
Citation:

Stanger, N and Whitaker, L and Backhouse, SH (2020) Testing the validity and reliability of the doping willingness in sport scale. Scandinavian Journal of Medicine and Science in Sports. ISSN 1600-0838 DOI: <https://doi.org/10.1111/sms.13683>

Link to Leeds Beckett Repository record:

<https://eprints.leedsbeckett.ac.uk/id/eprint/6721/>

Document Version:

Article (Accepted Version)

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.

Testing the validity and reliability of the doping willingness in sport scale

Nicholas Stanger¹, Lisa Whitaker² & Susan H. Backhouse¹

Leeds Beckett University

Affiliation of authors:

¹Carnegie School of Sport, Fairfax Hall, Headingley Campus, Leeds Beckett University, LS6 3QS, United Kingdom

²UK Coaching, Chelsea Close, Leeds, LS12 4HP, United Kingdom

Manuscript accepted: 25 March, 2020

Address correspondence to: Nicholas Stanger
Leeds Beckett University
Carnegie School of Sport
Fairfax Hall, Headingley Campus
Leeds, LS6 3QS, UK
Tel: (+44) 113 8123525
E-mail: N.Stanger@leedsbeckett.ac.uk

As accepted in *Scandinavian Journal of Medicine & Science in Sports*.

doi: 10.1111/sms.13683

Acknowledgements

This work was undertaken as a supplementary project with funding support from a World Anti-Doping Social Science Research Grant. The funder did not influence the design, conduct, analysis or interpretation of the data in any way. The authors would also like to thank Luke Barnes and Alexandra J. Potts for their help with data collection. The authors have no conflicts of interest to declare.

Abstract

Although research investigating doping in sport is burgeoning, there is still a lack of proxy measures of doping behaviour that have undergone extensive psychometric testing. To address this issue, we modified a previously used measure of doping willingness in sport and tested aspects of validity and reliability across four studies. In Study 1, we provided support for the face and content validity of the items, and then found support for the factor structure of the scale in a sample of athletes ($N = 205$) using confirmatory factor analysis. In Study 2, we collected data from an independent sample of athletes ($N = 236$) to provide further evidence for the factor structure of the scale using confirmatory factor analysis as well as provided evidence for concurrent and discriminant validity. In Study 3, a further independent sample of athletes ($N = 144$) completed the scale and provided support for discriminant and predictive validity of the scale. In Study 4, we collected data from a further independent sample ($N = 74$) to provide support for the test-retest reliability, and stability of items. Lastly, a confirmatory factor analysis was conducted on the samples across Studies 3 and 4, and the composite sample across all four studies which provided further support for the factor structure of the final 8-item scale. Taken together, these findings provide psychometric support for the scale to be used to measure the willingness of athletes to use banned substances to help facilitate future research investigating doping in sport.

Key words: Questionnaire; Scale development; Substance use;

Testing the validity and reliability of the doping willingness in sport scale

The use of banned substances poses a threat to the health and well-being of athletes,¹ and the integrity of sport.² To date, the magnitude of this threat has been difficult to quantify because of the inherent challenges of assessing a prohibited and socially undesirable behaviour. A recent review of the literature reported that the prevalence of performance enhancing substance use among athletes is variable, ranging from 5-31%.³ The relatively wide range of reported prevalence of substance use amongst athletes across studies may be due to the measures applied, the substances examined (banned and permitted) and the potential reluctance of athletes to admit their intention or engagement in this behaviour for fear of perceived repercussions of reporting such use, particularly for banned substances.⁴ Therefore, researchers have employed indirect or proxy measures of banned substance use such as the likelihood, susceptibility and willingness to use banned substance via scenarios.⁵⁻⁸ This is because participants may be more open to report their potential temptation to use banned substances and methods in such situations without directly disclosing their actual engagement.^{4,9} Although such approaches have provided valuable insight into understanding banned substance use in sport, most measures employed in the literature have not comprised of rigorous psychometric support. Thus, the development of a valid and reliable proxy measure of banned substance use is a critical pre-requisite to help facilitate understanding of doping in sport. The aim of this research was to modify and test the psychometric properties of a self-report measure assessing athletes' behavioural willingness to engage in doping.

Doping Willingness

Based on the premise of the prototype willingness model,¹⁰ decision making concerning risky behaviour, such as doping, is considered across two pathways. One pathway which is related to the theory of planned behaviour,¹¹⁻¹² whereby behavioural intentions is seen as a key proxy or determinant of behaviour. The second pathway termed the social

DOPING WILLINGNESS IN SPORT SCALE

reaction path suggests that behavioural willingness is a key determinant of behaviour. This latter path focuses on certain situations that facilitate the risky behaviour in question.

Specifically, behavioural willingness reflects an individuals' openness to opportunity under certain circumstances.⁸ That is, behavioural willingness recognises that although people may find a particular behaviour unfavourable and may have no intention to do it, they may be open to engage in a risk behaviour under certain (risk-conducive) circumstances.^{8,10}

Therefore, doping willingness is one theoretically relevant construct that could be considered to provide a proxy measure of doping in athletes. In the context of doping, we define doping willingness as an openness to take a banned substance in certain risk-conducive situations or contexts even if there was no prior intention to do so.

Previous research has highlighted that there are a range of risk-conducive circumstances that may make athletes consider taking a banned substance.^{8, 13} These include; (a) athletes' perceptions that taking a banned substance could increase their opportunities for selection, gaining a contract or funding; (b) athletes perceiving they are underperforming in competition and/or training; (c) injury; (d) perceptions that everyone else is doping; (e) perceptions that substances can quickly enhance one's physical condition; and, (f) being offered substances from trusted others.^{8, 13} Thus, under such circumstances, athletes may be more vulnerable and willing to dope. Therefore, a doping willingness scale to comprise of content validity (which refers to the extent to which the measure sufficiently covers the construct of interest)^{14,15} would benefit from including items that encompass a range of such risk-conducive situations that may make people vulnerable to taking banned substances.

Previous research measuring doping willingness have asked athletes to report their willingness to dope under situations that may make them more vulnerable and open to engage in doping.⁸ Specifically, such research has revealed that more favourable attitudes about doping and perceived subjective norms (an individual's perception about what significant

DOPING WILLINGNESS IN SPORT SCALE

others think they ought to do) which were favourable toward doping were both positively related with doping willingness.⁸ Although such findings provide valuable insight into correlates of doping willingness, it should be noted that the measure of doping willingness employed underwent limited psychometric testing. Accordingly, the aim of this multi-study research was to address this issue by modifying a measure of doping willingness used in a previous study⁸ to further facilitate the items encompassing a range of risk-conducive situations that were noted in previous research,^{8, 13} and then test various aspects of validity and reliability of the scale.

In sum, previous research examining doping behaviour, or proxies of doping behaviour (e.g., doping susceptibility, doping likelihood) including doping willingness have had shortcomings in terms of limited psychometric support for the scales used. Therefore, we first aimed to modify an existing unidimensional measure of doping willingness,⁸ and test the psychometric properties through testing various aspect of its validity and reliability.^{14,15}

Study 1

The aim of this first study was to modify the items of a pre-existing measure of doping willingness and test for content validity as well as face validity (which refers to whether the items are measuring the construct of interest) which are both typically assessed via expert opinion.^{14,15} Then, we aimed to provide initial evidence for the factorial validity via distributing the items to a sample of athletes to test the factor structure to inform a unidimensional doping willingness scale.

Method

Preliminary scale development. Nine items were developed that were refined or modified from an existing unidimensional scale used in previous research⁸, whereby each item was aimed to reflect at least one of the following risk-conducive or vulnerability factors that may make athletes more willing to engage in doping.^{8, 13} Specifically, the vulnerability

DOPING WILLINGNESS IN SPORT SCALE

factors were; a) perceptions that banned substances may increase opportunities for selection or gaining/ maintaining a professional contract or funding; b) underperforming in competition and/or training; c) injury; d) perceptions that ‘everyone else’ is taking banned substances; e) perceptions that it is necessary to quickly enhance strength or physical condition; f) being offered banned substances from trusted significant others. To test for face validity, the items were evaluated by four sport and exercise psychology academics with expertise in doping, evidenced through undertaking research projects, and publications, relating to doping in sport. Specifically, experts were asked to rate how representative each item was of the respective definition of doping willingness and whether they comprise of at least one of the vulnerability factors on a scale ranging from -3 (not at all representative) to +3 (very representative), and also invited to provide any comments on the items or the scale. Items had a median and mean of 2 or above, and an Aiken’s V score $\geq .83$ (Range = .83 to 1, $p < .05$)¹⁶, therefore all items were retained.

