K6 measures generalized distress. Participants rated K6 items describing feelings or experiences of psychological distress (e.g., "… *restless or fidgety*") from 0 (*none of the time*) to 4 (*all of the time*). We generated K6 mean scores for further analysis. Cronbach's alpha was .90 and .87 for Samples 1 and 2, respectively.

Pessimistic orientation. We used the 6-item *Revised Life Orientation Test* (LOT-R; Marshall et al., 1992) to assess the concurrent validity of the DRSS. Pessimistic orientation refers to an individual's tendency to have negative expectations of life events and experiences and is commonly assessed via the LOT-R. One would expect an athlete high in pessimistic orientation to be more likely to perceive the reality in his/her sport career or training/performance as not going in the way one wants it to (thus deflated). As such, pessimistic orientation is a good external criterion for the DRSS, and we predicted a positive correlation between the LOT-R and the DRSS. LOT-R items (e.g., "I hardly ever expect things to go my way", rated from "*0 – strongly disagree*" to "*4 – strongly agree*") were coded so that higher scores reflect greater pessimistic orientation for further analysis. Cronbach's alpha was .78 and .73 for Samples 1 and 2, respectively.

1.4. Study 1: Procedures

With institutional ethics approval, we recruited Sample 1 using online data collection and Sample 2 using in-person data collection to approach athletes from more diversified backgrounds and for independence between the two samples. For Sample 1, we built a Qualtrics survey and recruited participants through Prolific (i.e., the UK's largest cloud-sourcing research participation platform; https://www.prolific.co). We screened potential participants by asking them a set of questions about their sport experience (e.g., training experience, competition status, club/team membership). Only those who met our inclusion criteria were invited to the study survey, providing their consent through Qualtrics. The entire survey took approximately 15 min to complete, and we offered each participant a £1.5 incentive through Prolific (i.e., based on the minimum Prolific pay rate and the average survey completion time). For Sample 2, two student research assistants were involved in delivering and collecting the study survey. Specifically, the assistants contacted sport teams/clubs in their corresponding universities and local regions and visited the team/club upon approval of the club manager or head coach. A visit was then arranged for a research assistant to attend a meeting with each participating team/club to do a study briefing, followed by the collection of the written consent and delivery of the printed study survey. All participants had opportunities to ask questions prior to partaking in this study and were thanked and debriefed upon survey completion.

1.5. Study 1: Data analysis

We used the IBM SPSS Version 28 for data processing and the Mplus Version 8 (Muthén & Muthén, 2017) for the main analysis. Missingness, skewness, and Kurtosis in the data were first checked. We then conducted Bartlett's test of sphericity and a Kaiser-Meyer-Olkin (KMO) test of sampling adequacy to assess the appropriateness of our data for factor analysis (i.e., a significant Bartlett's test and over 0.80 KMO is required). To explore the factor structure of the SFPS and the DRSS, we performed Exploratory Factor Analysis (EFA) in Sample 1 using an oblique type of rotation (i.e., Geomin) that allows emerging factors to correlate with each other (Kline, 2016). We extracted the number of factor(s) and decided factor structure of the new scales based on eigenvalue $\geq$1.00 and the comparison of 1-, 2-, and 3-factor EFA models.

Then, we conducted Confirmatory Factor Analysis (CFA) on the identified factor structure using both Samples 1 and 2, which allowed us to assess the replicability of the identified factor structure and item performance. Robust Maximum Likelihood (MLR) estimation was used to mitigate any potential impacts of data non-normality and for more accurate fit indices. Following recommendations (Hu & Bentler, 1999), robust Chi-square ($R\chi^2$), comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA) were used to assess and compare the model fit, with $\geq$0.95 CFI, $\leq$0.08 SRMR, $\leq$0.06 RMSEA indicating good model fit whilst $\geq$0.90 CFI, close to 0.08 SRMR and RMSEA reflecting acceptable model fit.

Once we had confirmed the factor structure of the SFPS and the DRSS, we used the combined sample (i.e., containing both Samples 1 and 2) to assess the measurement invariance using multi-group CFA (see Byrne, 2012). Specifically, we tested measurement invariance of the new scales by assessing the identified models across different genders (i.e., male and female), sport types (i.e., team and individual sports), and participating levels (i.e., regional, national, and international) against three invariance criteria. These criteria included *configural* invariance (i.e., identical factor structure, thus the same items measure the same factor across groups), *metric* invariance (i.e., identical factor structure + invariant factor loadings, thus each item contributes equally when measuring a certain factor in different groups), and *scalar* invariance (i.e., identical factor structure + invariant factor loadings + equal intercepts/threshold across groups, thus different groups use the response scale in the same way). We tested invariance by progressively imposing the appropriate constraints described above to the identified model and examined ΔCFI at each step when imposing constraints for assessing the three levels of invariance. Following guidance (Cheung & Rensvold, 2002), ΔCFI values of less than .01 change indicate invariance when progressively imposing the relevant constraints (i.e., configural, metric, scalar). We also performed a test of Chi-square change (see Kline, 2016) to determine if models with progressively imposed constraints for configural, metric, and scalar invariance differed significantly from each other (i.e., non-significance indicates invariance).

Furthermore, we examined the convergent, discriminant, and concurrent validity of the SFPS and the DRSS, testing correlations between scores of the new scales and that of the selected measures (see Measures). A large correlation (i.e., $r > 0.50$; Cohen et al., 2003), a conceptually distinguishable correlation (i.e., $r < 0.90$; Vaughn & Daniel, 2012), and an at least moderate correlation (i.e., $r > 0.30$; Cohen et al., 2003) provided evidence for convergent, discriminant, and concurrent validity, respectively. Last, we assessed correlation and Cronbach's alphas of the two scales in two samples, with 0.70, 0.80, and 0.90 reflecting good, very good, and excellent internal consistency, respectively.

2. Study 1: Results

2.1. Study 1: Preliminary analysis

No missing data was found in either sample, whilst skewness was within ±1.03 and ±0.98, and kurtosis was within ±1.32 and ±1.35 for Samples 1 and 2, respectively. KMO value was 0.93, and Bartlett's test was significant ($\chi^2 = 1572.43$, $df = 55$, $p = .00$), suggesting appropriateness for conducting factor analysis.

