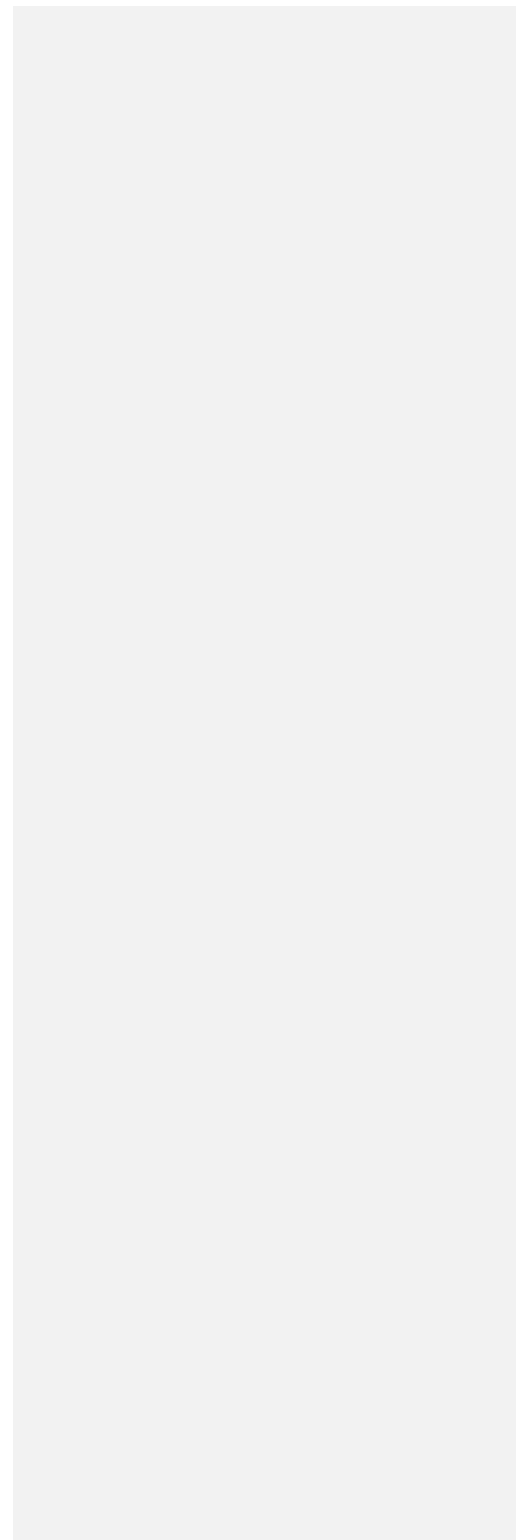


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Development of Moral Disengagement and Self-Regulatory Efficacy
Assessments Relevant to Doping in Sport and Exercise



Abstract

Objectives: To develop Moral Disengagement (MD) and Self-Regulatory Efficacy (SRE) instruments relevant to doping in sport and exercise and provide evidence for the validity and reliability of instrument scores.

Design: Cross-sectional, correlational.

Methods: Data were collected from male and female team- and individual-sport athletes and corporate- and bodybuilding-gym exercisers. Two samples ($n_{sample\ 1} = 318$; $n_{sample\ 2} = 300$) were utilized in instrument development and score validation and another ($n_{sample\ 3} = 101$) in examining test-retest reliability and stability of scores. Samples 1 and 2 responded to the newly developed items alongside others assessing theoretically-related variables, whereas Sample 3 completed the new instruments on two separate occasions.

Results: Factor analyses identified the final items and dimensional structures for the Doping Moral Disengagement Scale (DMDS), Doping Moral Disengagement Scale–Short (DMDS–S) and Doping Self-Regulatory Efficacy Scale (DSRES). The DMDS has six lower- and one higher-order factor, whereas the DMDS-S and DSRES are unidimensional. These structures were invariant by sex and sport/exercise context. Evidence supporting external validity, test-retest reliability, and stability of scores was also provided.

Conclusion: This research developed and provided evidence of score validity and internal consistency for three instruments relevant to doping in sport and exercise.

Keywords: Empathy, anticipated guilt, performance-enhancing drugs, measurement

1 Development of Moral Disengagement and Self-Regulatory Efficacy

2 Assessments Relevant to Doping in Sport and Exercise

3 Performance and Image Enhancing Drug (PIED) use can have detrimental health and
4 legal consequences for athletes and exercisers (McVeigh & Begley, in press; Pope, Wood,
5 Rogol, Nyberg, Bowers, & Bhasin, 2013) and is a behavior that raises substantive moral
6 questions (Donovan, Egger, Kapernick, & Mendoza, 2002; Petróczi & Aidman, 2008). Although
7 accurate prevalence rates are difficult to obtain, the estimated prevalence of PIED use in athletes
8 is 5-31% (Momaya, Fawal, & Estes, 2015). In light of this nontrivial prevalence, an important
9 aim for researchers is to understand psychological factors that influence PIED use, also known as
10 doping. Successful pursuit of such understanding requires the development of psychometric
11 instruments that provide valid and reliable scores for variables believed to contribute to doping.
12 Thus, we sought to develop psychometric instruments for assessing two psychological variables
13 of conceptual relevance to doping and validate their scores.

14 The theoretical framework for the current work was Bandura's (1991) social cognitive
15 theory of morality. Bandura proposed that harmful activities are deterred when people anticipate
16 negative emotional reactions (e.g., guilt) to engaging in them. However, people can reduce or
17 eliminate anticipation of such reactions through any of eight psychosocial mechanisms
18 collectively termed Moral Disengagement (MD). These mechanisms cognitively distort harmful
19 acts, reduce personal accountability for them and/or their consequences, distort/avoid their
20 consequences, or dehumanize or blame the victim of the act (Bandura 1991). Representing the
21 conditional endorsement of harmful acts, MD may facilitate doping by allowing sport and
22 exercise participants to use PIED without experiencing negative emotional reactions.

1 Qualitative research has shown that sport and exercise participants who dope demonstrate
2 MD when explaining their doping. For example, Boardley and Grix (2014) conducted semi-
3 structured interviews with nine bodybuilders who had doped. Analysis of the interview data
4 revealed evidence of six of the eight MD mechanisms. Boardley, Grix, and Dewar (2014)
5 expanded this work with 64 male bodybuilders from across England, all with experience of
6 doping. Content analysis again revealed evidence of the same six MD mechanisms. Boardley,
7 Grix, and Harkin (2015) extended this line of research by interviewing twelve male team- and
8 individual-sport athletes who had doped; data analysis again revealed the same six MD
9 mechanisms. Therefore, these three studies have provided consistent evidence for use of the
10 same six MD mechanisms across sport and exercise contexts. Further, there is considerable
11 consistency in the way in which sport and exercise participants use these six mechanisms,
12 supporting the potential benefits of developing a single measure of doping MD appropriate for
13 use in both contexts.

14 Definitions for each of these six MD mechanisms have been provided by Bandura (1991).
15 The first – *moral justification* – occurs when harmful activities are made personally and socially
16 acceptable by portraying how they achieve commendable social or moral purposes. Boardley et
17 al. (2014) found professional bodybuilders used this mechanism to justify PED use by suggesting
18 it allowed them to financially support their families: ‘So the ethics were skewed a bit towards
19 putting food on the table, rather than it is ethically right to take these and to do these things’
20 (p.838). The second – *euphemistic labelling* – diminishes the damaging nature of actions through
21 palliative or convoluted language. This mechanism was evidenced in Boardley et al. (2015) by a
22 mixed martial artist identifying use of a euphemistic term to hide the socially unaccepted nature
23 of doping: ‘Juice ... it kinda hides the fact that they’re frowned upon’ (p.7). The third –

1 *advantageous comparison* – makes detrimental conduct appear less damaging by comparing the
2 act to more heinous ones. Bodybuilders demonstrate this mechanism when making favorable
3 comparisons between doping and other harmful lifestyle behaviors, as shown by Boardley and
4 Grix (2014): ‘is doing this any worse than someone who goes out and has three or four pints [of
5 beer] every night’ (p.9).

6 The fourth mechanism – *displacement of responsibility* – diminishes personal
7 accountability for harmful behavior or its consequences by proffering the act resulted from social
8 pressures. Boardley et al. (2015) provided an example of this when a swimmer stated: ‘because it
9 wasn’t my idea to take them I feel ok about it’ (p.7). The fifth – *diffusion of responsibility* –
10 diminishes personal accountability for harmful acts or their outcomes through group decision
11 making or action. This mechanism is most likely to operate in environments where doping is
12 perceived as highly prevalent, as shown by a bodybuilder explaining the impact of such settings:
13 ‘the longer I’m here the more keen I am to do stuff I shouldn’t really or would never ever of
14 considered’ (Boardley & Grix, 2014, p.8). The final mechanism – *distortion of consequences* –
15 occurs when perpetrators of harmful acts avoid information relating to the harm caused or
16 downplay its significance. Such distortion was evidenced in Boardley et al. (2015) when an
17 American Footballer stated: ‘I didn’t ever see the people I played against as disadvantaged,
18 we’ve all got testosterone in our bodies, I just had more’ (p.7).

19 Moral disengagement has also been linked with doping in quantitative research. For
20 instance, researchers have identified positive links between MD, intention to dope, and reported
21 doping (e.g., Hodge, Hargreaves, Gerrard, & Lonsdale, 2013; Lucidi, Grano, Leone, Lombardo,
22 & Pesce, 2004; Lucidi, Zelli, & Mallia, 2013; Lucidi, Zelli, Mallia, Grano, Russo, & Violani,
23 2008; Ring & Kavussanu, in press) using cross-sectional and longitudinal designs. Although

1 these studies provide support for the potential importance of MD to doping in sport, prevalence
2 rates for doping were either low or not assessed. Also, to date researchers have not individually
3 examined the importance of the six relevant MD mechanisms in doping research.

4 Another variable from Bandura's (1991) theory that has been empirically linked with
5 doping is self-regulatory efficacy (SRE; Lucidi et al., 2008). Self-regulatory efficacy reflects the
6 belief in one's capabilities to resist personal and social pressures to engage in harmful conduct
7 (Bandura, Caprara, Barbaranelli, Pastorelli, & Regalia, 2001), and increases in SRE should lead
8 to less frequent engagement in such behavior (Bandura, 1991). This is because SRE should
9 increase one's ability to resist temptations and inducements to transgress. When applied to
10 doping, SRE represents a person's belief in his/her ability to forbear personal and social
11 pressures to dope. Athletes with elevated levels of doping SRE should be able to resist pressures
12 to dope as they are able to foresee the potential negative consequences of doping and formulate
13 alternative – licit – means of enhancing performance. In accord with this theorizing, Lucidi et al.
14 (2008) and Ring and Kavussanu (2017), respectively, found negative associations between
15 doping SRE and intention to dope and doping likelihood, in research with university students.