Participants. The initial sample to complete the doping willingness in sport scale (DWiSS) comprised 205 athletes (144 men; 61 women) with an average age of 19.53 (SD = 1.01) years from a range of team ($N = 150$) and individual sports ($N = 55$). The most common team sports athletes participated in were soccer ($N = 71$), rugby ($N = 31$), and cricket ($N = 14$), and the most common individual sport athletes competed in were athletics ($N = 14$), combat sports ($N = 9$) and triathlon ($N = 5$). Participants competed in their respective sport for an average of 8.66 (SD = 4.20) years at international/ national (22%), county/ regional (45%), and club (33%) level. Based on previous Monte Carlo simulation studies,¹⁷ to conduct confirmatory factor analyses for a unidimensional factor comprising 6 and 8 indicators/ items for anticipated factor loadings of at least .50, a samples size of 90 participants was required. Moreover, the required sample size typically is decreased with

DOPING WILLINGNESS IN SPORT SCALE

more indicators in the factor and for higher factor loadings.¹⁷ Therefore, our sample size was perceived adequate to test the factor structure of the scale.

Measures. Participants were asked to answer questions regarding their demographic characteristics (e.g., gender, main sport) and then responded to each of the nine items on the DWiSS following the stem “Would you be willing to use a banned substance if you...” on a 5 point Likert-type scale anchored from 1 (*not at all willing*) to 5 (*extremely willing*). All items are presented in Table 1.

Procedure. Following receipt of ethical approval from the university ethics committee, participants were approached in university classes or at training following previous permission from the university tutor or coach to approach players during sessions. Participants were informed about the purpose of the study, that participation was voluntary, that their responses would be kept confidential and anonymised, and that they had the right to withdraw. Participants were reminded that honesty was vital in their responses. After reading a participant information sheet and signing a consent form, participants were handed the questionnaire to complete. Once completed, participants sealed their completed questionnaire in an envelope provided and returned it directly back to the researcher.

Results and Discussion

Factor analysis. To test the unidimensional factor structure of the DWiSS, we conducted Confirmatory Factor Analysis (CFA) using Stata v14.^{14,18} During initial analysis, estimates of Mardia’s coefficient revealed multivariate non-normality (skewness = 36.03; kurtosis = 184.73). Therefore, we applied Maximum Likelihood with the robust Satorra-Bentler estimation for these analyses. Factor loadings were checked for each item and the model fit was evaluated using the chi-square test, the compare fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) and the standardised root mean squared residual (SRMR). Conventional criteria were used to assist

DOPING WILLINGNESS IN SPORT SCALE

with model assessment whereby chi-square test statistics were (ideally) not significant, CFI and TLI $> .90$, and RMSEA and SRMR $\leq .08$, were considered to reflect adequate model fit, whereas CFI and TLI $> .95$, and RMSEA and SRMR $\leq .06$, were considered to present excellent model fit.¹⁹ As displayed in Table 1, the CFA revealed that all items had a high factor loading (between .73 and .89), and the DWiSS had an excellent fit, SB $\chi^2(27) = 33.07$, $p = .20$; CFI = 0.99, TLI = 0.99, RMSEA = 0.03, SRMR = 0.03. Moreover, the DWiSS comprised of high internal consistency with both Cronbach's²⁰ alpha and McDonald's²¹ Omega (ω) coefficients of .94. The findings from this initial study provided support for the face and content validity, factor structure, and internal consistency of the DWiSS.

Study 2

To extend on the initial psychometric support provided in Study 1, we then aimed to further test the factor structure of the DWiSS in an independent sample, as well as provide support for other aspects of validity for the scale. A prominent issue when developing a measure is to provide support for criterion validity, which refers to the extent scores are correlated with variables (i.e., criterion variables) that one would expect it would be related with.^{14,15} Specifically, this includes concurrent validity which refers to when the criterion variable and the construct are measured at the same time-point.¹⁵ As the use of banned substances can be considered a form of morally questionable behaviour, particularly if undertaken under one's awareness and intentionally, it can be considered an antisocial behaviour which is defined as an act with the intention to harm or disadvantage another.²² Therefore, we examined whether the DWiSS score was associated with antisocial behaviour in sport. Moreover, we also tested whether the DWiSS score was associated with moral disengagement towards general transgressions in sport. Specifically, moral disengagement refers to a set of psychosocial mechanisms or processes that justify the use of antisocial behaviour and negates or eliminates the perpetrator experiencing the typical negative self-

DOPING WILLINGNESS IN SPORT SCALE

evaluative emotional consequences (i.e., guilt).²³ As expected, moral disengagement in sport has been linked with transgressive behaviour in sport,²⁴ been used to justify performance enhancing drug use,²⁵ and been linked with doping susceptibility in previous research.⁵ Therefore, if there were positive relationships for doping willingness with antisocial behaviour and moral disengagement in sport, then this would provide some support for the concurrent validity of the DWiSS.

Another aspect we aimed to investigate is the discriminant validity of the scale. Discriminant validity refers to the extent the scores on a scale are not (too highly) correlated with variables that are conceptually distinct, and there is evidence that the scale scores are independent of such variables.^{14,15,26} Thus, support for discriminant validity for the scale would be supported by being empirically and statistically distinct from other variables that they are correlated with, but yet conceptually distinct (e.g., antisocial behaviour and moral disengagement in sport), as well as weak correlations with variables that one would expect to be weakly linked with doping willingness. As previous research has highlighted that antisocial behaviour and prosocial behaviour tend to be unrelated, or weakly associated,^{27, 28} doping willingness scores being weakly associated with prosocial behaviour would provide some evidence of discriminant validity. In sum, the aim of the next study was to test the factor structure of the scale in an independent sample and provide support for the concurrent and discriminant validity of the DWiSS.

Method

Participants. Participants were 236 team sport players (157 men and 79 women), with an average age of 19.37 ($SD = 1.95$) years. They competed in soccer ($n = 82$), rugby ($n = 51$), field hockey ($n = 43$), netball ($n = 17$), cricket ($n = 17$), basketball ($n = 12$), volleyball ($n = 6$), lacrosse ($n = 4$) and American football ($n = 4$). Participants competed in their respective

DOPING WILLINGNESS IN SPORT SCALE

sports at international/ national (22%), regional/ county (60%) and club (18%) level for an average of 9.38 ($SD = 4.00$) years.

Measures.

Doping willingness. The 9-item DWiSS was used as per Study 1.

Prosocial and antisocial behaviour. The 20-item Prosocial and Antisocial Behaviour in Sport Scale²⁷ was used to measure athletes' prosocial and antisocial behaviour in sport. The PABSS comprises four subscales that measure: antisocial behaviour towards opponents (8 items; e.g., deliberately fouled an opponent), antisocial behaviour towards teammates (5 items; e.g., verbally abused a teammate), prosocial behaviour towards opponents (3 items; e.g., helped an injured opponent) and prosocial behaviour towards teammates (4 items; e.g., congratulated a teammate for good play). Participants were asked how often they engaged in each behaviour whilst competing in their team sport during the past 12 months on a 5-point Likert type scale anchored from 1 (*never*) to 5 (*very often*). Research has supported the convergent, discriminant, and factorial validity of the PABSS, and its internal consistency (alpha range of .73 to .86).^{27, 28}

Moral disengagement in sport. Moral disengagement in sport was measured using the 8-item Moral Disengagement in Sport Scale – Short.²⁹ Participants were asked to rate their level of agreement to statements on a 7-point Likert type scale, anchored from 1 (*strongly disagree*) to 7 (*strongly agree*). An example item is “bending the rules is a way of evening things up”. Each item assesses one of the eight mechanisms of moral disengagement. Psychometric support for the scale has been provided with alpha coefficients ranging from .80 to .85.²⁹

Social Desirability. Social desirability was measured using the 13-item short form Marlowe-Crown social desirability scale.³⁰ Participants were asked a series of questions (e.g., “I’m always willing to admit it when I make a mistake”) and asked to

DOPING WILLINGNESS IN SPORT SCALE

answer true or false to each item. For analysis, a sum score was calculated, based on true being coded as 1 and false coded as 0, but scores were reversed for items where false was the socially desirable response.