2.2. Study 1: Factor structure and validity

For both the SFPS and DRSS, EFA using Sample 1 revealed only one emerging factor yielded an eigenvalue greater than one (i.e., 5.95 eigenvalue for the first factor from the SFPS followed by 0.95 eigenvalue for the second potential factor; 5.81 eigenvalue for the first emerging factor from the DRSS followed by 0.96 eigenvalue for the second potential factor). Moreover, a comparison of 1-, 2-, and 3-factor EFA models suggested that the 1-factor model fit significantly better than 2- and 3-factor models (via Chi-square difference test; all $Ps < .01$).

However, we found one SFPS item (i.e., "*I believe in the existence of sporting heroes*") loaded consistently low (i.e., factor loading <0.32) and two other SFPS items (e.g., "*When I think about great athletes, I sometimes* *dream of how I could be like them*", "*At times, I imagine myself winning a difficult contest that most people would expect me to lose*") cross-loaded to two other factors (eigenvalues below 1). Similarly, we found two DRSS items (i.e., "*Performance not working out the way I want*", "*Disappointment in my sporting career despite strong ambition*") loaded consistently low (i. e., factor loading close to zero) and two other DRSS items (e.g., "*Foreseeing more performance setbacks than enhancements*", "*Suspecting performance setbacks might stay forever*") cross-loaded to two other factors (eigenvalues below 1). We, therefore, removed these items and retained all the others (see Table 1).

We then used CFA to assess the model fit of the 1-factor model for the final SFPS and DRSS suggested by EFA in Sample 1 and tested the replicability of the identified model using Sample 2. Results suggested that the identified model for SFPS fit excellently to both samples (*Sample 1 SFPS*: $R\chi^2 = 19.97$, df = 20, p = .46; RCFI = 1.00, SRMR = 0.03, RMSEA = 0.00; *Sample 2 SFPS*: $R\chi^2 = 21.90$, df = 20, p = .35; RCFI = 0.99, SRMR = 0.02, RMSEA = 0.02). Model fit for the final DRSS also performed well in both samples (*Sample 1 DRSS*: $R\chi^2 = 24.88$, *df* = 20, *p* = .21; RCFI = 0.99, SRMR = 0.03, RMSEA = 0.03; *Sample 2 DRSS*: $R\chi^2 =$ 37.85, *df* = 20, *p* = .01; RCFI = 0.96, SRMR = 0.04, RMSEA = 0.06). Table 1 displays all final items, factor loadings, and error variances of the SFPS and the DRSS.

2.3. Study 1: Measurement invariance

For the SFPS, *configural* invariance for gender (i.e., male/female) was supported though good to very good model fit (M1a: $R\chi^2 = 82.55$, *df* = 40; RCFI = 0.97, SRMR = 0.04, RMSEA = 0.06), with considerable support for *metric* (M1b) and *scalar* (M1c) invariance based on less than 0.01 ΔCFI from the baseline model. However, Chi-square difference tests revealed significant differences between M1c and M1b ($\Delta R\chi^2 = 18.96$, $\Delta df = 7$) and M1c and M1a ($\Delta R\chi^2 = 31.92$, $\Delta df = 14$). *Configural* invariance by sport type (i.e., team/individual; M2a: $R\chi^2 = 90.51$, *df* = 40; RCFI = 0.97, SRMR = 0.04, RMSEA = 0.07) and competitive level (i. e., regional/national/international; M3a: $R\chi^2 = 136.53$, *df* = 60; RCFI = 0.95, SRMR = 0.05, RMSEA = 0.08) was also supported, with considerable support for *metric* (M2b/M3b) and *scalar* (M2c/M3c) invariance based on less than 0.01 ΔCFI change from the baseline model. Chi-square difference tests suggested non-significant changes in model fit when progressively imposing constraints for testing invariance by sport type and competitive level. Overall, these results provided considerable evidence for the measurement invariance of the SFPS. Table 2 displays all statistics for measurement invariance testing of the SFPS.

For the DRSS, *configural* invariance by gender was also supported (M4a: $R\chi^2 = 79.63$, *df* = 40; RCFI = 0.96, SRMR = 0.04, RMSEA = 0.06), with considerable support for *metric* (M4b) and *scalar* (M4c) invariance based on less than 0.01 ΔCFI change. However, Chi-square difference tests revealed significant differences between M4b and M4a ($\Delta R\chi^2 =$ 16.22, $\Delta df = 7$), M4c and M4b ($\Delta R\chi^2 = 20.80$, $\Delta df = 7$), and M4c and M4a ($\Delta R\chi^2 = 36.38$, $\Delta df = 14$). *Configural* invariance by sport (i.e., team vs individual; M5a: $R\chi^2 = 95.46$, *df* = 40; RCFI = 0.95, SRMR = 0.04, RMSEA = 0.07) and competitive level (i.e., regional, national, international; M6a: $R\chi^2 = 164.35$, *df* = 60; RCFI = 0.91, SRMR = 0.06, RMSEA = 0.10) was supported, and support for *metric* (M5b/M6b) and *scalar* (M5c/M6c) invariance based on less than 0.01 ΔCFI change from the baseline model was also evidence. Chi-square difference tests showed non-significant changes in model fit when progressively imposing constraints for sport type, but significant differences were evident when comparing *metric* (M6b) and *scalar* (M6c) models for competitive level ($\Delta R\chi^2 = 26.95$, $\Delta df = 14$). Overall, these results provided considerable evidence for the DRSS's measurement invariance by gender and even stronger evidence for the DRSS's invariance for competitive level and sport type. Table 3 displays all statistics for measurement invariance of the DRSS.

Table 1
Study 1 standardised factor loadings, and error variances of the one-factor model for the final Sport Fantasy Proneness Scale (SFPS) and the Deflated Reality in Sport Scale (DRSS) items using confirmatory factor analysis (CFA).