16 Although Lucidi and colleagues (2008) developed psychometric instruments to assess
17 doping MD and doping SRE, several concerns exist regarding the development of these
18 instruments. One concern relates to the item-development process. Specifically, items were
19 developed based on interviews with 35 high-school students who played sport. However, no
20 information was provided as to whether any of these students had any experience with doping.
21 During these interviews, participants were asked to list situations in which (a) doping would or
22 should not be completely condemned (i.e., to inform doping MD items) and (b) doping would be
23 more likely (i.e., to inform doping SRE items). For the MD measure, the frequencies of common

1 situations were then summed and categorized into the MD mechanisms evoked, leading to the
2 selection of 21 items. Six of these items were then selected for use in the doping MD measure;
3 no information was provided on the process through which items were selected. For the SRE
4 measure, the researchers only described how 10 items were developed based on situations
5 described during the interviews. Thus, items for the two instruments were: (a) developed based
6 on interviews with sport participants with unknown experience of doping when psychometric
7 instruments should be developed using samples representative of intended end users (Clark &
8 Watson, 1995), (b) not appraised for content validity when this should be a key aspect of item
9 development (Haynes, Richard, & Kubany, 1995), and (c) selected based on unknown criteria.
10 Finally, the factor structure of neither instrument was appropriately examined, nor confirmed in a
11 second sample (see Fabrigar, Wegener, MacCallum, & Strahan, 1999).

12 Kavussanu, Hatzigeorgiadis, Elbe, and Ring (2016) also developed a measure of doping
13 MD, following a more rigorous development process than Lucidi et al. (2008). This scale – the
14 Moral Disengagement in Doping Scale (MDDS) – was developed across three studies with team
15 and individual sport athletes, and evidence supporting the validity and reliability of scale scores
16 was provided. The MDDS is a six-item unidimensional scale assessing doping MD in team and
17 individual sport. However, it should be acknowledged that different versions of the scale are
18 used to assess doping MD in team compared to individual sport. This is potentially problematic
19 for research making comparisons between team- and individual-sport athletes as score
20 comparisons may be confounded by differences in item content.

1 Although the above measures¹ exist to assess doping MD and doping SRE, there are
2 several ways in which assessment of these constructs could be further developed. First, although
3 doping is an issue in exercise as well as sport (see Sjöqvist, Garle, & Rane, 2008), there are
4 currently no instruments available to assess doping MD and doping SRE in exercise populations.
5 This is particularly concerning when one considers doping in exercisers is considered a public
6 health issue (see McVeigh & Begley, in press; Pope et al., 2013). Thus, there is a need for the
7 development of instruments assessing doping MD and doping SRE in exercise populations.

8 As exercise in gymnasias is often part of the training process in sport (i.e., strength and
9 conditioning), many sportspersons frequently interact with exercisers as part of their preparatory
10 activities. Given this, and the suggestion that MD is socially transmitted (see Bandura, 1991), it
11 is perhaps not surprising that research has demonstrated considerable consistency in the way in
12 which sport and exercise participants utilize MD with respect to doping (see Boardley & Grix,
13 2014; Boardley et al., 2014, 2015). These studies also evidence considerable consistency in the
14 personal and social pressures to dope as perceived by sport and exercise participants. Thus, it
15 would appear logical to develop instruments capable of assessing doping MD and doping SRE
16 across sport and exercise populations. Such instruments would make a significant contribution to
17 doping research not only by allowing researchers to investigate doping MD and doping SRE in
18 exercise populations, but also by facilitating research in which levels of these constructs can be
19 directly compared across different sport and exercise populations.

20 Assessment of doping MD could also be furthered through the development of a
21 multidimensional measure. Bandura (1991) proposed MD to be multidimensional, describing
22 eight mechanisms of MD. However, there is currently no instrument capable of measuring

¹ Mallia, Lazuras, Barkoukis, et al. (2016) also developed measures assessing doping MD and doping SRE. However, as these instruments were designed to capture team rather than individual assessments of these constructs, a detailed review of their development and properties is not included here.

1 individual mechanisms of doping MD in either sport or exercise contexts. This is particularly
2 important when one considers six different mechanisms of MD have been shown to be relevant
3 to doping (see Boardley & Grix, 2014; Boardley et al., 2014, 2015). A multidimensional measure
4 of doping MD capable of measuring individual mechanisms of doping MD would make a
5 significant contribution, as it would allow researchers to investigate whether certain MD
6 mechanisms are of greater importance to doping than others.

7 A key consideration when developing psychometric instruments is that they are validated
8 with samples representative of the target population/s (Clark & Watson, 1995). As such, doping-
9 specific instruments should be developed and validated with participant groups representing the
10 wide range of doping prevalence rates estimated across athletic populations (e.g., Momaya et al.,
11 2015). To ensure this, sub-samples should be specifically targeted to represent such populations
12 when developing measures, and doping prevalence should be assessed. To date, this has not been
13 the case in research developing doping MD scales (Kavussanu et al., 2016; Lucidi et al., 2008).

14 A prominent issue when developing measures is the provision of evidence supporting the
15 convergent validity of scores (Messick, 1995). Such evidence is established if scores correlate at
16 least moderately with scores from instruments assessing variables within the target construct's
17 nomological network (Vaughn & Daniel, 2012). One variable that should be associated with
18 doping MD and doping SRE is anticipated guilt, representing a distasteful emotional state
19 experienced as tension and regret, resulting from the personal responsibility felt – and empathic
20 feelings for – someone suffering anguish (Hoffman, 2000). Anticipation of guilt is thought to be
21 diminished by MD, as tension and regret are less likely to be anticipated for actions we view
22 favorably, do not feel responsible for, and/or are perceived to be harmless (Bandura, 1991). In
23 contrast, doping SRE should increase anticipated guilt as we are likely to feel highly responsible

1 for engaging in harmful behaviors we feel capable of resisting. As such, doping MD and doping
2 SRE, respectively, should have negative and positive associations with anticipated guilt.

3 Another variable that should be linked with doping MD is empathy, which represents a
4 tendency to vicariously experience emotional and cognitive responses to another individual's
5 emotional state (Davis, 1983). Two sub-components of empathy shown (e.g., Kavussanu, Stamp,
6 Slade, & Ring, 2009) to be of specific relevance to sport are perspective taking (i.e., tendency to
7 spontaneously adopt the psychological point of view of others) and empathic concern (i.e.,
8 tendency to experience feelings of sympathy, compassion, and concern for unfortunate others).
9 Combining these two sub-components to form an overall measure of empathy has been shown to
10 be a suitable approach to assessing empathy in sport (e.g., Kavussanu et al., 2009) and non-sport
11 (e.g., Carlo, Raffaelli, Laible, & Meyer, 1999) research. Empathy is thought to impair MD
12 because endorsement of harmful conduct is more problematic when one can anticipate and
13 experience the consequences of one's actions for others (Bandura, 1991; Hoffman, 2000). A
14 negative relation between doping MD and empathy would therefore be expected.

15 Evidence for the discriminant validity of scores obtained with new measures is also
16 important to establish (Messick, 1995). Discriminant validity represents the degree to which
17 scores are empirically distinguishable (i.e., $r \leq .90$) from those of closely related but conceptually
18 distinct constructs (Vaughn & Daniel, 2012). For doping MD, a suitable construct for examining
19 discriminant validity is sport MD. Sport MD represents use of MD to justify and rationalize
20 engagement in transgressive acts on the sports field (Boardley & Kavussanu, 2007). Doping MD
21 and sport MD are closely related due to the incorporation of equivalent mechanisms of MD, but
22 distinct due to the differences in the acts to which the mechanisms are applied. For doping SRE,
23 an apposite construct for assessing discriminant validity is peer pressure SRE, representing the

Commented [IB1]: Comment 2

Commented [IB2]: Comment 2

1 belief in one's ability to resist peer pressure to engage in high-risk activities (e.g., use of drugs;
2 Bandura et al., 2001). Doping SRE and peer pressure SRE represent beliefs in one's capacity to
3 resist personal and social pressure to engage in harmful behavior, yet are distinct due to their
4 focus on different transgressive acts.

5 The overall objective of the current project was to develop instruments assessing doping
6 MD and doping SRE for use across sport and exercise populations, and to evaluate the validity of
7 their scores. High quality instruments assessing these constructs will allow researchers to
8 investigate doping MD and doping SRE in both contexts. Further, for doping MD, we aimed to
9 develop long and short versions of the new instrument. Researchers with aims pertaining to
10 individual mechanisms of doping MD could therefore use the long version, whereas those aiming
11 to study overall doping MD could take advantage of the conciseness of the short version. Guided
12 by theory and instruments developed to assess MD in other contexts, we anticipated the long
13 version would assess six dimensions of MD (Bandura, 1991; Boardley & Kavussanu, 2007;
14 Boardley & Grix, 2014, Boardley et al., 2014, 2015). In contrast, we expected the short version
15 of the MD instrument and the doping SRE instrument would be unidimensional (Boardley &
16 Kavussanu, 2008; Bandura et al., 2001). We also anticipated the dimensionality of all three
17 instruments would be invariant by sex and across four sport and exercise contexts (Boardley &
18 Kavussanu, 2008; Bandura et al., 2001). Finally, we expected scores on the respective
19 instruments to show stability over brief time periods.

20 When developing these new instruments and validating their scores, we considered five
21 of the six aspects of construct validity identified by Messick (1995). These were content,
22 substantive, structural, generalizability, and external. The content aspect relates to the relevance,
23 representativeness and technical quality of item content and was assessed presently through

1 expert opinion. We considered the substantive aspect – relating to the theoretical rationale for the
2 observed test responses – by ensuring item content was consistent with qualitative research
3 studying the process being assessed (e.g., Boardley & Grix, 2015; Boardley et al., 2014, 2015)
4 and examining theory-based relations between scores generated with the new measures and
5 conceptually associated variables. The structural aspect pertains to the fidelity of the scoring
6 structure to the structure of the construct domains being assessed and was examined through
7 factor analysis. We addressed the generalizability aspect – the extent to which score properties
8 and interpretations generalize to and across groups and settings – through multisample analyses.
9 The external aspect relates to evidence for convergent and discriminant validity as well as
10 criterion relevance, and was considered currently through relations with theoretically relevant
11 instrument scores. Evidence for the final aspect of construct validity – consequential – is
12 represented through positive and negative consequences stemming from application of new
13 measures. As such, this final aspect was more relevant to future application of the measures
14 rather than initial instrument development and score validation.

15 In sum, through this research we sought to develop three psychometric instruments: a) the
16 Doping Moral Disengagement Scale (DMDS), b) the Doping Moral Disengagement Scale –
17 Short (DMDS-S), and c) the Doping Self-Regulatory Efficacy Scale (DSRES). Across the two
18 studies below we followed appropriate procedures for developing such instruments and
19 validating their scores (i.e., Clark & Watson, 1995; Fabrigar et al., 1999; Haynes et al., 1995).