Procedure. Following ethical approval from the university ethics committee, participants were approached at training, sporting events or seminar sessions. Participants were provided with an information sheet informing them about the purpose of the study, the voluntary nature of participation, assured responses would be kept confidential, and questionnaires were completed and stored anonymously and reminded about the right to withdraw. After completing a consent form, participants completed the questionnaire comprising of the measures described above. Once completed, participants sealed the questionnaire in an envelope and returned it directly back to a researcher.

Results and Discussion

Preliminary analysis. Only 0.05% of the data was missing for the DWiSS (i.e., one participant missed out one item for the scale). Estimates of Mardia's coefficient revealed multivariate non-normality (skewness = 40.63; kurtosis = 205.23). Also, 0.25% of data was missing for the other measures. As there a small amount of missing data and there was no apparent pattern to these cases so were assumed to be missing at random, missing values were replaced with the mean of the remaining items for the respective subscale in each individual case in subsequent analysis.³¹

CFA. To test for the factor structure of the DWiSS in this sample, we conducted a CFA which revealed high factor loadings for each item (see Table 1) and demonstrated an adequate-to-excellent model fit, $SB\chi^2(27) = 50.90$, $p < .01$, CFI = 0.96, TLI = 0.95, RMSEA = 0.06, SRMR = 0.04. Due to the significant χ^2 statistics, after checking the modification indices and standardized expected parameter change (SEPC) estimates, item 2 had high correlated errors with three other items (items 3, 4 and 8) whereby each

modification index estimate was significant and SEPC scores were all above .20.^{32, 33} After removal of this item, this resulted in a stronger local (i.e., chi-square test statistic) and global fit (e.g., CFI), $SB\chi^2(20) = 27.14, p = .13$, CFI = 0.99, TLI = 0.98, RMSEA = 0.04, SRMR = 0.03, factor loading remained high (see Table 1), and a more parsimonious 8-item scale which still reflected the range of risk-conducive situations noted in previous research.^{8,13}

Construct validity. To test for concurrent and discriminant validity of the DWiSS, we ran correlations to examine relationships between the DWiSS score with prosocial behaviour, antisocial behaviour, and moral disengagement. When testing for construct validity, and particularly discriminant validity it is favourable to conduct analyses using structuring equation modelling (SEM) as this also enables for consideration of measurement error.^{26, 34} Therefore, analysis was based on correlations between variables in SEM using Maximum Likelihood with the robust Satorra-Bentler estimator, their 95% confidence intervals (CIs) as well as comparing the average variance extracted (AVE; i.e., average amount of variance explained by the questionnaire items) for each construct with the shared variance between constructs (i.e., square of correlations). Concurrent validity for the DWiSS would be supported if there are significant positive correlations (i.e., confidence intervals do not cross zero) with both antisocial behaviours and moral disengagement. Discriminant validity will be supported if the AVE for each construct is greater than the shared variance with other constructs,^{26, 35} and if the 95% confidence intervals do not include -1 or 1.³⁶

As shown in Table 2, the 8-item DWiSS comprised of high internal consistency. Also, the DWiSS score was positively associated with moral disengagement in sport and antisocial behaviour towards both teammates and opponents, thereby providing support for the concurrent validity of the DWiSS. The DWiSS score was also negatively linked with social desirability with a small effect size ($r = -.19, p < .01$). Therefore, we also conducted (partial) correlations controlling for social desirability in the models. Doping willingness was still

DOPING WILLINGNESS IN SPORT SCALE

positively related with antisocial behaviour toward opponents and moral disengagement, though the link with antisocial behaviour toward teammates was weakened when controlling for social desirability. However, these results broadly provide further support for the concurrent validity of the scale.

Discriminant validity was also supported whereby AVEs for each construct were greater than the shared variance between doping willingness and other constructs, and no CIs comprised -1 or 1, including when controlling for social desirability in the models. The DWiSS score was also negligibly (or weakly) related with prosocial behaviour towards teammates and opponents both when, and when not, controlling for social desirability, thereby providing some further support for the discriminant validity of the scale.

Study 3

To extend on the support provided for concurrent and discriminant validity in Study 2, we aimed to further test the discriminant validity as well as the predictive validity of the DWiSS with other criterion variables. Predictive validity is a form of criterion validity which is supported when relationships are found with criterion variables that one would expect to be related with, when the criterion variable is measured at a later time.¹⁵ To test for predictive validity and discriminant validity, we examined relationships with criterion variables using a picture viewing task that participants completed at least three days after completing the DWiSS. Specifically, the picture viewing task is an established paradigm used in range of contexts to measure individuals' emotional reactivity to pictures that differ in emotional or thematic content.^{37, 38} This method is based on the biphasic theory,³⁹ which defines emotion around two motivational systems namely; valence (pleasant or appetitive vs unpleasant or aversive) and arousal (i.e., emotional activation), where processing picture content is assumed to prime a particular motivational system.³⁸ This paradigm has also been employed to assess emotional reactivity in sport in previous research via asking participants to report

DOPING WILLINGNESS IN SPORT SCALE

their valence and arousal toward pictures, which included those that depicted aggressive acts in sport.^{24, 40, 41} Such research has revealed that athletes reporting higher antisocial behaviour were associated with less unpleasant emotional reactions to content depicting such behaviour or players who were badly hurt.⁴¹ In this research, we developed a pool of pictures that depicted doping in sport to assess athletes' emotional reactivity towards doping. Based on previous research in relation to other forms of content including morally relevant behaviour,^{37, 41} it would be expected that people who are less willing to engage in doping would perceive doping represented in the pictures as being more unpleasant, due to priming an aversive state in relation to such conduct.

Moreover, we also aimed to examine whether moral judgment towards pictures of doping was linked with reported doping willingness measured at different time-points. Moral judgment has been referred to the cognitive evaluation about the morality of the conduct (e.g., whether it is right or wrong), and various theorists have argued that moral judgment is a key factor in moral action^{23, 42} whereby people are less likely to commit behaviours they perceive as wrong.²³ In previous research in sport, athletes who perceived antisocial behaviours as being more morally wrong were associated with engagement in less frequent antisocial behaviour.⁴³ Indeed, researchers have also argued that training moral judgment may be an important element in doping prevention.⁴⁴ Therefore, it would be expected that people who report being more willing to dope would also perceive that doping was less wrong. Therefore, we aimed to test relationships between DWiSS scores with emotional reactions (in terms of valence and arousal ratings) and moral judgment toward images depicting doping.

To test for discriminant validity, in addition to testing the independence between the DWiSS and other measures (i.e., valence, arousal and moral judgment toward doping), we also examined participants' emotional reactions and moral judgments towards pictures of

DOPING WILLINGNESS IN SPORT SCALE

other behaviours in sport (e.g., injurious acts) that have been used in previous research.^{24, 40}

Specifically, it was expected that the relationships for the score on the DWiSS with emotional reactions and moral judgments towards doping pictures will be stronger than those towards other forms of behaviour in sport (i.e., aggressive behaviours).

Method

Participants. Participants were 144 team sport players (99 men and 45 women), with an average age of 19.61 ($SD = 2.84$) years. They competed in soccer ($n = 48$), rugby ($n = 27$), field hockey ($n = 12$), netball ($n = 12$), tchoukball ($n = 9$), cricket ($n = 8$), basketball ($n = 6$), handball ($n = 6$), volleyball ($n = 5$), lacrosse ($n = 3$), American football ($n = 1$), and Gaelic football ($n = 1$). Participants competed in their respective sports at international/ national (14%), regional/ county (47%) and club (39%) levels for an average of 8.41 ($SD = 4.76$) years.