	Sample 1 (N = 255)	Sample 2 (N = 260)
SFPS items		
1. I have my own make-believe sporting abilities or skills.	.47 (.78)	.50 (.75)
2. I sometimes think about being a sporting idol or the greatest athlete in my sport.	.76 (.42)	.64 (.59)
3. I can spend a long time fantasizing or daydreaming about being an exceptional player.	.85 (.28)	.81 (.35)
4. At times, I imagine myself celebrating exceptional sporting achievements.	.82 (.33)	.75 (.44)
5. Many of my sport fantasies have a realistic intensity.	.64 (.59)	.54 (.71)
6. My sport fantasies are often as lifelike as a good movie.	.82 (.33)	.80 (.36)
7. When I recall my exceptional performance, I have very vivid and lively memories.	.58 (.67)	.55 (.70)
8. I often engage in sport fantasies when I am alone or have nothing to do.	.79 (.36)	.76 (.42)
DRSS items		
1. Not seeing the bright side of my training.	.62 (.62)	.66 (.57)
2. Failing to fulfil my goals or expectations.	.75 (.43)	.71 (.50)
3. Performance not developing the way I want them to.	.72 (.48)	.65 (.58)
4. Sustained unsatisfactory training outcome(s).	.73 (.47)	.65 (.58)
5. I am not realising in my potential.	.63 (.61)	.56 (.69)
6. Losing when I should win.	.47 (.79)	.55 (.80)
7. My performances are frequently not going as I expected.	.72 (.48)	.69 (.53)
8. Feeling unlikely to get my desired achievements in sport.	.60 (.64)	.51 (.74)

Note. **CFA for Sample 1 SFPS**: $R\chi^2 = 19.97$, *df* = 20, *p* = .46; RCFI = 1.00, SRMR = 0.03, RMSEA = 0.00. **CFA for Sample 2 SFPS**: $R\chi^2 = 21.90$, *df* = 20, *p* = .35; RCFI = 0.99, SRMR = 0.02, RMSEA = 0.02. **CFA for Sample 1 DRSS**: $R\chi^2 =$ 24.88, *df* = 20, *p* = .21; RCFI = 0.99, SRMR = 0.03, RMSEA = 0.03. **CFA for Sample 2 DRSS**: $R\chi^2 = 37.85$, *df* = 20, *p* = .01; RCFI = 0.96, SRMR = 0.04, RMSEA = 0.06. *df* = degrees of freedom; $R\chi^2$ = robust Chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation.

2.4. Study 1: Convergent, discriminant, and concurrent validity

We examined the convergent validity of SFPS and DRSS scores by assessing their correlations with grandiose fantasy (Pincus et al., 2009) and athlete psychological strain (Rice et al., 2019), respectively. In both samples, SFPS scores were correlated positively and strongly with grandiose fantasy (Sample 1: $r = 0.69$, $p = .00$; Sample 2: $r = 0.63$, $p =$ .00), whilst DRSS scores were correlated moderately to strongly with athlete psychological strain (Sample 1: $r = 0.40$, $p = .00$; Sample 2: $r =$ 0.52, $p = .00$). Collectively, evidence for convergent validity was stronger for the SFPS than the DRSS.

We also examined the discriminant validity of SFPS and DRSS scores by assessing their correlations with IPIP imagination (Goldberg, 1999) and psychological distress (Kessler et al., 2002), respectively. In both samples, SFPS (Sample 1: $r = 0.41$, $p = .00$; Sample 2: $r = 0.35$, $p = .00$) and DRSS (Sample 1: $r = 0.33$, $p = .00$; Sample 2: $r = 0.45$, $p = .00$) scores correlated positively and moderately-to-strongly to their target construct. These correlations were significant and distinguishable (i.e., $r < 0.90$; Vaughn & Daniel, 2012), suggesting individual differences in sport fantasy cannot be fully explained by one's dispositional imagination, and deflated reality cannot be fully explained by one's extent of psychological distress. Overall, evidence supports the discriminant validity of scores on the two new scales.

To establish evidence for concurrent validity, we used maladaptive daydreaming (Somer et al., 2016) and pessimistic orientation (Scheier et al., 1994) as the external criterion for the SFPS and the DRSS, respectively. As predicted, SFPS scores correlated significantly to

Table 2
Study 1 summary of fit indices for test of measurement invariance between male and female, and between team and individual sport for Sport Fantasy Proneness Scale (SFPS) using combined Samples 1 and 2 (N = 515).

Model	*Df*	Rχ^2	RCFI	SRMR	RMSEA	Comparison	ΔRχ^2	Δ*df*
Sex (male, female)								
M1a Configural Invariance	40	82.55	.97	.04	.06	M1b vs. M1a	12.93	7
M1b Metric Invariance	47	95.82	.96	.05	.04	M1c vs. M1b	18.96*	7
M1c Scalar Invariance	54	113.85	.96	.06	.07	M1c vs. M1a	31.92*	14
Sport (team, individual)								
M2a Configural Invariance	40	90.51	.97	.04	.07	M2b vs. M2a	3.12	7
M2b Metric Invariance	47	94.89	.97	.04	.06	M2c vs. M2b	10.68	7
M2c Scalar Invariance	54	105.95	.96	.04	.06	M2c vs. M2a	13.68	14
Level (regional, national, international)								
M3a Configural Invariance	60	136.53	.95	.05	.08	M3b vs. M3a	17.32	14
M3b Metric Invariance	74	152.49	.95	.07	.08	M3c vs. M3b	21.36	14
M3c Scalar Invariance	88	174.30	.94	.08	.08	M3c vs. M3a	38.35	28

Note. *df* = degrees of freedom; Rχ^2 = robust Chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; * indicates significant Chi-square change at 0.05 alpha level.

Table 3
Study 1 summary of fit indices for test of measurement invariance between male and female, and between team and individual sport for Deflated Reality in Sport (DRSS) using combined Samples 1 and 2 (N = 515).