20 **Study 1**

21 **Method**

22 **Item Development.** Informed by a review of existing instruments assessing the target
23 constructs, qualitative papers investigating the constructs in physical-activity contexts, and

1 consultation with relevant experts (i.e., sport psychologists, sport coaches, exercise leaders, sport
2 and exercise participants), we developed large pools of items for doping MD and doping SRE
3 (see Clark & Watson, 1995)². Thirty-eight items representing doping MD and 13 for doping SRE
4 were generated. These items then underwent initial pilot testing with a sample of sport and
5 exercise participants ($N = 280$). For DMDS/DMDS-S items a 7-point Likert scale anchored by 1
6 (*strongly disagree*) and 7 (*strongly agree*) was used during pilot testing and all subsequent data
7 collections. This is consistent with existing MD instruments in sport (e.g., Boardley &
8 Kavussanu, 2007, 2008) and supported by research suggesting this format offers the best
9 compromise between reliability, validity, discriminatory power, and respondent preference
10 (Preston & Colman, 2000). However, to be consistent with existing SRE instruments (e.g.,
11 Bandura et al., 2001) the DSRES used a 5-point Likert scale anchored by 1 (*no confidence*) and 5
12 (*complete confidence*). Inter-item correlations were computed with the pilot data to ensure mean
13 values were between .15 and .50 for each subscale (see Clark & Watson, 1995); items with inter-
14 item correlations inconsistent with this range were eliminated or adapted. As a result of these
15 analyses, 28 doping MD items were retained and 10 were adapted. Also, five additional items
16 were generated based on participant feedback and pilot data analyses. For doping SRE, 11 items
17 were retained and two adapted; no additional items were developed. Following these changes,
18 the item pool consisted of 43 doping MD³ and 13 doping SRE items.

19 The item pools were then subjected to content-validity assessment to determine whether
20 they characterized their intended domain (Dunn, Bouffard, & Rogers, 1999; Haynes et al., 1995).

² The term "athlete" was used to refer to participants regardless of sport or exercise context. This was based upon qualitative research that shows participants from a range of physical-activity contexts (e.g., sport, exercise) use this term to refer to participants (i.e., Boardley & Grix, 2014, Boardley et al., 2014, 2015).

³ Moral Justification = 8 items; Euphemistic Labelling = 9 items; Advantageous Comparison = 6 items; Displacement of Responsibility = 6 items; Diffusion of Responsibility = 6 items; Distortion of Consequences = 8 items

Commented [1B3]: Comment 1.

1 Content validity was examined through expert opinion, and the size and composition of the
2 expert panel was informed by relevant guidelines (Dunn et al., 1999). Assessment packs were
3 emailed to 12 academics⁴, none of whom were involved in the item-writing phase. Each expert
4 had a PhD in sport psychology or psychology, had published in peer-reviewed international sport
5 psychology journals, and was employed in a psychology, kinesiology, or sport science
6 department in Europe, North America, or Australia. Packs consisted of ten sections: a)
7 introduction and instructions, b) moral justification, c) euphemistic labelling, d) advantageous
8 comparison, e) displacement of responsibility, f) diffusion of responsibility, g) distortion of
9 consequences, h) response format, scale, and general comments (MD), i) doping SRE, and j)
10 response format, scale, and general comments (doping SRE). Sections b-g, and i consisted of a
11 definition of the relevant construct followed by the items developed to assess the relevant
12 construct. Experts were asked to rate each item using a Likert-type scale ranging from -3 (*Not at*
13 *all Representative*) to 3 (*Very Representative*), and were provided with the opportunity to
14 comment on the relevance of each item to the context and definition. Sections h and j presented
15 the proposed response format and scale, and asked the experts to comment on these, the match of
16 the instrument's attributes to its function, and suggested additions, deletions, and modifications.

17 For each item, we took the mean expert rating after removing any ratings that deviated
18 notably from the other judges (see Hambleton, 1980); such ratings were defined as those that
19 were two or more response options lower than the next lowest score (e.g., scoring -2 when the
20 next lowest was 0). Items with mean expert rating scores of 2.0 or above were retained; those
21 with scores below 2.0 were revised based on expert comments or removed. Of the 56 items
22 assessed, 26 were retained without change, 17 underwent **minor** content revisions, and 13 were

⁴ 16 potential experts were contacted, 12 of whom agreed to act as content-validity assessors.

1 removed; six new doping MD items were created⁵. **The content validity of revised and newly**
2 **created items was examined by members of the research team not involved in item creation.**

3 These 49 items (i.e., 36 doping MD) were then pilot tested with a sample of 122 sport and
4 exercise participants to: (a) ensure correlations among theoretically related items remained
5 consistent with the target mean inter-item correlation for subscales (i.e., .15 to .50), (b) examine
6 the internal consistency of scores from all subscales, and (c) obtain feedback on item difficulty
7 and wording (Clark & Watson, 1995). Correlation analyses demonstrated 86% of doping MD
8 items were correlated in the target range; items not correlated as expected were adapted to
9 improve wording. All doping SRE items were correlated in the target range. Cronbach alpha
10 coefficients ranged from .70 to .85 ($M = .78$) for the doping MD subscales and was .94 for the
11 doping SRE items. Feedback on item difficulty and wording was positive, with only minor
12 adjustments to wording needed for three items. Following item development, these 49 items
13 were taken forward into the main construct-validity phase.

14 **Participants.**

15 **Sample 1.** Participants were team- (e.g., American football, soccer) ($n = 181$) or
16 individual- (e.g., swimming, athletics) ($n = 70$) sport or bodybuilding⁶- ($n = 44$) or corporate⁷- (n
17 $= 23$) gym participants ($n_{\text{male}} = 203$; $n_{\text{female}} = 115$); ages ranged from 16 to 70 years ($M = 23.3$,
18 $SD = 8.2$). Sport participants competed at a recreational ($n = 6$), local ($n = 3$), university ($n =$
19 131), regional ($n = 20$), national ($n = 40$), or international ($n = 44$) level (seven failed to report).
20 Participants had been training/competing for an average of 7.3 years ($SD = 5.4$), spent an average

⁵ Only one of these items was ultimately selected for use in a final doping MD measure.

⁶ Bodybuilding gymnasias are those targeting and designed for serious and experienced bodybuilders. Such gymnasias have no aerobic fitness programmes and the equipment within them is largely free weights.

⁷ Corporate gymnasias are those targeting and designed for users training for general health and fitness. Such gymnasias have comprehensive aerobic fitness programmes and the equipment within them is largely resistance and aerobic fitness machines.

1 of 8.6 hours ($SD = 3.9$) per week training, and had trained in/with their current gym/team for an
2 average of 4.1 years ($SD = 4.5$). Self-reported lifetime prevalence of doping was 14.5%⁸.

3 **Sample 2.** Participants were team- (e.g., Australian rules football, soccer) ($n = 14$) or
4 individual- (e.g., athletics, triathlon) ($n = 99$) sport or bodybuilding- ($n = 89$) or corporate- ($n =$
5 98) gym participants ($n_{male} = 172$; $n_{female} = 128$); ages ranged from 17 to 73 years ($M = 29.6$, SD
6 = 12.4). Sport participants competed at a recreational ($n = 18$), local ($n = 18$), university ($n = 25$),
7 regional ($n = 16$), national ($n = 21$), or international ($n = 10$) level (five failed to report).
8 Participants had been training/competing for an average of 9.1 years ($SD = 8.6$), spent an average
9 of 7.9 hours ($SD = 5.0$) per week training, and had trained in/with their current gym/team for an
10 average of 3.8 years ($SD = 4.9$). Self-reported lifetime prevalence of doping was 15.3%.

11 **Measures.** To help evaluate the scores obtained with the new instruments, a series of
12 existing measures were administered when collecting data from Samples 1 and 2. These included
13 instruments assessing sport MD, peer pressure SRE, anticipated guilt, and empathy.

14 **Sport moral disengagement.** The eight-item moral disengagement in sport scale-short
15 (MDSS-S; Boardley & Kavussanu, 2008) was used to measure sport MD. Participants read
16 statements representing MD (e.g., 'Insults among players do not really hurt anyone') and
17 indicated their level of agreement using a Likert scale anchored by 1 (*strongly disagree*) and 7
18 (*strongly agree*). Scores on this instrument have shown very good levels of internal consistency
19 and evidence for factorial and convergent validity (Boardley & Kavussanu, 2008). Good levels
20 of internal consistency were demonstrated in the present study ($\alpha_{Sample 1} = .87$; $\alpha_{Sample 2} = .89$).

21 **Peer pressure SRE.** The five-item SRE scale (Bandura et al., 2001) was used to assess
22 peer-pressure SRE. Items (e.g., 'Resist peer pressure to drink beer, wine or alcohol') assess

⁸ Prevalence rates by gender and sport/exercise context were as follows: males = 19.3%; females = 8.2%; individual sport = 5.9%; team sport = 10.3%; hardcore gym = 39.8%; corporate gym = 7.4%.

1 peoples' beliefs regarding their ability to resist peer pressure to engage in high-risk activities
2 involving alcohol, drugs and transgressive behavior. For each item, participants rated their
3 confidence using a Likert scale anchored by 1 (*no confidence*) and 5 (*complete confidence*).
4 Scores on this instrument have shown good levels of internal consistency and evidence for
5 construct validity (Bandura et al., 2001). Acceptable levels of internal consistency were
6 demonstrated in the present study ($\alpha_{\text{Sample 1}} = .77$; $\alpha_{\text{Sample 2}} = .81$).

7 ***Anticipated guilt.*** To assess anticipated guilt in response to doping, participants were
8 asked to imagine being in the following situation:

9 *Having returned to training following a period of injury, you are feeling very out of shape.*
10 *As such, you feel the need to get back in shape as soon as possible. A friend who you train*
11 *with has been taking a training supplement that he/she says really helped him/her get back*
12 *in shape quickly following a similar injury. He/she offers to give you some and you decide*
13 *to take it. Subsequently you get back in shape much quicker than expected, but then*
14 *discover the supplement you have been taking is a banned performance-enhancing*
15 *substance. However, due to the improvements you have experienced, you decide to*
16 *continue taking the substance.*

17 Participants expressed how they anticipated feeling about continuing to dope by responding to
18 the five items (e.g., 'I would feel remorse, regret') of the guilt scale from the State Shame and
19 Guilt Scale (SSGS; Marschall, Saftner, & Tangney, 1994). Participants responded on a 5-point
20 scale ranging from 1 (*not at all*) to 5 (*extremely*). Marschall et al. (1994) provided evidence
21 supporting the internal consistency and validity of scores from this subscale. This method of
22 assessing anticipated guilt has been used successfully in sport (Stanger, Kavussanu, Boardley, &

1 Ring, 2013). Excellent levels of internal consistency were demonstrated in the present study
2 ($\alpha_{\text{Sample 1}} = .94$; $\alpha_{\text{Sample 2}} = .95$).