Dispositional Measures.

Doping willingness. The DWiSS was completed as per Studies 1 and 2 (8-items), and social desirability as per Study 2.

Picture Viewing Measures.

Emotional reactions to pictures. Each picture was rated on valence and arousal using the Self-Assessment Manikin.⁴⁵ Valence was rated on a 9-point scale with anchors from 1 (*very unpleasant*) to 9 (*very pleasant*), and arousal was rated on a 9-point scale with anchors from 1 (*very calming/ low arousal*) to 9 (*very activating/ high arousal*). This approach is similar to previous research investigating emotional reactions to sport stimuli.^{24, 40, 41}

Moral judgement to pictures. Moral judgments to the pictures were assessed by asking participants the extent to which they perceived the behaviour in the picture (or represented by the picture) to be morally wrong on a 9-point scale with anchors from 1 (*not*

1 *at all wrong*) to 9 (*very wrong*). Similar approaches have been adopted to assess moral
2 judgment in previous research.⁴³

3 **Picture Viewing Task.** The picture viewing task involved participants observing 24
4 pictures of which eight depicted doping. The doping pictures displayed athletes either taking
5 pills, injecting a substance via syringe, or comprised concise newspaper headlines reporting
6 incidences of athletes doping. Prior to testing, the pictures were piloted on a sample of 15
7 active researchers in sport and exercise psychology or a related discipline, who rated a larger
8 set of pictures in terms of whether the picture was clear (from 1 = *not at all*; to 9 =
9 *extremely*), whether the behaviour in the pictures could clearly represent doping in sport
10 (from 1 = *not at all*; to 9 = *extremely*), as well as their emotional involvement on valence and
11 arousal to each picture using the Self-Assessment Manikin. This was conducted to test for the
12 content and face validity of the pictures. The eight pictures that scored highest for content
13 validity (i.e., picture was clear; could depict doping) and were rated as most unpleasant out of
14 17 pictures from this piloting were included in the study. In addition to the doping-related
15 pictures, eight pictures depicted severe aggressive behaviours (e.g., bad tackles), four pictures
16 depicted mild fouls (shirt pulling), and four pictures were neutral or pleasant stimuli (e.g.,
17 players celebrating). The mild fouls, and neutral or pleasant pictures were included as fillers
18 to reduce the potential for familiarisation, habituation or reporting bias. These non-doping
19 related pictures were similar to those used in previous research.^{24, 40, 41}

20 The pictures were presented in a randomised order that was fixed for all participants.
21 To help control for some order effects in terms of the presentation of the pictures, the order
22 presentation of pictures was reversed for half the participants. Thus, there were two picture
23 orders, half the participants observed the pictures in the original fixed order whereas the other
24 half observed the pictures in the fixed reverse order. Each picture was presented for six
25 seconds with an inter-picture interval of 30 seconds. During the inter-picture interval

DOPING WILLINGNESS IN SPORT SCALE

participants were asked to rate their emotional involvement (for valence and arousal) and moral judgements (in terms of the extent the behaviour was morally wrong) toward the pictures. This inter-picture interval was found appropriate for participants to make their responses to each of the three questions, and then fixate on a cross on the screen ready for the presentation of the next picture.

Procedure. Following ethical approval, participants were recruited and after reading the participant information sheet and completing a consent form to confirm their agreement to take part, participants provided demographic information and completed the DWiSS and the measure for social desirability. Then, at least three days (but no more than 30 days) later, participants attended a computer-based session. At the start of this session, participants were provided with an information sheet and a second consent form to confirm that they still agreed to participate and understood what participation would involve. Participants then completed a second demographics sheet, so responses could be matched with the previous measures to also help maintain anonymity.

Participants were then instructed that they will complete a picture viewing task, which included pictures comprising of different behaviours in sport, including pictures of athletes intentionally taking a banned substance or headlines that represented athletes taking banned substances, players celebrating or players engaging in bad tackles aiming to harm their opponent. Participants were asked to sit back in their chair whilst viewing each picture and to view the picture for the full duration that it was presented before making their responses following each picture. Participants were also given a one-minute break half-way through the presentation of pictures (i.e., after 12 pictures) to help reduce potential fatigue.

Results and Discussion

Preliminary analysis. Similar with Studies 1 and 2, estimates of Mardia's coefficient revealed multivariate non-normality for the DWiSS (skewness = 28.11; kurtosis = 130.95). We

DOPING WILLINGNESS IN SPORT SCALE

conducted preliminary analyses to test the factor structure of the valence ratings, arousal ratings, and moral judgments toward the respective doping and aggressive behaviour pictures. Specifically, we conducted a series of CFAs and removed any pictures whereby any rating (valence, arousal or moral judgment) had low factor loadings. Two of the eight pictures for both doping and aggressive content were removed. The subsequent model fits are presented in Table 3 whereby each measure comprised of an adequate-to-excellent or excellent model fit.

Correlational analysis. Descriptive statistics, internal consistencies for each variable, and correlations with the 8-item DWiSS are presented in Table 4. Identical to Study 2, analyses were based on correlations between variables in SEM using Maximum Likelihood with the robust Satorra-Bentler estimator, their 95% confidence intervals (CIs) as well as comparing the AVE for each construct with the shared variance between constructs.^{26, 35, 36}

Correlational analyses revealed that the DWiSS score was positively associated with valence ratings (i.e., higher doping willingness was associated with less unpleasant reactions toward doping), negatively associated with perceptions that doping was morally wrong (i.e., higher doping willingness was associated with judgment that doping was less wrong), providing support for the predictive validity of the scale. The DWiSS score was not linked with arousal ratings toward doping. This finding is aligned to previous research showing a lack of relationship between morally relevant variables and the arousal dimension of emotion toward pictorial stimuli.⁴¹

The DWiSS scores were not associated with emotional reactions (neither valence nor arousal ratings) or moral judgements towards pictures of other (i.e., aggressive) acts in sport. As such, these weaker and more negligible relationships between DWiSS scores with valence ratings and moral judgment toward pictures depicting aggressive content (i.e., a distinct behaviour from doping) than toward pictures of doping content provides further support for the discriminant validity of the scale. Similar to Study 2, doping willingness was negatively

DOPING WILLINGNESS IN SPORT SCALE

linked with social desirability ($r = -.40, p < .001$). When controlling for social desirability and number of days between completion of the doping willingness scale and the picture viewing task by including these as covariates in the respective models, the strength and direction of the links between doping willingness scores with emotional reactions (valence and arousal ratings) and moral judgment toward the pictures remained similar.

Discriminant validity was also supported whereby AVEs for each construct were greater than the shared variance between the DWiSS scores and other constructs, and no CIs comprised -1 or 1, including when controlling for social desirability in the models.

Study 4

The results from Studies 1 to 3 supported the face, factorial, and external validity of the scale. Another psychometric property that is important to consider when developing a measure is testing for test-retest reliability and stability of the items. This is typically estimated by examining how scores are correlated and reproducible when administering the same measure at two-time points, and thereby testing whether scores are resistant to change on repeated administrations, when no change is expected.^{46,47} Therefore, we tested for test-retest reliability and stability over a two-week interval. We also aimed to provide further examination of the factorial validity of the scale on an independent sample by conducting a CFA for the samples across Studies 3 and 4 as well as undertake preliminary gender invariance testing based on the composite sample across all 4 studies to see if the DWiSS performed equally for men and women.

Method

Participants. Participants were seventy-four (52 men; 22 women) athletes, with an average age of 19.01 ($SD = 1.75$) years who competed in predominantly team sport (55 team sport athletes; 19 individual sport athletes). The most common team sports were soccer ($N = 26$), rugby ($N = 17$) and lacrosse ($N = 3$), and the most common individual based sports were

DOPING WILLINGNESS IN SPORT SCALE

athletics ($N = 5$), weightlifting ($N = 3$), and martial arts ($N = 3$). Participants competed in their respective sports at international/ national (16%), regional/ county (50%) and club level (34%) levels for an average of 7.85 ($SD = 4.36$) years.