Model	*df*	Rχ^2	RCFI	SRMR	RMSEA	Comparison	ΔRχ^2	Δ*df*
Sex (male, female)								
M4a Configural Invariance	40	79.63	.96	.04	.06	M4b vs. M4a	16.22*	7
M4b Metric Invariance	47	95.86	.95	.06	.07	M4c vs. M4b	20.80*	7
M4c Scalar Invariance	54	115.28	.94	.07	.07	M4c vs. M4a	36.38*	14
Sport (team, individual)								
M5a Configural Invariance	40	95.46	.95	.04	.07	M5b vs. M5a	3.38	7
M5b Metric Invariance	47	97.75	.95	.05	.07	M5c vs. M5b	6.17	7
M5c Scalar Invariance	54	105.32	.95	.05	.06	M5c vs. M5a	9.16	14
Level (regional, national, international)								
M6a Configural Invariance	60	164.35	.91	.06	.10	M6b vs. M6a	14.47	14
M6b Metric Invariance	74	174.84	.91	.07	.09	M6c vs. M6b	26.95*	14
M6c Scalar Invariance	88	202.83	.90	.08	.09	M6c vs. M6a	39.11	28

Note. *df* = degrees of freedom; Rχ^2 = robust Chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; * indicates significant Chi-square change at 0.05 alpha level.

maladaptive daydreaming (Sample 1: $r = 0.46$, $p = .00$; Sample 2: $r = 0.30$, $p = .00$), and DRSS scores correlated significantly to pessimistic orientation (Sample 1: $r = 0.28$, $p = .00$; Sample 2: $r = 0.28$, $p = .00$).

Table 4
Study 1 correlations of the Sport Fantasy Proneness Scale (SFPS) and the Deflated Reality in Sport Scale (DRSS) scores with target measures for assessing convergent, discriminant, and concurrent validity.

	Sample 1 (N = 255)		Sample 2 (N = 260)	
	SFPS	DRSS	SFPS	DRSS
Grandiose fantasy	.69	.34	.63	.32
IPIP imagination	.41	.24	.33	.03 (*ns*)
Maladaptive daydreaming	.46	.33	.30	.28
Athlete Psychological Strain	.30	.40	.27	.52
Psychological distress	.13	.35	.04 (*ns*)	.45
Pessimistic orientation	.02 (*ns*)	.28	−.14	.28
SFPS	–	.26	–	.18
DRSS	.26	–	.18	–
Cronbach's alpha	.90	.87	.85	.82

Note. Grandiose fantasy, international personality item pool (IPIP) imagination, and maladaptive daydreaming were used to test convergent, discriminant, and concurrent validity for SFPS, respectively. Athlete psychological strain, psychological distress, and pessimistic orientation were used to test convergent, discriminant, and concurrent validity for DRSS, respectively. All correlations were significant at 0.01 alpha level unless marked *ns*.

These results support the concurrent validity of the new scales. Table 4 presents all details for assessing convergent, discriminant, and concurrent validity.

2.5. Study 1: Internal consistency and other correlations

Cronbach's alphas were very good to excellent for the SFPS and the DRSS across the two samples. Alpha values for the SFPS were .90 and .85 in Samples 1 and 2, respectively. Alpha values for the DRSS were 0.87 and 0.82 in Samples 1 and 2, respectively. Overall, the results provided strong support for the internal consistency of the SFPS and the DRSS.

Additionally, SFPS scores correlated moderately and consistently with athlete psychological strain (Sample 1: $r = 0.30$, $p = .00$; Sample 2: $r = 0.27$, $p = .00$), whilst DRSS scores correlated moderately and consistently with grandiose fantasy (Sample 1: $r = 0.34$, $p = .00$; Sample 2: $r = 0.32$, $p = .00$) and maladaptive daydreaming (Sample 1: $r = 0.32$, $p = .00$; Sample 2: $r = 0.28$, $p = .00$). The correlation between SFPS and DRSS scores was small to moderate (Sample 1: $r = 0.26$, $p = .00$; Sample 2: $r = 0.18$, $p = .00$). Table 4 displays Cronbach's alphas of the SFPS and DRSS and correlations between all study measures.

2.6. Study 1: Discussion

To summarize Study 1 briefly, we developed an item pool measuring sport fantasy proneness (i.e., SFPS) and deflated reality in sport (i.e., DRSS) and went through a process engaging an expert panel to assess the content validity of the new items. We then revised and improved items for the new scales and tested their psychometric properties, including factorial validity, construct validity (i.e., convergent, discriminant, concurrent), measurement invariance, and internal consistency in two independent samples of competitive athletes (total $N = 515$). Results provided support for the robust psychometric properties of the new scales. However, due to the cross-sectional nature of Study 1, we could not assess the test-retest reliability and temporal stability of the SFPS and the DRSS. Therefore, we employed a longitudinal design to conduct these tests in the following study.

3. Study 2: Method

3.1. Study 2: Participants

118 competitive athletes (58 males) who on average were 28 years old (SD = 6.50) and had 12 years of training experience (SD = 8.14) participated in the study. These athletes were from either team (n = 87; e.g., football, rugby, netball) or individual sport (n = 31; e.g., athletics, weightlifting, swimming), competing at regional (n = 89) or national/international level (n = 29) at the time of data collection. All participants completed the SFPS and the DRSS twice, two weeks apart. Sample size calculation for reliability studies (see Borg et al., 2022) suggested that we need a minimum of 99 participants to retain 0.80 power in detecting high reliability; that is, an intra-class correlation (ICC) of two repeated measures equal to 0.85, with minimum acceptable ICC set at 0.75 (i.e., moderate reliability). The Study 2 sample fulfils the need for assessing the test-retest reliability of the SFPS and DRSS.

3.2. Study 2: Measures

We used the final SFPS and DRSS developed in Study 1 for assessing sport fantasy proneness and deflated reality in sport, respectively, in the test and retest two weeks apart. Cronbach's alphas of the SFPS (i.e., 0.87 at both time points) and the DRSS (i.e., 0.84 at Time 1, 0.88 at Time 2) indicated very good internal consistency.