3 **Empathy.** The 7-item perspective taking (e.g., ‘before criticizing somebody, I try to
4 imagine how I would feel if I were in their place’) and 7-item empathic concern (e.g., ‘I am often
5 quite touched by things that I see happen’) subscales of the Interpersonal Reactivity Index
6 (Davis, 1983) were used to measure empathy. Participants indicated how well each statement
7 described them by responding on a scale from 1 (*does not describe me well*) to 7 (*describes me*
8 *very well*). Scores from these respective subscales have shown internal consistency reliability
9 and validity (Davis, 1983). In line with previous research (Kavussanu, Stanger, & Boardley,
10 2013), we combined these subscales to produce one empathy score. Good levels of internal
11 consistency were demonstrated in the present study ($\alpha_{\text{Sample 1}} = .86$; $\alpha_{\text{Sample 2}} = .84$).

12 **Reported Doping.** Reported doping was assessed using a method based on that of Lucidi
13 et al. (2008). Specifically, participants were provided with a list of nine categories of doping
14 substances (e.g., Ephedrine stimulants) and methods (e.g., Blood manipulation) and asked to
15 indicate which they currently used, had used in the past three months, had used prior to the past
16 three months, or had never used. The list was based on substances/methods banned by the World
17 Anti-Doping Agency. Lifetime use of doping was indicated by use of one or more of the
18 substances/methods during any of the time points indicated.

19 **Procedures.**

20 **Recruitment and data collection.** Recruitment for the primary samples commenced once
21 approved by the ethics committee of the first author’s institution. Our approach to recruitment
22 differed for sport versus exercise participants. For sport participants, we contacted sport coaches
23 and sought permission to visit a designated training session to introduce the project to athletes

1 and invite them to participate. For exercise participants, we contacted gym managers and sought
2 permission to visit their gymnasias to introduce the project to exercisers and invite them to
3 participate. Before completing the questionnaire, all volunteers were informed that honesty in
4 responses was vital to the study, and that all responses would be kept confidential and be used
5 only for research purposes. Participants signed an informed consent form prior to participating
6 and the questionnaire pack took 10–15 minutes to complete. Data were collected across two
7 phases. Data from the first phase (i.e., Sample 1) were analyzed before the second phase (i.e.,
8 Sample 2) commenced. This allowed for item adjustment/generation between phases.

9 ***Factorial, convergent, and discriminant validity, and internal consistency.*** To provide
10 further evidence for construct validity, we sought to establish evidence for the factorial,
11 convergent and discriminant validity of scores from the new instruments, as well as for internal
12 consistency. Factorial validity relates to the number of separate dimensions represented in a
13 measure and was analyzed by identifying the factor structure in the first sample before
14 confirming it in the second. To establish evidence for convergent validity, associations of
15 DMDS, DMDS-S, and DSRES scores with empathy and anticipated guilt were computed. To
16 establish evidence for discriminant validity, we analyzed correlations of DMDS and DMDS-S
17 scores with sport MD scores, and of DSRES scores with peer pressure SRE scores. Here and
18 elsewhere, the magnitudes of correlation coefficients were interpreted based on the guidelines
19 provided by Cohen (1992), with coefficients of 0.10, 0.30, and 0.50 representing small,
20 moderate, and large effect sizes, respectively. Finally, the internal consistency of scores for the
21 overall scales plus the DMDS subscales were estimated using Cronbach's alpha.

22 **Results**

1 **Preliminary analyses.** Only 0.83% of data points were missing, and missing data were
2 assumed missing at random (see Enders, 2006). The expectation maximization algorithm was
3 used to impute missing values. Before seeking to establish evidence for the factorial validity of
4 instrument scores, we used a two-stage process to identify the most effective items for measuring
5 each construct; selected items were retained for use in subsequent testing. First, inter-item
6 correlations were examined within each construct and all item scores were intercorrelated as
7 anticipated (see Clark & Watson, 1995).

8 Exploratory factor analysis (EFA) was then conducted on each of the seven constructs
9 (i.e., six MD mechanisms plus doping SRE) using principal axis extraction, with extraction
10 based on an eigenvalue ≥ 1.00 . Prior to conducting these analyses, we determined the
11 appropriateness of the matrices using the following criteria as suggested by Dziuban and Shirkey
12 (1974): (a) a significant Bartlett's test of sphericity and (b) a Kaiser-Meyer-Olkin measure of
13 sampling adequacy of $> .80$. All seven matrices satisfied these criteria. Each EFA produced a
14 unidimensional factor structure, and except for one item, all items had factor loadings ≥ 0.61 .
15 The one exception was a displacement of responsibility item (i.e., "Seeing athletes achieve goals
16 through doping encourages others to dope too"), with a factor loading of $.32$. Given this, we
17 removed the item from further analyses and replaced it with a newly developed item (i.e.,
18 "Athletes shouldn't be held responsible for doping if they feel pressured to do it to keep up with
19 others") before data collection for Sample 2 commenced. All other items were retained.

20 **Factorial validity.** Confirmatory factor analysis was used in the next step because it
21 offers a rigorous and appropriate method for confirming hypothesized factor structures (Fabrigar
22 et al., 1999). EQS 6.1 (Bentler & Wu, 2002) software and the Maximum Likelihood method
23 were used. In initial analyses, the normalized estimate of Mardia's coefficient indicated deviation

⁹ This item was not used in either of the final MD instruments.

Commented [IB4]: Comment 6

1 from multivariate normality. Thus, the Robust Maximum Likelihood estimation method was
2 used for all analyses. This method provides more accurate standard errors, chi-squared values,
3 and fit indices when data are non-normally distributed (Bentler & Wu, 2002). The cases with the
4 largest contribution to normalized multivariate kurtosis minimally impacted the analyses and
5 therefore no cases were deleted. Standard indices and criteria were used to estimate model fit
6 (Hu & Bentler, 1999). Fit indices for all models appear in Table 1.

7 In the first DMDS model, six items were specified for each of the six MD mechanisms,
8 except for displacement of responsibility which was represented by five (M1a). Results showed
9 an inadequate fit for the model (Row 1). Subsequently, 17 items implicated in large modification
10 indices as indicated by the Lagrange multiplier (LM) test and/or large standardized residuals
11 were removed in a series of CFAs. This iterative process was guided by an aim to develop an
12 instrument that would contain 18 items (i.e., three for each MD mechanism). A final model
13 (M1b) with 18 items produced a six-factor solution with excellent fit (Row 2). Theoretically the
14 six factors represent different dimensions of an overriding construct, and as such we also
15 examined whether the associations amongst the six first-order factors could be represented by a
16 higher-order factor. Although the fit of this model (Row 3) was reduced compared to that of the
17 corresponding first-order model (Row 2), as the fit of a second-order model cannot be better than
18 the fit of the equivalent first-order one, it was sufficient to support the presence of a second-order
19 structure (Marsh, 1987). This second-order factor was named doping MD.

20 Although our hypothesized model was supported, it was important to rule out alternative
21 models. For instance, the development of MD instruments for other contexts has shown that pairs
22 of MD mechanisms sometimes converge to form single factors (e.g., Boardley & Kavussanu,
23 2007) or can all converge to form a unidimensional measure (e.g., Bandura, Barbaranelli,

1 Caprara, & Pastorelli, 1996). Thus, once the final item content for the DMDS was confirmed, we
2 compared the fit of the six-factor model with two other possible structures based on those seen in
3 existing instruments. Specifically, these were a three-factor model in which the six mechanisms
4 were grouped according to the aspect/s of detrimental conduct they operate upon (M2; Osofsky,
5 Bandura, & Zimbardo, 2005) and a unidimensional model in which all items loaded on a single
6 factor (M3; Bandura et al., 1996). Also, based on very strong factor correlations between
7 advantageous comparison/distortion of consequences (i.e., .93) and displacement/diffusion of
8 responsibility (i.e., .91), we also tested a four-factor model (M4) in which these respective
9 mechanism pairs were combined into single factors. As shown in Table 1, the fit of the six-factor
10 model (M1b) was superior to any of these alternative models. Thus, the six-factor model was
11 accepted as the best model for the DMDS. Factor correlations for this model can be found in
12 Table 2, and items, factor loadings, and error variances are shown in Table 3.

13 Although the 18-item instrument provides the capability of measuring the six MD
14 mechanisms individually, some users may only need to measure overall doping MD. For such
15 instances, a short version would reduce the time needed for completion. For these reasons, we
16 also developed the DMDS-S. Our aim was to develop a six-item instrument with one item for
17 each of the six relevant mechanisms of MD. Two main steps were involved in selecting items for
18 the DMDS-S. Potential items were first selected based on item content, with the 12 (i.e., two for
19 each MD mechanism) shortest and simplest items retained for further analysis. Then, the factor
20 structure underlying these 12 items was examined using EFA and CFA with the data from
21 Sample 1; the suitability of the matrix for EFA was determined using the same approach as
22 described earlier, and EFA conducted using the same approach. This resulted in a single factor
23 being extracted. Then, for each of the six MD mechanisms the item with the stronger factor

1 loading was retained. Once these six items had been identified, a single-factor CFA was
2 specified (i.e., M5a). As seen in Table 1 (Row 7), this model had a very good fit. However, the
3 LM Test results indicated the presence of a correlated error between the advantageous
4 comparison and distortion of consequences items. Testing of a subsequent model (M5b) with this
5 correlation specified resulted in an excellent model fit. Specifying correlated errors when present
6 is important to prevent possible inaccurate parameter estimates (see Kline, 2015). However, such
7 associations can be sample specific so this model was accepted under the proviso that the
8 presence of this correlated error would be confirmed in Sample 2.

9 When developing psychometric instruments, it is important to confirm factor structures
10 using a separate sample (Fabrigar et al., 1999). As such, we used the data from Sample 2 to
11 confirm the factor structures for the DMDS and DMDS-S. As shown in Table 1, the final DMDS
12 model from the Sample 1 analyses showed excellent model fit (Row 11) and again supported the
13 presence of a second-order doping MD factor (Row 12). This was also the case for the DMDS-S
14 model (Row 13), where the significant correlated error identified in Sample 1 (i.e., $r = .30$, p
15 $< .05$) was again present (i.e., $r = .25$, $p < .05$).

16 Similar procedures were followed for the development of the DSRES. Using the Sample
17 1 data, initially all 13 items were specified to load on a single factor (M6a). This model had an
18 inadequate fit (Row 9). Guided by modification indices and/or standardized residuals, 7 items
19 were removed in a series of CFAs. This iterative process was guided by an aim to develop an
20 instrument that would contain items representing the six main personal and social influences on
21 doping use identified in past research (e.g., Boardley & Grix, 2014; Boardley et al., 2014, 2015).
22 A final six-item model (M6b) demonstrated an excellent model fit (Row 10). Subsequent testing
23 of this model using the data from Sample 2 also resulted in excellent fit (Row 14), confirming

1 the structure of the DSRES scores. Items, factor loadings and error variances for the DMDS-S
2 and DSRES appear in Table 4.