Procedure. Following ethical approval, participants were recruited via teaching sessions at university. After reading an information sheet and completing a consent form, participants completed the DWiSS. Two-weeks later, participants were asked to complete a second consent form to confirm their continued agreement to take part and a demographics sheet to match responses from the initial questionnaire completion. In total, 50 athletes (36 males, 14 females; 34 team sport players, 16 individual sport athletes) completed measures on both occasions. The difference in response rate between the two-points was due to non-attendance during the second session rather than choosing not to participate.

Results and Discussion

Test-retest. Test-retest reliability was examined on the subsample using intraclass correlations between the mean scores across the two-time points (i.e., 2-week interval). Results revealed a strong intraclass correlation of .92 ($p < .001$; 95% CI; 0.86 to 0.95) suggesting excellent test-retest reliability for the DWiSS.⁴⁸ We also tested for the stability of the individual items as per recommendations from Nevill and colleagues⁴⁷ (see Table 5). For each item, we have reported the test-retest differences and the minimum and maximum difference scores, the proportion of agreement which refers to the percentage of proportion of test-retest differences within ± 1 . It is suggested for 5-point scales that score stability is supported when the proportion of agreement scores are 90% or above.⁴⁷ All items had at least 90% proportion of agreement apart from item 4 with a proportion of agreement score of 86%. We also conducted a Sign-Rank test to determine the degree of potential bias of scores increasing or decreasing from test to retest. The results revealed no significant bias between scores for most items, however the score for item 3 tended to increase over the two-week

interval ($p = .04$). Thus, besides some slight differences in responses over the 2-week interval, overall these results provide general support for the test-retest reliability and stability of the DWiSS items.

Confirmatory factor analysis. To provide further evidence for the factor structure of the scale, we conducted a CFA on the 218 participants across Study 3 and Study 4 (including responses only from the 74 participants at time point 1), which revealed an excellent model fit, $SB\chi^2(20) = 24.03$, $p = .24$, CFI = 1.00, TLI = 0.99, RMSEA = 0.03, SRMR = 0.02. As presented in Table 1, factor loadings ranged from .65 to .88, and both Cronbach's ²⁰ alpha ($\alpha = .93$) and McDonald's ²¹ Omega ($\omega = .94$) indicated high internal consistency.

We also conducted a final CFA on the composite sample of 659 participants across studies 1 to 4, which revealed overall an excellent model fit, $SB\chi^2(20) = 56.48$, $p = < .001$, CFI = 0.98, TLI = 0.97, RMSEA = 0.05, SRMR = 0.02. Although the chi-square statistics were significant, it is often advocated that this statistic can be over-sensitive and over-inflated with large sample sizes.⁴⁹ In addition, the Cronbach alpha (α) and McDonald's omega (ω) of .93 across the composite sample provided further support for the internal consistency of the DWiSS.

Gender invariance. As our sample size comprised of a sufficient sample size of men ($N = 452$) and women ($N = 207$) across the four studies, we conducted some initial measurement invariance testing via multi-group SEM which are presented in Table 6. Please note that due to Stata software not producing robust estimator model fit indices for metric (constraining factor loadings) and scalar (also constraining measurement intercepts) invariance testing apart from for SRMR, we can only provide SRMR results for these analyses. Therefore, these results do need to be considered with this cautionary note in mind.

DOPING WILLINGNESS IN SPORT SCALE

We first tested the factor structure separately for each gender, which confirmed an excellent model fit for men and women. As noted above the chi-square statistic can be sensitive to over-inflation in larger sample sizes,⁴⁹ and thereby the significant chi-square statistic in men may be due to the larger number of men in the sample. Then, we tested for configural invariance whereby no constraints were placed on parameters, which indicated this was supported via an excellent model fit. We then conducted metric invariance which resulted in an increase in SRMR of .018 from the configural invariance model. As our sample comprised unequal size groups (452 men and 207 women), but an adequate sample size, we considered both of these aspects in the consideration of evaluating invariance testing. Based on Chen⁵⁰, the recommended criteria for SRMR to support metric invariance is an increase of $<.025$ for unequal size groups and $<.03$ for adequate sample sizes. Therefore, based on these criteria, the estimates for SRMR provide some promising initial support for metric invariance. In terms of scalar invariance, support has been suggested to be provided when an increase in SRMR is $<.005$ for unequal size groups or $<.010$ for adequate sample size.⁵⁰ Invariance testing in our sample revealed a SRMR increase of .004, and thereby provided some promising preliminary support for scalar invariance.

General Discussion

Research on doping in sport has burgeoned over recent years,⁵¹ yet there are limited proxy measures of doping that comprise of suitable psychometric testing. Such measures are important to help facilitate understanding about such a complex and multifaceted behaviour as doping in sport.⁵² Accordingly, the present research refined a measure of willingness to use banned substances in sport and provided evidence for various aspects of validity and reliability.

Based on the premises of the prototype willingness model,¹⁰ and recent research,^{8, 13} we modified and tested a measure that assesses athletes' behavioural willingness to engage in doping via banned substance use that considers six key risk-conducive situations. A series of

DOPING WILLINGNESS IN SPORT SCALE

confirmatory factor analyses across independent samples provided evidence for the unidimensional factor structure of the DWiSS that comprised an excellent fit (Studies 1, 2 and 4), with all items possessing high factor loadings. We also provided preliminary evidence for measurement invariance to offer some initial support that the scale may be equally suitable to measure doping willingness in both men and women.

Concurrent validity of the DWiSS score was evidenced by the positive associations noted with antisocial behaviour towards teammates and opponents (Study 2) which is expected given that these behaviours (particularly if doping is undertaken intentionally) could be considered transgressive-related behaviour. Moreover, the DWiSS score was also positively linked with moral disengagement in sport measured via the moral disengagement in sport scale-short²⁹, which is aligned to previous research showing a positive relationship between this moral disengagement scale and doping susceptibility.⁵

Predictive validity for the scale was also evidenced via DWiSS scores being positively linked with valence ratings (i.e., less unpleasant reactions), and associated with perceiving doping as less morally wrong in relation to viewing pictures of doping in sport at a later time point (Study 3). Based on the social cognitive theory of moral thought and action,²³ people are less likely to engage in behaviours that they anticipate experiencing negative emotional reactions towards, or behaviours that they perceive, are morally wrong. Therefore, these findings are aligned to expectations whereby those who possess lower doping willingness are associated with more unpleasant reactions towards doping and judgments that such behaviours are more morally wrong.

Discriminant validity was also evidenced in this research whereby the DWiSS was shown to be independent of a range of constructs across Studies 2 and 3, which demonstrated that the AVEs were greater than any shared variance with all the other constructs measured in this research. Additionally, DWiSS scores were not significantly linked with prosocial

behaviours, but was significant with antisocial behaviour, particularly toward opponents (Study 2). These findings are aligned to previous research indicating the prosocial and antisocial behaviour are independent behaviours and weakly related or unrelated in previous research,^{27, 28} and thereby offers some additional support for discriminant validity. To further support the discriminant validity of the DWiSS, the scores for doping willingness were also more strongly linked with valence ratings and moral judgments towards doping pictures than towards pictures reflecting other behaviours (i.e., aggressive acts) in Study 3.

We also provided evidence for the internal consistency and test-retest reliability of scores obtained from the DWiSS. The internal consistency was high across all subsamples in this research whereby alpha (α) and omega (ω) coefficients were both above .90. Also, test-retest reliability scores via intraclass correlations was high across the two-week interval and substantially above the recommended .70.⁴⁷ Moreover, on the whole, the majority of items across the scale were stable across the two-week interval. Therefore, these findings indicate the scores obtained using the scale are relatively stable over time and are resistant to situational factors that can impart error on scale responses.