3.3. Study 2: Procedures

With institutional ethics approval, we used an identical approach to data collection for Sample 1 in Study 1 to recruit participants from the UK's largest crowdsourcing platform for research participants (i.e., Prolific) for Study 2. Only those who met our inclusion criteria (e.g., currently engaging in sport training and competing for a certain club/team) and did not participate in Study 1 were eligible for participation. These participants received the study link through Prolific to access the study online survey built in Qualtrics. Once the first survey (Time 1) was completed, a scheduled release of the invite was set up to allow anyone who completed the Time 1 survey to access and complete the Time 2 survey in an exact two-week time (i.e., 14 calendar days). All participants provided consent before participation and were debriefed on completing each survey. Those who completed each survey received £0.50 as compensation.

3.4. Study 2: Data analysis

We used the IBM SPSS Version 28 for the test-retest analysis and took an identical approach to Study 1 when checking missingness, skewness, and Kurtosis in data prior to running further analysis. We then calculated ICCs of the repeatedly measured SFPS and DRSS for assessing test-retest reliability, with 0.50, 0.75, and 0.90 ICC indicating acceptable, good, and excellent reliability, respectively (see Borg et al., 2022). We also assessed the temporal stability of each SFPS and DRSS item following recommendations (see Nevill et al., 2001). Specifically, we reported mean, SD, minimum/maximum changes, and test-retest differences for each item. We also calculated the proportion of agreement (PA; i.e., percentage of participants whose test-retest differences in score were within ±1). According to Nevill et al. (2001), ≥0.90 PA on a 5-point scale when test-retested within two-week time indicates good score stability. However, the PA is normally significantly lower than 0.90 for a 7-point scale like the SFPS and the DRSS. Thus, we used 0.90 as a reference score rather than a strict criterion, with closer to 0.90 PA reflecting greater item stability (see also Boardley et al., 2018).

4. Study 2: Results

No missing data was found in both scales. Skewness and kurtosis were within ±0.67 and ±1.16, respectively, for both scales in the test-retest. ICC for the average scores of SFPS and DRSS over the two testing time points was .94 and .87, respectively, indicating very good to excellent test-retest reliability. For the SFPS items, scores of approximately 80% participants remained stable in the test-retest (i.e., test-retest score differences within ±1 assessed by PA). Three items achieved over 0.80 PA (i.e., items 3, 7, 8), four items achieved 0.77-0.80 PA (i.e., items 1, 2, 5, 6), and one item achieved 0.72 PA (i.e., item 5). For the DRSS items, scores of approximately 84.9% participants remained stable in the test-retest. Two items achieved 0.90 PA (i.e., items 6, 7), five items achieved over 0.80 PA (i.e., items 1, 2, 3, 4, 8), and one item achieved 0.77 PA. Table 5 displays all details of the temporal stability test.

5. General discussion

Research has highlighted the potential importance of an individual's generalized fantasy-prone characteristics and pessimistic orientation in health (e.g., Klinger et al., 2009; Marshall et al., 1992; Somer et al., 2016) and performance contexts (e.g., Bacon et al., 2013; Gaudreau & Blondin, 2004; Henderson & Wilson, 1991; Nicholls et al., 2008). Knowledge is also available regarding risks associated with high fantasy prone characteristics and its incongruence with one's perceived reality (McCain et al., 2015). However, little is known regarding sport-specific fantasy proneness (i.e., sport fantasy proneness) and pessimistic perception of sporting experience (i.e., deflated reality in sport) and their influences on competitive sport participants, nor are there any

Table 5
Study 2 temporal stability of the Sport Fantasy Proneness Scale (SFPS) and the Deflated Reality in Sport Scale (DRSS) items (N = 118).

Item	Time 1		Time 2		Time 2 – Time 1 Differences		Proportion Agreement
	Mean	SD	Mean	SD	Min.	Max.	% of within ±1 change
SFPS1	3.54	1.73	3.67	1.79	−3	4	79.8%
SFPS2	4.17	1.92	4.31	1.75	−3	4	77.4%
SFPS3	4.42	1.72	4.23	1.89	−4	3	81.5%
SFPS4	4.98	1.56	4.81	1.69	−5	3	78.8%
SFPS5	4.23	1.54	4.14	1.39	−3	4	72.3%
SFPS6	4.13	1.71	4.16	1.66	−3	4	77.2%
SFPS7[a]	4.87	1.45	4.68	1.47	−4	3	87.4%
SFPS8	4.05	1.72	4.08	1.58	−3	3	83.2%
DRSS1[a]	3.58	1.32	3.83	1.45	−6	3	80.6%
DRSS2	3.86	1.10	3.91	1.17	−6	3	80.7%
DRSS3	4.11	1.01	4.25	1.14	−2	3	89.0%
DRSS4	3.63	1.13	3.72	1.21	−3	3	84.0%
DRSS5[a]	3.95	1.35	4.17	1.40	−3	3	76.5%
DRSS6	4.06	1.13	4.14	1.13	−4	4	90.7%
DRSS7	3.58	1.13	3.64	1.16	−3	3	93.3%
DRSS8	3.67	1.30	3.83	1.28	−3	3	84.0%

Note.
[a] Indicates significant change from Time 1 to Time 2 at .05 alpha level.

instruments that can assess these constructs in sport. Therefore, in the present research, we developed and validated two psychometric scales, namely the *Sport Fantasy Proneness Scale* (SFPS) and the *Deflated Reality in Sport Scale* (DRSS), to measure the two important, relevant but overlooked constructs.

5.1. Research highlights

In two studies involving three independent athletic samples (cross-sectional Sample 1 $N = 255$, cross-sectional Sample 2 $N = 260$, longitudinal Sample 3 $N = 118$), we demonstrated evidence for excellent factorial structure, good to very good convergent, concurrent and discriminant validity, very good measurement invariance (especially across sports and competitive levels), and very good to excellent internal consistency and test-retest reliability.

To expand, support for the unidimensional nature of the SFPS and DRSS was excellent, aligning with their conceptualizations and related measurements in general life domains (i.e., Beck et al., 1974; Merckelbach et al., 2001). Sport fantasy proneness was consistently correlated strongly to grandiose fantasy (supporting convergent validity), moderately to strongly with imagination (supporting discriminant validity) and maladaptive daydreaming[1] (supporting concurrent validity). Deflated reality in sport was consistently correlated moderately to strongly with athlete psychological strain (partially supporting convergent validity)[2], moderately to strongly with psychological distress (supporting discriminant validity), and weakly to moderately with pessimistic orientation (supporting concurrent validity).