3 **Multisample analyses.** When developing instruments for use in diverse populations, it is
4 important to determine their measurement invariance across sub-groups. As such, we tested for
5 measurement invariance by sex and across the four sport and exercise groups represented using
6 multisample analyses. Different aspects of invariance can be tested depending on the research
7 question (Cheung & Rensvold, 2002). As we were interested in construct validity and whether
8 the instruments were appropriate for making comparisons amongst groups, we tested three
9 relevant aspects of invariance (Byrne, 2006): a) configural invariance (i.e., items are indicators
10 of the same factors in all groups), b) metric invariance (i.e., all factor loadings are equal across
11 groups), and c) equivalence of construct variance and covariance (ECVC; i.e., variances and
12 covariances of latent variables are equivalent across groups). Prior to invariance testing, we
13 estimated baseline model fit for each sub-group (see Byrne, 2006). We then tested for configural
14 invariance, metric invariance and ECVC, respectively, by progressively imposing the appropriate
15 constraints. We examined ΔCFI at each step; values of less than 0.01 indicate no significant
16 difference between models (Cheung & Rensvold, 2002). Results are presented in Table 5.

17 **Sex invariance.** For the DMDS, model fit for the baseline models was very good for
18 male participants and acceptable-to-good for female participants, and configural invariance was
19 demonstrated by good model fit. Metric invariance was also established by a ΔCFI of .00. The
20 ECVC was not established though, as imposing such constraints resulted in a $\Delta CFI > .01$. For the
21 DMDS-S, model fit for baseline models was very good for male participants and excellent for
22 female participants, and configural invariance was demonstrated by the very good fit of the
23 relevant model. Metric invariance was also established, as shown by a $\Delta CFI < .01$. However, like

1 the DMDS, the ECVC was not established as $\Delta CFI > .01$. Finally, for the DSRES, model fit for
2 the baseline models was excellent for male and female participants, and configural invariance
3 was demonstrated by excellent model fit. Like the DMDS and DMDS-S, metric invariance was
4 established, as shown by a $\Delta CFI < .01$, and the ECVC was not supported, as $\Delta CFI > .01$.

5 ***Sport/exercise group invariance.*** For the DMDS, model fit for the baseline models
6 ranged from acceptable-to-good for corporate-gym exercisers to excellent for bodybuilding-gym
7 exercisers and individual-sport athletes. Configural invariance was established through good
8 model fit. Metric invariance was also established, as shown by a $\Delta CFI < .01$. However, the
9 ECVC was not established as $\Delta CFI > .01$. For the DMDS-S, model fit for the baseline models
10 was excellent for all groups, and configural invariance was demonstrated by excellent model fit.
11 Complete metric invariance was not established, as the LM test results for this model indicated
12 two of the specified constraints (i.e., factor loadings for the advantageous comparison and
13 diffusion of responsibility items between corporate gym exercisers and individual-sport athletes)
14 led to an increase in χ^2 of $\geq 5.0/df$. Respecification of the model with these constraints released
15 led to excellent model fit and a $\Delta CFI < .01$. This supports partial metric invariance (see Byrne,
16 Shavelson, & Muthén, 1989). However, ECVC was again not established. Finally, for the
17 DSRES, model fit for the baseline models was excellent for all groups, and configural invariance
18 was demonstrated by excellent model fit. Metric invariance was also established, as shown by a
19 $\Delta CFI < .01$. However, the ECVC was again not supported.

20 **Convergent and discriminant validity.** We examined convergent validity by computing
21 associations of DMDS, DMDS-S, and DSRES scores with scores for empathy and anticipated
22 guilt. As shown in Table 6, correlations of DMDS, DMDS-S, and DSRES scores with empathy
23 and anticipated guilt scores support convergent validity in both samples. This is also true for all

1 DMDS subscale scores, although the degree of convergence was weaker for the euphemistic
2 labelling subscale in comparison to the other five subscales. Collectively these correlations also
3 provide evidence of distinct predictive capabilities, supporting some degree of conceptual
4 separation between the subscales despite their largely strong inter-correlations (see Table 2).

5 We examined discriminant validity by computing the correlations of sport MD (Boardley
6 & Kavussanu, 2008) with DMDS and DMDS-S scores and of peer pressure SRE with DSRES
7 scores. In both samples, DMDS (Sample 1 $r = .59, p < .01$; Sample 2 $r = .58, p < .01$) and
8 DMDS-S (Sample 1 $r = .56, p < .01$; Sample 2 $r = .58, p < .01$) scores were positively related to
9 sport MD, and DSRES scores were positively related to peer pressure SRE (Sample 1 $r = .41, p$
10 $< .01$; Sample 2 $r = .56, p < .01$). Thus, overall the findings support the convergent and
11 discriminant validity of DMDS, DMDS-S, and DSRES scores.

12 **Internal consistency.** Cronbach alpha values showed internal consistency to be either
13 good or very good for all subscales of the DMDS in both samples (see Table 2). Alpha values for
14 overall doping MD for the DMDS were excellent in both samples (i.e., Sample 1 = .95; Sample 2
15 = .96). Similarly, alpha values were very good for the DMDS-S in both samples (i.e., Sample 1
16 = .86; Sample 2 = .89) and excellent for the DSRES in both samples (i.e., Sample 1 = .93;
17 Sample 2 = .94). Thus, overall the findings support the internal consistency reliability of DMDS,
18 DMDS-S, and DSRES scores.

19 Study 2

20 Method

21 **Participants.** Participants were team- (e.g., netball, soccer) ($n = 9$) or individual- (e.g.,
22 athletics, triathlon) ($n = 78$) sport or bodybuilding- ($n = 5$) or corporate- ($n = 9$) gym exercisers
23 ($n_{male} = 60$; $n_{female} = 41$); ages ranged from 16 to 70 years ($M = 35.2, SD = 13.5$). Sport

1 participants competed at a recreational ($n = 25$), local ($n = 21$), university ($n = 7$), regional ($n =$
2 14), national ($n = 9$), or international ($n = 10$) level (15 didn't report). Participants had been
3 training/competing for an average of 9.4 years ($SD = 7.4$), spent an average of 7.9 hours ($SD =$
4 4.8) per week training, and had trained in/with their current gym/team for an average of 4.4 years
5 ($SD = 5.0$)¹⁰.

6 **Measures.**

7 ***Doping Moral Disengagement.*** Doping moral disengagement was measured using the
8 18-item DMDS developed in Study 1.

9 ***Doping Self-Regulatory Efficacy.*** Doping self-regulatory efficacy was measured using
10 the 6-item DMDS developed in Study 1.

11 **Procedures.** Test-retest reliability represents consistency of scores across time in a
12 population with stable scores on the construct being assessed (Lohr, 2002). It can be examined
13 by administering a measure to the same sample twice and examining the relation between the
14 two scores (Pedhazur & Schmelkin, 1991). Given both doping MD and doping SRE should be
15 stable over the short term, evidence of score reliability and stability would support the temporal
16 reproducibility of scores obtained using the DMDS, DMDS-S, and DSRES. To determine this,
17 doping MD and doping SRE data were collected on two occasions using paper and online
18 versions of the new instruments. Recruitment procedures for face-to-face data collections at
19 Time 1 matched procedures used in Study 1. Recruitment of online participants involved
20 advertising the study through sport-club and gymnasium websites, and discussion groups, social
21 media, and personal contacts. Adverts included the basic information on the study, as well as a

¹⁰ Prevalence of doping was not assessed in Sample 3. This was because some respondents who acknowledged doping at Time 1 may subsequently have been reluctant to participate again at Time 2. Further, prevalence of doping was not central to the aims of this phase, as Study 3 data were only used to determine temporal stability of instrument scores and not for instrument development.

1 link to the survey webpage; the webpage provided full details of the study. Prior to starting the
2 online questionnaire, participants were informed that by clicking on the link to start the
3 questionnaire they were providing informed consent to participate. For both face-to-face and
4 online collections, participants were informed that participation involved providing data on two
5 occasions, and that data provided on the two occasions would be linked.

6 Relatively short between-administration intervals are recommended for examining test-
7 retest reliability (Pedhazur & Schmelkin, 1991). Such intervals ensure any differences in scores
8 are largely due to random measurement error rather than changes in participants' scores on the
9 assessed constructs. As such, we used an inter-administration interval of nine to 16 days. For
10 face-to-face participants, arrangements were made to collect data again within this window. For
11 online participants, reminder emails were sent nine days following initial completion, and
12 continued each day until day 16. At which point the online questionnaire closed for participants
13 who had not completed their second administration. Of the 101 participants who completed the
14 instruments on two occasions, 87 were online participants. Overall, 50% of those who
15 participated on one occasion completed the instruments again on a second occasion within the
16 allotted timeframe (online = 63%; face-to-face = 22%¹¹).

17 **Results**

18 Intraclass correlation coefficients for overall scores obtained using the DMDS, DMDS-S
19 and DSRES were .94, .93 and .87, respectively, and ranged from .87 to .93 ($M = .91$) for the
20 individual DMDS subscales. Overall, the findings demonstrate very good to excellent levels of
21 score reliability for all three instruments. We also investigated individual-item stability in
22 accordance with the recommendations of Nevill, Lane, Kilgour, Bowes, and Whyte (2001; see

¹¹ The low retention rate for face-to-face collections was largely due to non-attendance at the training session where the second collection took place, as opposed to athletes choosing not to participate.

1 Table 7). For each item, we calculated test-retest differences and reported the minimum and
2 maximum difference scores. In addition, we calculated the percentage of participants whose test-
3 retest difference scores were within ± 1 , termed proportion of agreement. For 5-point scales (i.e.,
4 DSRES items) acceptable levels of score stability are evidenced by proportion of agreement
5 scores of $\geq 90\%$ (Nevill et al., 2001). Although such criterion values are not available for 7-point
6 scales (i.e., DMDS and DMDS-S items), past research applying this method with such scales has
7 shown proportion of agreement scores significantly below 90% (e.g., Uphill, Lane, & Jones,
8 2012). This is logical given the greater potential for scores to fluctuate between administrations
9 with 7-point scales compared to 5-point scales. For DMDS items, for 11 items at least 90% of
10 participants had test-retest difference scores within ± 1 , for six items 80-90% of participants had
11 test-retest difference scores within ± 1 , whereas for one item the percentage in this range was just
12 below 80%. For DMDS-S items, for four items at least 90% of participants had test-retest
13 difference scores within ± 1 , whereas for two items 80-90% of participants had test-retest
14 difference scores in this range. For DSRES items, for all 6 items at least 90% of participants had
15 test-retest difference scores within ± 1 . Finally, we computed the median-sign test (i.e., the
16 number of participants with test-retest differences above and below zero) to ascertain the amount
17 of possible bias in scores increasing or decreasing from test to retest. The median-sign test
18 indicated the scores of five DMDS items, two DMDS-S items, and one DSRES item decreased
19 significantly between test and retest.