Limitations and Future Research

The present research provided psychometric support for the DWiSS across four studies. Although we tested various aspects of validity, one area that warrants examination is testing the convergent validity of the DWiSS, which refers to the extent that scores are associated with measures of the same or similar construct.¹⁴ At the moment, there are limited measures that have undergone rigorous psychometric testing that may be suitable to assess for convergent validity, though researchers could consider using measures applied to assess doping likelihood in previous research as an initial step to test this.^{6, 9}

It should be acknowledged that we did not include testing for aspects of validity with other scales measuring factors within the prototype willingness model¹⁰ or theory of planned

DOPING WILLINGNESS IN SPORT SCALE

behaviour^{11, 12} which could have provided further evidence for construct validity. This was mainly due to the lack of validated measures for subjective norms, doping intention and behaviour which have under-gone rigorous psychometric development. That said, testing relationships with doping attitudes via the Performance Enhancement Attitudes Scale (PEAS),⁵³ would provide further support for the concurrent validity of the scale. Therefore, as scale development is an on-going process future research may wish to address this issue. Though researchers may wish to consider adopting an abbreviated version of the PEAS scale due to some issues noted regarding the factor structure of the original 17-item scale.⁵⁴

We provided some evidence for the predictive validity of scores by investigating relationships with variables taken at a later time point (Study 3). However, this time point was not particularly long and varied between participants due to difficulties with keeping the time-frame consistent with participants attending a controlled laboratory-based session (i.e., between three and 30 days). However, we did control for number of days between DWiSS completion and the picture viewing task in analyses to help address this. Moreover, we employed the picture viewing task to examine emotional reactions and moral judgments toward doping (and aggressive behaviour) in Study 3. Although this is based on an established paradigm and we provided support for some aspects of validity of this task, the pictures employed particularly for doping had not undergone previous extensive psychometric testing which may not be ideal for criterion variables in psychometric testing of new measures. We adopted this approach to offer some variation in the methods employed to measure criterion variables¹⁴ (i.e., beyond a typical self-report questionnaire) and due to rather limited validated doping-specific measures available particularly at study conception. That said, future research could consider investigating the relationships between scores from the DWiSS with variables over a longer (and if possible consistent) time period using measures that have also received previous extensive psychometric support.

Finally, we provided support for the configural invariance of the DWiSS for gender, and some preliminary support was provided for metric and scalar invariance. However, we relied solely on SRMR results for metric and scalar variance due to the statistical software employed which does not currently compute robust estimates for other model fit indices in multi-group analyses with constraints. It is recommended to consider a range of fit indices, particularly CFI, to make more informed judgments when evaluating measurement invariance.^{50, 55} Therefore, the results do need to be carefully considered with this limitation in mind and researchers may wish to address this in future evaluations of the scale. Further testing of the DWiSS would benefit from further invariance testing for gender with ideally equal size groups, and also comprise of suitable samples to test for sport type (team vs. individual) and cross-cultural invariance.

Perspective

Banned substance use can have adverse consequences on the health and well-being of athletes,¹ as well as question the integrity of those involved.² To help facilitate research investigating factors surrounding banned substance use, we modified a psychometric scale of doping willingness in sport and provided support for its reliability and validity. Specifically, evidence has been provided for the factor structure as well as the face, content, concurrent, discriminant and predictive validity of the DWiSS, and for its internal consistency, test-retest reliability and stability of its items. The DWiSS provides the first validated measure for athletes' behavioural willingness to engage in doping through banned substance use and provides a proxy measure of doping in sport that has gone through psychometric testing. Though measurement development is an ongoing process, we look forward to seeing this scale used, and continually evaluated, in future research to facilitate understanding about the complex nature of doping in sport.

Endnotes

- (1) When testing for the face and content validity in Study 1, one discrepant rating from an expert for item 9 was removed from analysis due to it deviating so notably from the other experts.⁵⁶ Specifically, as this single rating was *more than 2* response options lower on the scale than all the other experts, this discrepant rating was removed similar to previous research.⁵⁷ Therefore, we decided to retain this item based on majority opinion, and because the Aiken *V* score using the ratings from the remaining 3 experts was significant (Aiken's $V = 0.89, p < .05$).
- (2) To further check the clarity of the items for content validity, a sample of 27 athletes (14 men; 13 women) with an average age of 23.52 ($SD = 6.30$) years who competed in their main sport for an average of 11.70 ($SD = 5.19$) years, completed the scale and also asked to rate the clarity of wording for each item by indicating whether each item was "*easy to understand*", "*moderately difficult to understand*" or "*difficult to understand*". A comments box was included to invite any qualitative comments about the wording of the items. All participants rated items 1 to 8 as easy to understand so were retained, and 25 of the 27 participants rated item 9 as easy to understand. As still the vast majority (93%) rated item 9 as easy to understand, we decided to also retain this item.
- (3) In Study 1, a supplementary exploratory factor analysis (EFA) was conducted on the nine items of the DWiSS to also check that the scale was unidimensional using principal axis extraction and direct oblimin rotation. The EFA supported the appropriateness of the matrices (Bartlett test of sphericity $p < .001$; Kaiser-Meyer Olkin = .94)⁵⁸ and revealed a unidimensional solution with only one factor with an eigenvalue of above 1 (eigenvalue = 6.13), with all items having a factor loading of at least .74.
- (4) Due to the removal of item 2 in Study 2, we also cross-checked the modification indices and SEPC estimates for Study 1 whereby item 2 comprised significant modification indices and high SEPC ($> .20$) with a different item from that in Study 2 (item 3). When removing item 2 from Study 1, this resulted in an improved model fit, $SB\chi^2(20) = 22.65, p = .31, CFI = 1.00, TLI = .99, RMSEA = 0.03, SRMR = 0.03$.
- (5) For consideration of sample sizes in Studies 2 and 3, based on Monte Carlo simulation in a study by Wolf et al.,¹⁷ a model comprising two factors (with a moderate correlation of .30) in structural equation modelling where both factors have at least 6 indicators, with factor loadings of .50 and .65, adequate sample sizes were 190 and 120 participants, respectively. Considering the factor loadings for the DWiSS in Study 1 and 2, and development of the PABSS (with scales comprising 3, 4, and 5 items for prosocial opponent, prosocial teammate and antisocial teammate respectively) were mainly above .65, based on Wolf et al.¹⁷, 200 participants was found to be an adequate sample size for two factors.
- (6) The samples in Study 2 and Study 3 also completed other measures including for doping moral disengagement⁵⁹ and moral identity⁶⁰, which were completed to address another study purpose reported in another manuscript. Data for social desirability from these samples used in this paper was also used in this other research paper.

References

1. Pope, H. G., Wood, R. I., Rogol, A., Nyberg, F., Bowers, L., & Bhasin, S. (2013). Adverse health consequences of performance-enhancing drugs: An Endocrine Society scientific statement. *Endocrine Reviews*, 35, 341–375. doi: 10.1210/er.20131058.
2. Donovan, R. J., Egger, G., Kapernick, V., & Mendoza, J. (2002). A conceptual framework for achieving performance enhancing drug compliance in sport. *Sports Medicine*, 32, 269-284. doi:10.2165/00007256-200232040-00005
3. Momaya, A., Fawal, M., & Estes, R. (2015). Performance-enhancing substances in sports: A review of the literature. *Sports Medicine*, 45, 517–531. doi: 10.1007/s40279-015-0308-9
4. Petroczi, A. (2016). Indirect measures in doping behavior research. In V. Barkoukis, L. Lazaras, & H. Tsorbatzoudis (Eds.). *The psychology of doping in sport* (pp. 93-110). New York: Routledge.
5. Hodge, K., Hargreaves, E. A., Gerrard, D., & Lonsdale, C. (2013). Psychological mechanisms underlying doping attitudes in sport: Motivation and moral disengagement. *Journal of Sport & Exercise Psychology*, 35, 419-432.
6. Kavussanu, M., & Ring, C. (2017). Moral identity predicts doping likelihood via moral disengagement and anticipated guilt. *Journal of Sport & Exercise Psychology*, 39, 293-301. doi: 10.1123/jsep.2016-0333
7. Ring, C., & Kavussanu, M. (2017). The role of self-regulatory efficacy, moral disengagement and guilt on doping likelihood: A social cognitive theory perspective. *Journal of Sports Sciences*, 36, 578-584.
8. Whitaker, L., Long, J., Petróczi, A., & Backhouse, S. (2014). Using prototype willingness model to predict doping in sport. *Scandinavian Journal of Medicine & Science in Sports*, 24, e398-e405.
9. Ring, C., Kavussanu, M., Simms, M., & Masanov, J. (2018). Effects of situational costs and benefits on projected doping likelihood. *Psychology of Sport & Exercise*, 34, 88-94. Doi:10.1016/j.psychsport.2017.09.012
10. Gibbons, F. X., Gerrard, M., & Lane, D. J. (2003). A social reaction model of adolescent health risk. In J. M. Suls, K. A. Wallaston (Ed.), *Social psychological foundations of health and illness*. Oxford: Blackwell.
11. Ajzen, I. (1985). From intentions to actions: A theory of planned behaviour: In J. Kuhl, &