Sport fantasy proneness also manifested a consistent, moderate correlation with athlete psychological strain but inconsistent correlations with psychological distress (significant in Sample 1 and insignificant in Sample 2) and pessimistic orientation (insignificant in Sample 1 and significant in Sample 2). These findings agree with the literature on negative influences of fantasy proneness on mental health (e.g., Klinger et al., 2009; Somer et al., 2016) and further support the context-specific nature of sport fantasy proneness as it was more related to athlete-specific psychological strain rather than psychological distress and pessimistic orientation in general life domains. Similarly, deflated reality in sport was consistently and moderately associated with grandiose fantasy and maladaptive daydreaming in general (non-sport) domains, but inconsistently associated with one's imagination (significant in Sample 1 and insignificant in Sample 2). Considering the undesirable outcomes linked to grandiose fantasy (e.g., impaired self-esteem and increased aggression; Pincus et al., 2009) and maladaptive daydreaming (e.g., more frequent anxiety and depression symptoms; Somer et al., 2016), our findings suggested that pessimistic perception of one's sport experience, or a sense of deflated reality in sport, may exert negative mental health implications beyond sport (in general life).

The new scales also demonstrated sound measurement invariance using Cheung and Rensvold's (2002) criterion and the additional Chi-square difference test (Kline, 2016). Specifically, evidence of invariance from multi-group CFA was strong for the two scales when measuring across individual and team sports (i.e., less than .01 ΔCFI from the baseline configural model to the scalar model and insignificant Chi-square differences). This suggests the two scales perform well for comparisons between individual versus team sports, as individuals from different sport types demonstrated identical factor structure (i.e., configural invariance), equal weight of scale items (i.e., metric invariance), and similar response pattern and scale interpretation (i.e., scalar invariance). When assessing invariance across varied competitive levels, evidence was strong for the SFPS on scalar invariance and for the DRSS on the metric invariance (considerable evidence for scalar invariance in DRSS based on less than 0.01 ΔCFI from the baseline configural model to the scalar model but insignificant Chi-square differences). This suggests the SFPS performs well for comparison among athletes competing at different levels (i.e., fulfilling the three levels of invariances we tested), but one should pay some attention when using the DRSS for cross-level comparison as the DRSS items may be interpreted differently by athletes competing at different levels (i.e., relative weak scalar invariance but good configural/metric invariance). When assessing invariance across different genders, evidence was strong for both scales on configural invariance, with considerable support for invariance at metric and scalar levels. This suggests the two scales hold identical factorial structure in males and females, which fulfils the basic requirement for invariance (see Kline, 2016). However, one should avoid direct across-gender comparison using scores from the two scales, because some of the scale items could weight or be interpreted differently between males and females (i.e., lack of metric and scalar invariance).

Moreover, internal consistency and test-retest reliability were very good to excellent for the new scales. Within-person variation (i.e., 1 – ICC for test-retest) was larger in DRSS (0.13) than in SFPS (i.e., 06), which suggests that deflated reality in sport contains more state-like components than sport fantasy proneness and thus is likely more sensitive to change over time. However, temporal stability measured by the proportion of agreement (i.e., percentage of participants whose test-retest differences in score were within ±1) appeared to be higher for the DRSS than the SFPS. This finding suggests one's sport fantasy proneness may be more susceptible to random response-error or acute change (e.g., pre-/post-training or competition) for some – but not the majority of – individuals across a short timeframe.

Overall, we provided valid and reliable psychometrics for assessing sport-specific fantasy proneness (i.e., SFPS) and deflated reality (i.e., DRSS) through this research. The new scales enable researchers and practitioners to examine how sport fantasy and deflated reality influence athletes' training and performance, morality, mental wellbeing, and other performance- or health-related outcomes in sport, especially when a high disparity between fantasy and reality exists. Future research would also do well to explore and investigate possible mediators and moderators of the effects of sport fantasy proneness and deflated reality in sport, which should provide useful information regarding how to maximise athletes' training, performance, and health.

6. Limitations and future direction

While providing two new, useful psychometric tools for assessing sport fantasy proneness (i.e., SFPS) and deflated reality in sport (i.e., DRSS), we are aware that the current project is not without limitation. One limitation in the scale development is that the draft of the item pool for the new scales was not underpinned by prior qualitative investigation due to a lack of such research. However, we provided a working definition for sport fantasy proneness and deflated reality in sport based on literature conceptualizations of generalized fantasy-prone characteristics (Wilson & Barber, 1982) and pessimistic orientation (Marshall et al., 1992), respectively. We then developed initial items for the new scales by rewriting appropriate items from well-established measures for generalized fantasy proneness (Merckelbach et al., 2001) and hopelessness (Beck et al., 1974) to fit into sport and performance contexts guided by our working definition of the two constructs. The working definition of the constructs and initial items pool went through a strict

[1] Thanks for a comment from a reviewer, it is noteworthy that the maladaptive daydreaming scale used in this study contains items assessing interference in daily performance resulting from daydreaming. For comparison, our SFPS does not include such items because the purpose of the SFPS is to measure one's disposition towards fantasy. This explains why the correlation between the two was not strong and further rationalise why we used maladaptive daydreaming for assessing concurrent (i.e., correlation to scores of an external, relevant construct) not convergent validity (i.e., correlation to scores of a conceptually similar construct).

[2] We acknowledge the correlation between deflated reality in sport and athlete psychological strain might not be strong enough to support convergent validity. We have discussed this issue in more details in the limitations and future directions section.

expert panel review following guidance (Almanasreh et al., 2019), based on which we made revisions and revised the two new scales for further validation. We, therefore, believe the two new scales are robust in their content validity. Future research would do well to adopt a qualitative method to explore the manifestation of sport fantasy proneness and deflated reality in sport.