20 **Discussion**

21 Research has highlighted the potential importance of doping MD and doping SRE to
22 doping in sport and exercise (e.g., Boardley & Grix, 2014; Boardley et al., 2014, 2015; Lucidi et
23 al., 2004, 2008). However, valid instruments are needed to measure these constructs and

1 substantively advance the knowledge base. The relevant extant psychological assessments tied to
2 doping have shortcomings, and there is currently no multidimensional measure of doping MD
3 available nor any instrument to assess doping MD in exercise populations. Therefore, we sought
4 to develop psychometrically sound instruments assessing doping MD and doping SRE, pursuing
5 item development and assessment of score validity and reliability using an expert panel, two
6 pilot samples, and three primary samples.

7 Bandura (1991) described eight mechanisms of MD and research in sport and exercise
8 contexts has shown six of these mechanisms to be utilized to rationalize and justify doping
9 (Boardley & Grix, 2014; Boardley et al., 2014, 2015). Consequently, we developed items for
10 these six mechanisms and expected scores from the final instrument to evidence six lower-order
11 factors. Consistent with this expectation, results from both samples suggested doping MD – as
12 assessed by the DMDS – incorporates six lower-order dimensions. This is consistent with the
13 only other multidimensional measure of MD developed for use in a sport context – the MDSS
14 (Boardley & Kavussanu, 2007) – which also has six lower-order factors. However, it is
15 important to acknowledge the six dimensions in the DMDS represent six MD mechanisms,
16 whereas the six dimensions in the MDSS represent eight mechanisms.

17 Convergent validity of overall DMDS scores was evidenced by the strong negative
18 correlation between doping MD and anticipated guilt, and the moderate negative correlation
19 between doping MD and empathy. These associations are consistent with theory and research
20 (Bandura, 1991; Bandura et al., 1996). Evidence for the convergent validity of scores for most
21 DMDS subscales was also provided, with relationships of the subscales with anticipated guilt
22 and empathy generally consistent with those obtained using all 18 instrument items. However,
23 evidence for the convergent validity of scores obtained for euphemistic labelling was weaker

Commented [IB5]: Comment 8

1 than for scores obtained with the other five subscales. Because use of euphemistic terms
2 regarding doping (e.g., gear, juice, etc.) is a key aspect of doping culture (see Andrews, Sudwell,
3 & Sparkes, 2005), people may at times use such terminology to fit in with this culture and not
4 exclusively for making doping appear less harmful.

5 Discriminant validity of the DMDS was evidenced through the associations of DMDS
6 and MDSS-S scores, which showed the DMDS scores to be related to – but distinct from –
7 MDSS-S scores. This supports the context-specific nature of MD proposed by Bandura (1991).
8 Evidence for discriminant validity was also provided internally by the strength of the
9 associations amongst the DMDS subscale scores. Euphemistic labelling had weaker relationships
10 with the other five mechanisms than those mechanisms did amongst themselves, suggesting
11 discriminant validity was highest for this mechanism. In contrast, the very strong correlations
12 between the remaining mechanisms demonstrated substantial redundancy. The highest
13 redundancy was observed between displacement and diffusion of responsibility and between
14 advantageous comparison and distortion of consequences. These findings are not out of line with
15 past research showing similar levels of convergence between these mechanisms (Boardley &
16 Kavussanu, 2007), and the six-factor DMDS model was superior to alternative models.

17 We also developed a short version of the DMDS termed the DMDS-S, enabling a concise
18 measure of overall doping MD. Following procedures used successfully to develop previous
19 short versions of MD instruments (e.g., Boardley & Kavussanu, 2008), we found evidence for
20 the factorial validity of DMDS-S scores, as well as discriminant and convergent validity.

21 Specifically, CFA supported the unidimensionality of scale scores, negative correlations with
22 empathy and guilt evidenced convergent validity, and strong – but not overly strong – positive
23 relations with MDSS-S scores evinced discriminant validity. Importantly, the instrument has

Commented [b6]: Comment 9.

1 items representing all six of the relevant mechanisms of MD (see Boardley & Grix, 2014;
2 Boardley et al., 2014, 2015), ensuring that the DMDS-S generates doping MD scores that are
3 equally representative of each of the six mechanisms. Although for us the DMDS is the preferred
4 option for most research given it is representative of underlying theory (i.e., Bandura, 1991), the
5 development of the DMDS-S provides a generalized and brief assessment suitable for use when
6 the DMDS is not practical to administer.

7 Another important context-specific variable from Bandura's (1991) theory is doping
8 SRE. Although past research had identified the potential importance of this variable to doping
9 (e.g., Lucidi et al., 2008), prior to the current research scores of doping SRE measures had not
10 been appropriately validated. Based on theory (Bandura, 1991) and the factorial structure of
11 existing instruments assessing SRE (e.g., Bandura et al., 1996), the DSRES was constructed and
12 scores demonstrated the expected unidimensional structure. Moreover, **positive** associations with
13 empathy and guilt supported the convergent validity of DSRES scores, while **positive moderate-**
14 **to-strong (i.e., Sample 1) and strong (i.e., Sample 2), yet empirically distinguishable,** associations
15 with peer-pressure SRE supported the discriminant validity of DSRES scores.

16 We additionally assessed internal consistency and test-retest reliability of scores obtained
17 using the three new instruments. The internal consistency of all higher- and lower-order scales
18 surpassed the minimum criterion level recommended when developing new instruments (i.e.,
19 0.80; Clark & Watson, 1995). Also, test-retest reliability levels across a nine- to 16-day period,
20 using a separate sample, were good to excellent. Thus, scores obtained using the DMDS, DMDS-
21 S and DSRES are reliable over the short term. On the whole, item stability analyses suggest the
22 majority of items in the three scales are stable across this period too. These findings suggest that
23 scores obtained using the three **instruments** are relatively stable over the short term, and that

Commented [IB7]: Comment 10

Commented [SA8]: Added this to align w/ how we describe discriminant validity evidence early in the paper. The idea is that correlations are strong, but not too strong.

Commented [IB9]: Comment 10

Commented [b10]: Comment 11.

1 situational factors may not impart much error into scale responses. Overall, across all samples
2 and analyses, the instruments performed well in measuring their target constructs.

3 Beyond the main validity and reliability analyses, we also utilized multisample analyses
4 to examine the measurement invariance of scores from all three instruments. In each case, we
5 performed two sets of multisample analyses. The first examined measurement invariance by sex
6 whereas the second examined measurement invariance across four sport/exercise groups. All
7 three instruments showed configural invariance, meaning across all groups the same subsets of
8 items are associated with the same constructs (Cheung & Rensvold, 2002). In all but one case,
9 complete metric invariance was established. This demonstrates the strength of the relationship
10 between all items and their underlying constructs was the same (Cheung & Rensvold, 2002). The
11 one exception was across the four sport/exercise groups with the DMDS-S, where partial metric
12 invariance was evidenced (Byrne et al., 1989). This is present when most items for a given latent
13 variable have loadings that are invariant across groups. If this is the case, cross-group
14 comparisons can still safely be made (Reise, Widaman, & Pugh, 1993). Thus, the DMDS,
15 DMDS-S and DSRES are suitable for research testing substantive hypotheses regarding group
16 differences by sex and among the four sport/exercise types tested.

17 Equivalence of construct variance exists when the range of responses given to each item
18 is the same across groups, whereas equivalence of construct covariance is apparent when
19 structural relationships are equivalent across groups (Cheung & Rensvold, 2002). As with our
20 analyses, these two forms of invariance are often tested simultaneously (see Byrne, 2006). In
21 contrast to the other forms of equivalence tested, our analyses showed ECVC was not evident for
22 any of the instruments by sex or across the four sport/exercise groups. This suggests the range of
23 item responses and/or the strength of the relations between MDMS subscales varied among these

1 groups. Importantly, prevalence of doping differed markedly between genders and across
2 sport/exercise context. Given that levels of doping MD and doping SRE are likely to be more
3 extreme in athletes with experience of doping, these differences in prevalence may have led to
4 greater construct variance in samples with higher prevalence rates. In addition, the degree of
5 association between MD subscales may be influenced by doping prevalence. Future research
6 should address what may lead to differences in construct variance and covariance across groups.

7 **Limitations and Future Directions**

8 The current project developed three psychological assessments relevant to doping in sport
9 and exercise that each provide scores with good psychometric properties. Nevertheless,
10 limitations relating to certain aspects of the research should be acknowledged. First, whilst we
11 achieved our aim of sampling from populations showing doping prevalence rates in line with
12 existing estimates (see Momaya et al., 2015), the use of self-report to assess doping prevalence
13 has known limitations when applied to socially sensitive behaviors such as PIED use.
14 Specifically, assessing doping through self-report is thought to underestimate the true prevalence
15 of doping. Accordingly, the prevalence of doping in our samples may be higher than reported.
16 Next, although the test-retest analyses presented evidence for fairly good levels of short-term
17 reliability and stability of DMDS, DMDS-S, and DSRES scores, the sample for these analyses
18 largely consisted of individual-sport athletes. Future research should assess test-retest reliability
19 with samples more representative of the other three sub-populations. Also, difficulties in
20 accessing the same participants on two occasions across the stipulated period resulted in a low
21 percentage of athletes completing the measure twice, especially for face-to-face collections.
22 These difficulties also required us to allow some flexibility on the period between test and retest

1 completions, when ideally this would have been standardized across participants. In future work
2 attempts should be made to increase retention rates and standardize the test-retest period.

3 Validation is a continuous process (Clark & Watson, 1995) and further aspects of validity
4 remain to be examined, such as predictive validity over time and associations with key variables
5 not assessed here. For example, associations of test scores with socially situated emotions such
6 as shame could be examined. Further, validity and reliability of test scores in other doping
7 populations and cultures require investigation. For instance, researchers could examine their
8 measurement invariance across different categories of bodybuilders (see Christiansen, Vinther, &
9 Liokaftos, 2017). Finally, the multidimensional nature of the DMDS – and the divergent
10 associations of its subscales with anticipated guilt – present the opportunity for research
11 determining which MD mechanisms have the greatest potential to facilitate doping.

12 **Conclusion**

13 Through a rigorous set of processes, we developed three psychometric instruments
14 relevant to the psychology of doping, and supported the validity of scores obtained with them.
15 The final versions of the DMDS, DMDS-S, and DSRES are found in the Appendix. Evidence for
16 the construct validity, internal consistency and test-retest reliability of scores for all three
17 instruments was provided. Items for the instruments have high levels of face and content validity,
18 making them particularly suitable for use with athletes and exercisers who have experience in
19 environments where doping is a salient aspect of the culture. The DMDS makes a particular
20 contribution to the literature, being the first measure capable of capturing the individual doping
21 MD mechanisms, and for use with exercise populations. The DSRES is also the first doping-
22 contextualized instrument developed for use with exercise populations. We look forward to
23 seeing these instruments employed and further evaluated in future research.

Commented [b11]: Comment 12.