- J. Beckman (Eds.), *Action control: From cognitions to behaviour* (pp.11-39). New York: Springer.
12. Ajzen, I., & Fishbein, M. (1980). *Understanding attitude and predicting social behaviour*. New Jersey: Prentice-Hall.
13. Whitaker, L., Backhouse, S., & Long, J. (2017). Doping vulnerabilities, rationalisations and contestations: The lived experience of national level athletes. *Performance Enhancement & Health*, 5, 134-141. doi:10.1016/j.peh/2017.06.001
14. Kline, R.B. (2016). *Principles and practice of structural equation modeling*. (4th Ed.). New York: The Guilford Press.
15. Price, P.C., Jhangiani, R. S., & Chiang, I-C, A. (2015). *Research methods in psychology* (2nd ed.). Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
16. Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142. doi:10.1177/0013164485451012
17. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution property. *Educational & Psychological Measurement*, 76, 913-934. doi:10.1177/0013164413495237
18. StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.
19. Marsh, H. W., Hau, K. T., & Wen, Z. L. (2004). In search of golden rules: comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers of overgeneralising Hu & Bentler (1999) findings. *Structural Equation Modeling*, 11, 320-341. doi:10.1207/s15328007sem1103_2
20. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334
21. McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum
22. Sage, L., Kavussanu, M., & Duda, J. L. (2006). Goal orientations and moral identity as predictors of prosocial and antisocial functioning in male association football players. *Journal of Sports Sciences*, 24, 455–466.

23. Bandura, A. (1991). *Social cognitive theory of moral thought and action*. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 45-103). Hillsdale, NJ: Erlbaum.
24. Stanger, N., Kavussanu, M., Boardley, I. D., & Ring, C. (2013). The influence of moral disengagement and negative emotion on antisocial sport behavior. *Sport, Exercise & Performance Psychology*, 2, 117-129. doi:10.1037/a0030585
25. Boardley, I. D., Grix, J., & Dewar, A. J. (2014). Moral disengagement and associated processes in performance-enhancing drug use: a national qualitative investigation. *Journal of Sports Sciences*, 32, 836-844. doi:10.1080/02640414.2013.862842
26. Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research*, 63, 324-327. doi:10.1016/j.jbusres.2009.05.003
27. Kavussanu, M & Boardley, I, D (2009). The prosocial and antisocial behavior in sport scale. *Journal of Sport & Exercise Psychology*, 31, 97-117. doi:10.1123/jsep.31.1.97
28. Kavussanu, M., Stanger, N., & Boardley, I. D. (2013). The prosocial and antisocial behaviour in sport scale: Further evidence for construct validity and reliability. *Journal of Sports Sciences*, 31, 1208-221. doi:10.1080/02640414.2013.775473
29. Boardley, I. D., & Kavussanu, M. (2008). The moral disengagement in sport scale - short. *Journal of Sports Sciences*, 26, 1507-1517. doi:10.1080/02640410802315054
30. Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-crowne social desirability scale. *Journal of Clinical Psychology*, 38, 119-125. doi:10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118.3.0.CO;2-I
31. Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka, & W.F. Velicer (Eds.), *Research methods in psychology* (pp. 87-112). New York, NY: Wiley.
32. Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications. *Structural Equation Modeling*, 16, 561-582. doi: 10.1080/10705510903203433
33. Whitaker, T. A. (2012). Using the modification index and standardized expected parameter change for model specification. *The Journal of Experimental Education*, 80, 26-

44. doi:10.1080/00220973.2010.531299
34. Bollen, K.A. *Structural equations with latent variables*. New York, NY: John Wiley; 1989.
35. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50. doi: 10.2307/3151312.
36. Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36, 421-458. doi:10.2307/2393203
37. Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1, 276–298. doi:10.1037//1528-3542.1.3.276
38. Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological Psychology*, 84, 437-450. doi:10.1016/j.biopsycho.2009.10.007
39. Lang, P. J. (1985). The cognitive psychophysiology of emotion: fear and anxiety. In A. H. Tuma, & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp.131-170). Hillsdale, NJ: Erlbaum.
40. Stanger, N., Kavussanu, M., & Ring, C. (2012). Put yourself in their boots: Effects of empathy on emotion and aggression. *Journal of Sport & Exercise Psychology*, 34, 208-222. doi:10.1123/jsep.34.2.208
41. Stanger, N., Kavussanu, M., Willoughby, A., & Ring, C. (2012). Psychophysiological responses to sport-specific affective pictures: A study of morality and emotion in athletes. *Psychology of Sport & Exercise*, 13, 840-848. doi:10.1016/j.psychosport.2012.06.004
42. Rest, J.R. (1984). The major components of morality. In W. Kurtines & J. Gewirtz (Eds.) *Morality, moral behavior, and moral development* (pp. 556-629). New York: Wiley.
43. Kavussanu, M., Stanger, N., & Ring, C. (2015). The effects of moral identity on moral emotion and antisocial behavior in sport. *Sport, Exercise, & Performance Psychology*, 4, 268-279.
44. Melzer, M., Elbe, A-M., & Brand, R. (2010). Moral and ethical decision-making: A chance for doping prevention in sports? *Nordic Journal of Applied Ethics*, 4, 69-85. doi:10.5324/eip.v4i1.1741

45. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy & Experimental Psychiatry*, 25, 49e59. doi:10.1016/0005-7916(94)90063-9.
46. Lane, A. M., Nevill, A. M., Bowes, N., & Fox, K. R. (2005). Investigating indices of stability using the task and ego orientation questionnaire. *Research Quarterly for Exercise and Sport*, 76, 339–346.
47. Nevill, A., Lane, A. M., Kilgour, L., Bowes, N., & Whyte, G. (2001). Stability of psychometric questionnaires. *Journal of Sports Sciences*, 19, 273–278.
48. Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaire. *Journal of Clinical Epidemiology*, 60, 34–42.
49. Lewis, T. F. (2017). Evidence regarding the internal structure: Confirmatory factor analysis. *Measurement and evaluation in counselling and development*, 50, 239-247. doi:10.1080/07481756.2017.1336929
50. Chen F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. doi: 10.1080/10705510701301834
51. Backhouse, S.H., Whitaker, L., Patterson, L., Erickson, K., McKenna, J. (2016). Social psychology of doping in sport: a mixed-studies narrative synthesis. Retrieved from http://www.wada-ama.org/sites/default/files/resources/files/literature_review_update_-_final_2016.pdf
52. Backhouse, S.H., Griffiths, C. & McKenna, J. (2018). Tackling doping in sport: a call to take action on the dopogenic environment. *British Journal of Sports Medicine*, 52, 1485-1486. doi:10.1136/bjsports-2016-097169
53. Petróczi, A., & Aidman, E. (2009). Measuring explicit attitude toward doping: Review of the psychometric properties of the Performance Enhancement Attitude Scale. *Psychology of Sport & Exercise*, 10, 390-396. doi:10.1016/j.psychsport.2008.11.001
54. Nicholls, A. R., Madigan, D. J., & Levy, A. R. (2017). A confirmatory factor analysis of the performance enhancement attitude scale for adult and adolescent athletes. *Psychology of Sport & Exercise*, 28, 100-104. doi:10.1016/j.psychsport.2016.10.010
55. Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and

reporting: the state of the art and future directions for psychological research.