Second, the use of athlete psychological strain as a convergent validity measure for the DRSS might invite concern. While the assessment of convergent validity requires a test of correlation between the target measure (e.g., DRSS) and another test that assesses the same or similar construct (Vaughn & Daniel, 2012), there is a lack of established measures in sport that measure deflated reality. Considering the context-specific nature of deflated reality in sport, we identified athlete psychological strain (Rice et al., 2019) as being one of the closest measures to deflated reality in sport. This is because, by our definition, a sense of deflated reality in sport can be amplified by concerns about performance, self-regulatory issues, and maladaptive coping that can be assessed by the measure of athlete psychological strain (Rice et al., 2019). However, athletes' concerns captured in Rice et al.'s athlete psychological strain measure (e.g., "I found training more stressful") do not capture the pessimistic, hopeless states that are central to our measure of deflated reality in sport (e.g., "Not seeing the bright side of my training"). This may explain why the correlation between deflated reality in sport and athlete psychological strain (i.e., 0.40-0.52) was not as strong as one would like to see for convergent validity (i.e., greater than 0.50; Cohen et al., 2003; Vaughn & Daniel, 2012).

Also, despite conceptualizing sport fantasy proneness and deflated reality in sport as two relevant concepts (i.e., athletes holding a great sense of deflated reality may be more fantasy-prone in sport), we did not make a specific hypothesis and examine the interplay between the two constructs in the present research. Nevertheless, it is noteworthy that scores of the two new scales correlated to each other relatively weakly (i.e., 0.18-0.26), and thus a mixed profile of sport fantasy proneness and sense of deflated reality in sport could exist (i.e., low-low, low-high, high-low, high-high). It is possible that potential influences of sport fantasy proneness (especially negative ones) on training- and performance-related outcomes can be magnified by the sense of deflated reality in sport. This is because highly fantasy-prone individuals are more likely to engage in fantasy as an approach to compensate for the unpleasant and unwanted reality (Lawrence et al., 1995; Rhue & Lynn, 1987), which in sport, may link to behavioral avoidance (Corr, 2013), intention to gain unfair advantages (Zhang & Boardley, 2022), and other undesirable outcomes that have impacts on training and performance, especially in face of harsh reality. We encourage future research to consider the potential interaction effects between sport fantasy proneness and deflated reality in sport.

7. Conclusion

In the present research, we developed and validated the *Sport Fantasy Proneness Scale* (SFPS) and the *Deflated Reality in Sport Scale* (DRSS) to capture the two important but previously overlooked psychological constructs in sport, through a rigorous set of processes (see Table 1 for the final items of the two instruments). The new scales fill in a gap in sport psychology literature and provide researchers and practitioners with robust psychometric instruments to examine new research questions and tackle issues relevant to athletes' sport-specific fantasy proneness and perception of deflated reality. We welcome future studies to further assess psychometric properties and measurement invariance of the two scales in different cultures and languages which should facilitate wider collaboration in research and practice using the two new scales.

Declaration of competing interest

This statement is to confirm that the authors do not have any conflicts or interests that might be interpreted as influencing the research. The first author of the paper will be serving as the corresponding author for this manuscript. All of the authors listed have agreed to the by-line order and to submission of the manuscript in the current form.

Data availability

Data will be made available on request.

References

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy, 15*, 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Bacon, A. M., Walsh, C. R., & Martin, L. (2013). Fantasy proneness and counterfactual thinking. *Personality and Individual Differences, 54*, 469–473. https://doi.org/10.1016/j.paid.2012.09.023

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42*, 861–865. https://doi.org/10.1037/h0037562

Boardley, I. D., Smith, A. L., Mills, J., Grix, J., Wynne, C., & Wilkins, L. (2018). Development of moral disengagement and self-regulatory efficacy assessments relevant to doping in sport and exercise. *Psychology of Sport and Exercise, 36*, 57–70. https://doi.org/10.1016/j.psychsport.2018.01.007

Borg, D. N., Bach, A. J. E., O'Brien, J. L., & Sainani, K. L. (2022). Calculating sample size for reliability studies. *PM&R, 14*, 1018–1025. https://doi.org/10.1002/pmrj.12850

Boulter, M. W., Roberts, R., & Hardy, J. (2022). Upsetting the apple cart: Within-team profiles of intragroup conflict and their associations with narcissism. *Psychology of Sport and Exercise, 63*. https://doi.org/10.1016/j.psychsport.2022.102291

Boulter, M. W., & Sandgren, S. S. (2022). Me, myself, and my muscles: Associations between narcissism and muscle dysmorphia. *Eating Disorders, 30*, 110–116. https://doi.org/10.1080/10640266.2021.1930348

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge/Taylor & Francis Group.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the bahvioural sciences* (3rd ed.). Hillsdale: Erlbaum.

Corr, P. (2013). Approach and avoidance behaviour: Multiple systems and their interactions. *Emotion Review, 5*, 285–290. https://doi.org/10.1177/1754073913477507

Costa, P. T., & McCrae, R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five factor model (NEO-FFI) professional manual*. Psychological Assessment Center.

Fraser, L., Burnell, M., Salter, L. C., Fourkala, E.-O., Kalsi, J., Ryan, A., Gessler, S., Gidron, Y., Steptoe, A., & Menon, U. (2014). Identifying hopelessness in population research: A validation study of two brief measures of hopelessness. *BMJ Open, 4*, Article e005093. https://doi.org/10.1136/bmjopen-2014-005093

Gaudreau, P., & Blondin, J.-P. (2004). Differential associations of dispositional optimism and pessimism with coping, goal attainment, and emotional adjustment during sport competition. *International Journal of Stress Management, 11*, 245–269. https://doi.org/10.1037/1072-5245.11.3.245

Goldberg. (1999). A broad-bandwidth, public domain, personality inventory. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in europe, 7* pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Henderson, B. B., & Wilson, S. E. (1991). Intelligence and curiosity in preschool children. *Journal of School Psychology, 29*, 167–175. https://doi.org/10.1016/S0022-4405(05)80009-3

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

Jones, B. D., Woodman, T., Barlow, M., & Roberts, R. (2017). The darker side of personality: Narcissism predicts moral disengagement and antisocial behavior in sport. *The Sport Psychologist, 31*, 109–116. https://doi.org/10.1123/tsp.2016-0007

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959–976. https://doi.org/10.1017/S0033291702006074

Kline, R. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford.