References

- 1
2 Andrews, G. J., Sudwell, M. I., & Sparkes, A. C. (2005). Towards a geography of fitness: An
3 ethnographic case study of the gym in British bodybuilding culture. *Social Science &*
4 *Medicine, 60*(4), 877–891. DOI: 10.1016/j.socscimed.2004.06.029
- 5 Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modelling in practice: A review
6 and recommended two-step approach. *Psychological Bulletin, 103*, 411-423. DOI:
7 10.1037/0033-2909.103.3.411
- 8 Bandura, A. (1991). Social cognitive theory of moral thought and action. In W.M. Kurtines & J.
9 L. Gewirtz (Eds.), *Handbook of moral behavior and development: Theory, research, and*
10 *applications* (Vol. 1, pp. 71-129). Hillsdale, NJ: Erlbaum.
- 11 Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral
12 disengagement in the exercise of moral agency. *Journal of Personality and Social*
13 *Psychology, 71*, 364-374. DOI: 10.1037/0022-3514.71.2.364
- 14 Bandura, A., Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Regalia, C. (2001).
15 Sociocognitive self-regulatory mechanisms governing transgressive behavior. *Journal of*
16 *Personality and Social Psychology, 80*, 125-135. DOI: 10.1037/0022-3514.80.1.125
- 17 Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide*. Multivariate Software.
- 18 Boardley, I. D., & Grix, J. (2014). Doping in bodybuilders: A qualitative investigation of
19 facilitative psychosocial processes. *Qualitative Research in Sport, Exercise and Health, 6*,
20 422-439. DOI: 10.1080/2159676X.2013.766809
- 21 Boardley, I. D., Grix, J., & Dewar, A. J. (2014). Moral disengagement and associated processes
22 in performance-enhancing drug use: a national qualitative investigation. *Journal of Sports*
23 *Sciences, 32*, 836-844. DOI: 10.1080/02640414.2013.862842

- 1 Boardley, I. D., Grix, J., & Harkin, J. (2015). Doping in team and individual sports: A qualitative
2 investigation of moral disengagement and associated processes. *Qualitative Research in*
3 *Sport, Exercise and Health*, 7, 698-717. DOI: 10.1080/2159676X.2014.992039
- 4 Boardley, I. D., & Kavussanu, M. (2007). Development and validation of the Moral
5 Disengagement in Sport Scale. *Journal of Sport & Exercise Psychology*, 29, 608-628. DOI:
6 10.1123/jsep.29.5.608
- 7 Boardley, I. D., & Kavussanu, M. (2008). The Moral Disengagement in Sport Scale-Short.
8 *Journal of Sports Sciences*, 26, 1507-1517. DOI: 10.1080/02640410802315054
- 9 Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and*
10 *programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- 11 Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor
12 covariance and mean structures: The issue of partial measurement invariance.
13 *Psychological Bulletin*, 105, 456-466. DOI: 10.1037/0033-2909.105.3.456
- 14 Carlo, G., Raffaelli, M., Laible, D. J., & Meyer, K. A. (1999). Why are girls less physically
15 aggressive than boys? Personality and parenting mediators of physical aggression. *Sex*
16 *Roles*, 40, 711-729. DOI: 10.1023/A:1018856601513
- 17 Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing
18 measurement invariance. *Structural Equation Modeling*, 9, 233-255. DOI:
19 10.1207/S15328007SEM0902_5
- 20 Christiansen, A. V., Vinther, A. S., & Liokaftos, D. (2017). Outline of a typology of men's use
21 of anabolic androgenic steroids in fitness and strength training environments. *Drugs:*
22 *Education, Prevention and Policy*, 24, 295-305. DOI: 10.1080/09687637.2016.1231173

- 1 Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale
2 development. *Psychological Assessment*, 7(3), 309-319. DOI: 10.1037/1040-3590.7.3.309
- 3 Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- 4 Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a
5 multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
6 DOI: 10.1037/0022-3514.44.1.113
- 7 Donovan, R. J., Egger, G., Kapernick, V., & Mendoza, J. (2002). A conceptual framework for
8 achieving performance enhancing drug compliance in sport. *Sports Medicine*, 32, 269-284.
9 DOI: 10.2165/00007256-200232040-00005
- 10 Dunn, J. G., Bouffard, M., & Rogers, W. T. (1999). Assessing item content-relevance in sport
11 psychology scale-construction research: Issues and recommendations. *Measurement in*
12 *Physical Education and Exercise Science*, 3, 15-36.
- 13 Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor
14 analysis? Some decision rules. *Psychological Bulletin*, 81, 358-361.
- 15 Enders, C.K. (2006). A primer on the use of modern missing-data methods in psychosomatic
16 medicine research. *Psychosomatic Medicine*, 68, 427-436. DOI:
17 10.1097/01.psy.0000221275.75056.d8
- 18 Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use
19 of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-
20 299. DOI: 10.1037/1082-989X.4.3.272
- 21 Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. A. Berk, (Ed.),
22 *Criterion-referenced measurement: The state of the art* (pp. 80-123). Baltimore, MD: The
23 Johns Hopkins Press.

- 1 Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological
2 assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7,
3 238-247. DOI: 10.1037/1040-3590.7.3.238
- 4 Hodge, K., Hargreaves, E. A., Gerrard, D., & Lonsdale, C. (2013). Psychological mechanisms
5 underlying doping attitudes in sport: Motivation and moral disengagement. *Journal of*
6 *Sport and Exercise Psychology*, 35, 419-432. DOI: 10.1123/jsep.35.4.419
- 7 Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*.
8 New York, NY: Cambridge University Press. DOI: 10.1017/CBO9780511805851
- 9 Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:
10 Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. DOI:
11 10.1080/10705519909540118
- 12 Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological*
13 *Measurement*, 34, 111-117.
- 14 Kavussanu, M., Hatzigeorgiadis, A., Elbe, A. M., & Ring, C. (2016). The moral disengagement
15 in doping scale. *Psychology of Sport and Exercise*, 24, 188-198. DOI:
16 10.1016/j.psychsport.2016.02.003
- 17 Kavussanu, M., Stamp, R., Slade, G., & Ring, C. (2009). Observed prosocial and antisocial
18 behaviors in male and female soccer players. *Journal of Applied Sport Psychology*, 21,
19 S62-S76. DOI: 10.1080/10413200802624292
- 20 Kavussanu, M., Stanger, N., & Boardley, I. D. (2013). The Prosocial and Antisocial Behaviour in
21 Sport Scale: Further evidence for construct validity and reliability, *Journal of Sports*
22 *Sciences*, 31, 11, 1208-1221, DOI: 10.1080/02640414.2013.775473

- 1 Kline, R. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York:
2 The Guilford Press.
- 3 Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: Attributes and review
4 criteria. *Quality of Life Research, 11*, 193-205. DOI: 10.1023/A:1015291021312
- 5 Lucidi, F., Grano, C., Leone, L., Lombardo, C., & Pesce, C. (2004). Determinants of the
6 intention to use doping substances. *International Journal of Sport Psychology, 35*, 133-
7 148.
- 8 Lucidi, F., Zelli, A., Mallia, L., Grano, C., Russo, P. M., & Violani, C. (2008). The social-
9 cognitive mechanisms regulating adolescents' use of doping substances. *Journal of Sports*
10 *Sciences, 26*, 447-456. DOI: 10.1080/02640410701579370
- 11 Lucidi, F., Zelli, A., & Mallia, L. (2013). The contribution of moral disengagement to
12 adolescents' use of doping substances. *International Journal of Sport Psychology, 44*, 493-
13 514.
- 14 Mallia, L., Lazuras, L., Barkoukis, et al. (2016). Doping use in sport teams: The development
15 and validation of measures of team-based efficacy beliefs and moral disengagement from a
16 cross-national perspective. *Psychology of Sport and Exercise, 25*, 78-88. DOI:
17 10.1016/j.psychsport.2016.04.005
- 18 Marschall, D. E., Saftner, J., & Tangney, J. P. (1994). *The State Shame and Guilt Scale*. Fairfax,
19 VA: George Mason University.
- 20 Marsh, H.W. (1987). The hierarchical structure of self-concept and the application of
21 hierarchical confirmatory factor analysis. *Journal of Educational Measurement, 24*, 17-39.
22 DOI: 10.1111/j.1745-3984.1987.tb00259.x

- 1 McVeigh, J. & Begley, E. (in press). Anabolic steroids in the UK: An increasing issue for public
2 health. *Drugs: Education, Prevention and Policy*. DOI: 10.1080/09687637.2016.1245713.
- 3 Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons'
4 responses and performances as scientific inquiry into score meaning. *American*
5 *Psychologist*, 50, 741.
- 6 Momaya, A., Fawal, M., & Estes, R. (2015). Performance-enhancing substances in sports: A
7 review of the literature. *Sports Medicine*, 45, 517-531. DOI: 10.1007/s40279-015-0308-9
- 8 Nevill, A. M., Lane, A. M., Kilgour, L. J., Bowes, N., & Whyte, G. P. (2001). Stability of
9 psychometric questionnaires. *Journal of Sports Sciences*, 19, 273-278. DOI:
10 10.1080/026404101750158358
- 11 Olrich, T. W., & Ewing, M. E. (1999). Life on steroids: bodybuilders describe their perceptions
12 of the anabolic-androgenic steroid use period. *The Sport Psychologist*, 13, 299-312. DOI:
13 10.1123/tsp.13.3.299
- 14 Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the
15 execution process. *Law and Human Behavior*, 29, 371-393. DOI: 10.1007/s10979-005-
16 4930-1
- 17 Pedhazur, E. J., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated*
18 *approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 19 Petróczy, A., & Aidman, E. (2008). Psychological drivers in doping: the life-cycle model of
20 performance enhancement. *Substance Abuse Treatment, Prevention, and Policy*, 3, 7. DOI:
21 10.1186/1747-597X-3-7

- 1 Pope, H. G., Wood, R. I., Rogol, A., Nyberg, F., Bowers, L., & Bhasin, S. (2013). Adverse
2 health consequences of performance-enhancing drugs: An Endocrine Society scientific
3 statement. *Endocrine Reviews*, *35*, 341-375. DOI: 10.1210/er.2013-1058.
- 4 Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales:
5 reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*,
6 *104*, 1-15.
- 7 Probert, A., Palmer, F., & Leberman, S. (2007). The fine line: an insight into 'risky' practices of
8 male and female competitive bodybuilders. *Annals of Leisure Research*, *10*, 272-290. DOI:
9 10.1080/11745398.2007.9686767
- 10 Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item
11 response theory: Two approaches for exploring measurement invariance. *Psychological*
12 *Bulletin*, *114*, 552-566. DOI: 10.1037/0033-2909.114.3.552
- 13 Ring, C., & Kavussanu, M. (in press). The role of self-regulatory efficacy, moral disengagement
14 and guilt on doping likelihood: A social cognitive theory perspective. *Journal of Sports*
15 *Sciences*. DOI: 10.1080/02640414.2017.1324206
- 16 Sjöqvist, F., Garle, M., & Rane, A. (2008). Use of doping agents, particularly anabolic steroids,
17 in sports and society. *The Lancet*, *371*, 1872-1882. DOI: 10.1016/S0140-6736(08)60801-6
- 18 Stanger, N., Kavussanu, M., Boardley, I. D., & Ring, C. (2013). The influence of moral
19 disengagement and negative emotion on antisocial sport behavior. *Sport, Exercise, and*
20 *Performance Psychology*, *2*, 117-129. DOI: 10.1037/a0030585
- 21 Uphill, M. A., Lane, A. M., & Jones, M. V. (2012). Emotion Regulation Questionnaire for use
22 with athletes. *Psychology of Sport and Exercise*, *13*, 761-770. DOI:
23 10.1016/j.psychsport.2012.05.001