Developmental Review, 41, 71-90. doi:10.1016/j.dr.2016.06.004

56. Hambleton, R. K. (1980). Test score validity and standard setting methods. In R.A. Berk (Ed.). *Criterion-referenced measurement. The state of the art* (pp. 80-123). Baltimore, MD: The Johns Hopkins Press.
57. 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 & Exercise*, 36, 57-70. doi:10.1016/j.psychsport.2018.01.007.
58. Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 81, 358–361.
59. Lucidi, F., Zelli, A., Mallia, L., Grano, C., Russo, P. M., & Violani, C. (2008). The social cognitive mechanisms regulating adolescents' use of doping substances. *Journal of Sports Sciences*, 26, 447–456. doi:10.1080/02640410701579370.
60. Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality & Social Psychology*, 83, 1423-1440. doi:10.1037/0022-3514.83.6.1423

DOPING WILLINGNESS IN SPORT SCALE

Table 1. Scale items, and factor loadings from CFA across studies 1 to 4.

Would you be willing to use a banned substance if....	Study 1 (9-item)	Study 2 (9-item)	Study 2 (8-item)	Studies 3 & 4 (8-item)
1. It increased your chances to gain a professional contract or funding	.80 (.36)	.74 (.53)	.75 (.52)	.82 (.43)
2. <i>You have been heavily underperforming</i>	.84 (.29)	.76 (.24)	-	-
3. You suffered an injury and needed to recover quickly	.79 (.38)	.71 (.53)	.69 (.56)	.71 (.59)
4. You thought everyone you were competing against was using a banned substance and getting away with it	.80 (.37)	.75 (.39)	.73 (.41)	.78 (.51)
5. You were struggling to keep up in training/competition with those around you	.86 (.26)	.84 (.20)	.82 (.22)	.88 (.17)
6. You were told that you needed to bulk up because all the other players/ athletes were much bigger and stronger than you	.73 (.46)	.78 (.40)	.79 (.38)	.85 (.30)
7. You were offered them by someone you trusted (e.g., coach, friend, team mate, family member)	.73 (.46)	.76 (.34)	.78 (.33)	.83 (.35)
8. It increased your chances of getting selected (for the team)	.88 (.22)	.87 (.26)	.88 (.23)	.88 (.24)
9. You became more attractive to others	.76 (.42)	.71 (.41)	.72 (.40)	.65 (.92)

Note: Error variances are reported in parentheses.

DOPING WILLINGNESS IN SPORT SCALE

Table 2. Descriptive statistics, internal consistency, and relationships with doping willingness using structural equation modelling in Study 2.

Variable	<i>M</i>	<i>SD</i>	<i>a</i>	ω	<i>AVE</i>	<i>Correlation</i>	<i>95% CI</i>	<i>Shared variance</i>	<i>Partial correlation</i>	<i>95% CI</i>	<i>Shared variance (partial)</i>
Doping willingness	1.62	0.78	.92	.92	.60	-	-	-	-	-	-
Prosocial teammate	4.25	0.50	.70	.72	.40	.06	-0.08 to 0.20	.00	.07	-0.07 to 0.20	.00
Prosocial opponent	3.09	0.87	.75	.78	.56	.10	-0.05 to 0.24	.01	.12	-0.02 to 0.27	.01
Antisocial teammate	2.29	0.75	.82	.82	.48	.15*	0.01 to 0.29	.02	.10	-0.04 to 0.25	.01
Antisocial opponent	2.58	0.76	.83	.83	.38	.30***	0.15 to 0.45	.09	.27***	0.12 to 0.43	.07
Moral disengagement	3.32	1.05	.78	.79	.33	.31***	0.18 to 0.45	.10	.28***	0.13 to 0.42	.08

Note: Partial correlations include social desirability as a covariate in the model. *a* = Cronbach's alpha coefficient; ω = McDonald's Omega coefficient; AVE = average variance extracted. * $p < .05$, ** $p < .01$, *** $p < .001$.

DOPING WILLINGNESS IN SPORT SCALE

Table 3. CFAs on valence, arousal and moral judgment rating scales toward pictures in Study 3.

Variable	<i>df</i>	$SB\chi^2$	CFI	TLI	RMSEA	SRMR
Valence (doping)	8	11.18 (ns)	0.98	0.97	0.05	0.04
Arousal (doping)	8	15.97*	0.98	0.97	0.08	0.04
Judgement (doping)	9	4.23 (ns)	1.00	1.06	0.00	0.03
Valence (aggressive)	9	11.01 (ns)	0.97	0.95	0.04	0.05
Arousal (aggressive)	9	17.27*	0.98	0.97	0.08	0.03
Judgment (aggressive)	9	13.33 (ns)	0.95	0.92	0.06	0.04

Note: Valence and arousal ratings for doping pictures included one correlated error between 2 items due to a significant modification index and a high SEPC ($> .30$). ns = not significant, * $p = .04$.

DOPING WILLINGNESS IN SPORT SCALE

Table 4. Descriptive statistics, internal consistency, and relationships with doping willingness using structural equation modelling in Study 3.

Variable	<i>M</i>	<i>SD</i>	<i>a</i>	ω	AVE	Correlation	95% CI	Shared variance	Partial Correlation	95% CI	Shared variance (partial)
Doping willingness	1.68	0.81	.93	.93	.63	-	-	-	-	-	-
Valence (doping)	3.33	1.09	.76	.77	.38	.30***	0.13 to 0.48	.09	.28***	0.11 to 0.46	.08
Arousal (doping)	5.14	1.95	.88	.88	.57	.01	-0.15 to 0.16	.00	-.08	-0.23 to 0.07	.01
Judgment (doping)	7.38	1.26	.77	.79	.41	-.41***	-0.61 to -0.21	.17	-.39***	-0.59 to -0.18	.15
Valence (aggressive)	2.14	1.06	.74	.75	.33	-.12	-0.27 to 0.02	.01	-.07	-0.22 to 0.08	.00
Arousal (aggressive)	6.35	2.10	.92	.93	.67	.04	-0.14 to 0.22	.00	-.04	-0.21 to 0.13	.00
Judgment (aggressive)	7.86	1.03	.78	.79	.39	-.10	-0.31 to 0.10	.01	-.11	-0.33 to 0.11	.01

Note: Partial correlations included social desirability and number of days (between DWiSS completion and picture viewing task) as covariates in the respective models. *a* = Cronbach's alpha coefficient; ω = McDonald's omega coefficient; AVE = Average variance extracted. * $p < .05$, ** $p < .01$, *** $p \leq .001$

DOPING WILLINGNESS IN SPORT SCALE

Table 5. Results of item stability in Study 4.

Item	Time 1		Time 2		Test-retest difference		PA	Median Sign Test		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Min	Max		≥ 1	0 diff	≤ -1
1	1.94	1.20	2.00	1.32	-3	+3	92%	12	30	8
3	1.68	1.00	2.00	1.28	-3	+2	92%	16*	27	7
4	1.94	1.22	1.70	1.04	-3	+2	86%	4	35	11
5	1.55	0.89	1.61	0.95	-1	+1	100% (2)	6	38	4
6	1.54	0.91	1.59	0.91	-1	+1	100% (1)	7	37	5
7	1.80	1.03	1.80	1.10	-2	+2	90% (1)	7	36	6
8	1.68	1.02	1.80	1.15	-2	+2	94% (1)	10	34	5
9	1.68	1.13	1.65	1.15	-2	+2	92% (1)	6	35	8

Note: PA = Proportion of Agreement. Numbers in parentheses in the PA column represent the number of missing data for that item across the two time-points (all missing values, apart from one, were from the same participant at time-point 2). Original item 2 has been removed from the scale as noted in Study 2. * $p < .05$ (on median sign test).

DOPING WILLINGNESS IN SPORT SCALE

Table 6. Model fit indices for gender invariance testing.

Variable	<i>df</i>	$SB\chi^2$	CFI	TLI	RMSEA	SRMR
Men (baseline)	20	50.02***	.979	.970	.058	.025
Women (baseline)	20	23.73	.993	.990	.030	.028
Configural invariance	40	74.73**	.982	.975	.051	.027
Metric invariance	-	-	-	-	-	.045
Scalar invariance	-	-	-	-	-	.049

Note: * $p < .05$, ** $p < .01$, *** $p < .001$