Klinger, E., Henning, V. R., & Janssen, J. M. (2009). Fantasy-proneness dimensionalized: Dissociative component is related to psychopathology, daydreaming as such is not. *Journal of Research in Personality, 43*, 506–510. https://doi.org/10.1016/j.jrp.2008.12.017

Lawrence, T., Edwards, C., Barraclough, N., Church, S., & Hetherington, F. (1995). Modelling childhood causes of paranormal belief and experience: Childhood trauma and childhood fantasy. *Personality and Individual Differences, 19*, 209–215. https://doi.org/10.1016/0191-8869(95)00034-4

Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers, R. R. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology, 62*, 1067–1074. https://doi.org/10.1037/0022-3514.62.6.1067

McCain, J., Gentile, B., & Campbell, W. K. (2015). A psychological exploration of engagement in geek culture. *PLoS One, 10*, 1–38. https://doi.org/10.1371/journal.pone.0142200

Merckelbach, H., Horselenberg, R., & Muris, P. (2001). The creative experiences Questionnaire (CEQ): A brief self-report measure of fantasy proneness. *Personality and Individual Differences, 31*, 987–995. https://doi.org/10.1016/S0191-8869(00)00201-4

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*, 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén www.StatModel.com.

Nevill, A. M., Lane, A. M., Kilgour, L. J., Bowes, N., & Whyte, G. P. (2001). Stability of psychometric questionnaires. *Journal of Sports Sciences, 19*, 273–278. https://doi.org/10.1080/026404101750158358

Nicholls, A. R., Polman, R. C. J., Levy, A. R., & Backhouse, S. H. (2008). Mental toughness, optimism, pessimism, and coping among athletes. *Personality and Individual Differences, 44*, 1182–1192. https://doi.org/10.1016/j.paid.2007.11.011

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the pathological narcissism inventory. *Psychological Assessment, 21*, 365–379. https://doi.org/10.1037/a0016530

Plante, C. N., Reysen, S., Groves, C. L., Roberts, S. E., & Gerbasi, K. (2017). The fantasy engagement scale: A flexible measure of positive and negative fantasy engagement. *Basic and Applied Social Psychology, 39*, 127–152. https://doi.org/10.1080/01973533.2017.1293538

Raskin, R., & Novacek, J. (1991). Narcissism and the use of fantasy. *Journal of Clinical Psychology, 47*, 490–499. https://doi.org/10.1002/1097-4679

Reardon, C. L., Hainline, B., Aron, C. M., Baron, D., … Engebretsen, L. (2019). Mental health in elite athletes: International Olympic Committee consensus statement. *British Journal of Sports Medicine, 53*, 667–699. https://doi.org/10.1136/bjsports-2019-100715

Rhue, J. W., & Lynn, S. J. (1987). Fantasy proneness and psychopathology. *Journal of Personality and Social Psychology, 53*, 327–336. https://doi.org/10.1037/0022-3514.53.2.327

Rice, S. M., Parker, A. G., Mawren, D., Clifton, P., Harcourt, P., Lloyd, M., Kountouris, A., Smith, B., McGorry, P. D., & Purcell, R. (2019). Preliminary psychometric validation of a brief screening tool for athlete mental health among male elite athletes: The athlete psychological strain Questionnaire. *International Journal of Sport and Exercise Psychology*, 1–16. https://doi.org/10.1080/1612197X.2019.1611900

Roberts, R., Woodman, T., & Sedikides, C. (2018). Pass me the ball: Narcissism in performance settings. *International Review of Sport and Exercise Psychology, 11*, 190–213. https://doi.org/10.1080/1750984X.2017.1290815

Somer, E., Lehrfeld, J., Bigelsen, J., & Jopp, D. S. (2016). Development and validation of the maladaptive daydreaming scale (MDS). *Consciousness and Cognition, 39*, 77–91. https://doi.org/10.1016/j.concog.2015.12.001

Twenge, J. M., & Campbell, W. K. (2009). *The narcissism epidemic: Living in the age of entitlement*. New York: Simon and Schuster.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Generational increases in agentic self-evaluations among American college students, 1966–2009. *Self and Identity, 11*, 409–427. https://doi.org/10.1080/15298868.2011.576820

Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the narcissistic personality inventory. *Journal of Personality, 76*, 875–902. https://doi.org/10.1111/j.1467-6494.2008.00507.x

Vaughn, B. K., & Daniel, S. R. (2012). Conceptualizing validity. In G. Tenenbaum, R. C. Eklund, & A. Kamata (Eds.), *Measurement in sport and exercise psychology* (pp. 33–40). Champaign, IL: Human Kinetics.

Wallace, H. M., Ready, C. B., & Weitenhagen, E. (2009). Narcissism and task persistence. *Self and Identity, 8*, 78–93. https://doi.org/10.1080/15298860802194346

Wilson, S. C., & Barber, T. X. (1982). The fantasy-prone personality: Implications for understanding imagery, hypnosis, and parapsychological phenomena. *PSI Research, 1*, 94–116.

Zhang, S., & Boardley, I. (2022). *The 'selves' in banned performance enhancement: Investigating narcissism and compassion in the context of doping.* Performance Enhancement & Health, Article 100243. https://doi.org/10.1016/j.peh.2022.100243

Zhang, S., Roberts, R., Woodman, T., & Cooke, A. (2020). I am great, but only when I also want to dominate: Maladaptive narcissism moderates the relationship between adaptive narcissism and performance under pressure. *Journal of Sport & Exercise Psychology, 42*, 323–335. https://doi.org/10.1123/JSEP.2019-0204

Zhang, S., Roberts, R., Woodman, T., Pitkethly, A., English, C., & Nightingale, D. (2021). Foresee the glory and train better: Narcissism, goal-setting, and athlete training. *Sport, Exercise, and Performance Psychology, 10*, 381–393. https://doi.org/10.1037/spy0000264