- 1 Vaughn, B. K., & Daniel, S. R. (2012). Conceptualizing validity. In G. Tenenbaum, R. C.
- 2 Eklund, & A. Kamata (Eds.), *Measurement in sport and exercise psychology* (pp. 33-40).
- 3 Champaign, IL: Human Kinetics.
- 4 Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using
- 5 short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R.
- 6 E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists*, (pp. 39-61).
- 7 Washington, DC: American Psychological Association. DOI: 10.1037/12350-003

Table 1

Summary of Fit Indices for All CFA Models Tested During Development of the Doping Moral Disengagement Scale (DMDS), the Doping Moral Disengagement Scale – Short (DMDS-S) and the Doping Self-Regulatory Efficacy Scale (DSRES)

Model	<i>df</i>	$R\chi^2$	RCFI	SRMR	RMSEA
<i>Sample 1 (N = 318)</i>					
<i>DMDS Models</i>					
1. M1a, 35 items	545	1183.39	.873	.050	.061
2. M1b, 18 items	120	222.54	.955	.034	.052
3. 2 nd order, M1b	129	333.48	.910	.052	.071
<i>Alternative DMDS Models</i>					
4. M2, 18 items	132	667.24	.765	.079	.113
5. M3, 18 items	135	779.76	.717	.088	.123
6. M4, 18 items	129	303.92	.923	.043	.065
<i>DMDS-S Models</i>					
7. M5a, 6 items	9	21.35	.968	.031	.066

DOPING IN SPORT AND EXERCISE

45

8. M5b, 6 items	8	7.38 (ns)	1.000	.020	.000
-----------------	---	-----------	-------	------	------

DSRES Models

9. M6a, 13 items	65	210.48	.871	.060	.084
------------------	----	--------	------	------	------

10. M6b, 6 items	9	12.17 (ns)	.991	.028	.033
------------------	---	------------	------	------	------

Sample 2 (N = 300)

DMDS Models

11. M1b, 18 items	120	248.14	.974	.032	.061
-------------------	-----	--------	------	------	------

12. 2 nd order, M1b	129	272.04	.953	.046	.061
--------------------------------	-----	--------	------	------	------

DMDS-S Model

13. M5b, 6 items	8	15.83	.987	.023	.058
------------------	---	-------	------	------	------

DSRES Model

14. M6b, 6 items	9	12.44 (ns)	.991	.020	.036
------------------	---	------------	------	------	------

Note. *df* = degrees of freedom; $R\chi^2$ = Satorra–Bentler scaled chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; ns = $p > .05$. M1 = 6-factor DMDS model; M2 = 3-factor DMDS model; M3 = 1-factor DMDS model, M4 = 4-factor DMDS model; M5 = 1-factor DMDS-S model; M6 = 1-factor DSRES model.

Table 2

CFA Factor Correlations for the Doping Moral Disengagement Scale (DMDS) Subscales in Sample 1 ($N = 318$) and Sample 2 ($N = 300$)

Factor	1	2	3	4	5	6
1. Moral Justification	.86/.93	.55	.76	.88	.86	.83
2. Euphemistic Labelling	.51	.86/.90	.55	.57	.51	.49
3. Advantageous Comparison	.80	.54	.82/.87	.74	.81	.92
4. Displacement of Responsibility	.88	.48	.72	.91/.95	.90	.76
5. Diffusion of Responsibility	.84	.46	.79	.91	.88/.90	.84
6. Distortion of Consequences	.84	.44	.93	.72	.78	.83/.86

Note. Sample 1 correlations are below the diagonal and those from Sample 2 are above. Alpha coefficients in Sample 1 / Sample 2 are presented on the diagonal. For all correlations, $p < .01$.

Table 3

Items, Standardized Factor Loadings and Error Variances for the Doping Moral Disengagement Scale (DMDS)

Factor Item	Factor Loading	Error Variance
<i>Moral Justification</i>		
1. It is okay to dope if it helps an athlete to provide for his/her family.	.76/.82	.65/.57
2. Doping is okay if it helps an athlete advise others on how to do it right.	.87/.95	.49/.32
3. It is acceptable to dope if knowledge gained helps an athlete advise others on safe doping.	.87/.96	.49/.30
<i>Euphemistic Labelling</i>		
4. Saying you "take steroids" feels worse than saying you "use some gear".	.67/.75	.75/.66
5. Using words like "roids", "gear" and "pinning" makes doping feel more acceptable.	.93/.93	.37/.36
6. Using terms such as "gear" or "juice" makes doping sound less harmful.	.91/.93	.41/.37
<i>Advantageous Comparison</i>		
7. Compared to most lifestyles in the general public, doping isn't that bad.	.82/.86	.57/.51
8. Compared to smoking, doping is pretty safe.	.79/.87	.62/.49
9. Compared to physical violence, doping isn't that serious.	.75/.76	.66/.65
<i>Displacement of Responsibility</i>		

10. Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	.87/.92	.50/.39
11. An athlete shouldn't be blamed for doping if a member of his/her training group has encouraged it.	.92/.94	.39/.33
12. An athlete shouldn't be held responsible for doping if his/her coach encouraged him/her to do it.	.88/.92	.47/.39
<hr/> <i>Diffusion of Responsibility</i>		
13. If most athletes in a sport dope, no one athlete should be held responsible for doing it.	.77/.81	.64/.59
14. It's not right to condemn individuals who dope when many in their sport are doing the same.	.87/.92	.50/.38
15. If an athlete trains/competes in an environment in which doping is the norm, he/she shouldn't be held accountable for doing it.	.90/.87	.43/.50
<hr/> <i>Distortion of Consequences</i>		
16. Risks associated with doping are exaggerated.	.85/.86	.53/.52
17. Doping doesn't really harm anyone else.	.70/.77	.72/.64
18. The negative aspects of doping are exaggerated by the media.	.82/.85	.58/.53

Note. Factor Loadings and Error Variances are presented as follows: Sample 1 / Sample 2.

Table 4

Items, Standardized Factor Loadings and Error Variances for the Doping Moral Disengagement Scale – Short (DMDS-S) and Doping Self-Regulatory Efficacy Scale (DSRES)

Item (<i>mechanism</i>)	Factor Loading	Error Variance
DMDS-S		
Doping is okay if it helps an athlete advise others on how to do it right (<i>moral justification</i>).	.61/.86	.79/.52
Using terms such as "gear" or "juice" makes doping sound less harmful (<i>euphemistic labelling</i>).	.44/.52	.90/.86
Compared to most lifestyles in the general public, doping isn't that bad (<i>advantageous comparison</i>).	.70/.76	.72/.65
Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it (<i>displacement of responsibility</i>).	.76/.88	.65/.48
It's not right to condemn individuals who dope when many in their sport are doing the same (<i>diffusion of responsibility</i>).	.83/.88	.56/.48
Risks associated with doping are exaggerated (<i>distortion of consequences</i>).	.67/.74	.74/.68
DSRES		
1. ...resist doping even if your training group encouraged you to do it?	.84/.83	.55/.56
2. ...resist doping even if you knew you could get away with it?	.82/.83	.57/.55
3. ...ignore the temptation to dope even if you knew it would improve your performance?	.86/.88	.51/.48
4. ...resist peer pressure to dope?	.80/.87	.60/.50
5. ...reject doping even if most of your training partners did it?	.84/.87	.54/.50
6. ...ignore the temptation to dope when feeling down physically?	.81/.77	.59/.63

Note. Factor Loadings and Error Variances are presented as follows: Sample 1 / Sample 2.

Table 5

Fit Indices for multisample analyses on the Doping Moral Disengagement Scale (DMDS), Doping Moral Disengagement Scale-Short (DMDS-S) and Doping Self-Regulatory Efficacy Scale (DSRES)

Model	<i>df</i>	$R\chi^2$	RCFI	SRMR	RMSEA
Sex DMDS					
Baseline Male	120	245.22	.968	.040	.053
Baseline Female	120	214.90	.922	.040	.057
Configural Invariance	240	461.00	.950	.040	.055
Metric Invariance	252	470.88	.950	.044	.053
ECVC	273	567.21	.933	.223	.059
Sex DMDS-S					
Baseline Male	8	21.48	.981	.028	.067
Baseline Female	8	12.96(ns)	.978	.023	.051
Configural Invariance	16	34.69	.977	.025	.062
Metric Invariance	21	41.43	.975	.049	.056
ECVC	22	62.26	.951	.225	.077
Sex DSRES					
Baseline Male	9	13.43(ns)	.992	.021	.036
Baseline Female	9	13.60(ns)	.981	.036	.046
Configural Invariance	18	27.04(ns)	.989	.029	.040
Metric Invariance	23	34.04(ns)	.986	.054	.039
ECVC	24	48.74	.969	.398	.058
Sport/Exercise Group DMDS					
Baseline Corporate	120	166.66	.925	.047	.057
Baseline Bodybuilding	120	209.66	.970	.050	.075
Baseline Team	120	172.91	.948	.038	.048

Baseline Individual	120	157.59	.966	.060	.043
Configural Invariance	480	703.75	.951	.050	.055
Metric Invariance	516	760.55	.946	.071	.056
ECVC	579	911.77	.926	.282	.061
Sport/Exercise Group DMDS-S					
Baseline Corporate	8	11.00 (ns)	.978	.028	.056
Baseline Bodybuilding	8	8.27 (ns)	1.000	.023	.016
Baseline Team	8	10.68 (ns)	.985	.035	.042
Baseline Individual	8	9.25 (ns)	.992	.035	.030
Configural Invariance	32	39.67 (ns)	.991	.030	.040
Metric Invariance	47	78.85	.961	.093	.066
Metric Invariance Revised	45	56.66 (ns)	.986	.076	.041
ECVC	48	121.78	.910	.342	.100
Sport/Exercise Group DSRES					
Baseline Corporate	9	12.38 (ns)	.970	.050	.056
Baseline Bodybuilding	9	14.53 (ns)	.985	.023	.068
Baseline Team	9	11.75 (ns)	.989	.042	.040
Baseline Individual	9	9.16 (ns)	.998	.028	.010
Configural Invariance	36	47.06 (ns)	.983	.037	.045
Metric Invariance	51	66.53 (ns)	.976	.088	.045
ECVC	54	77.07	.964	.242	.053

Note. df = degrees of freedom; $R\chi^2$ = Satorra–Bentler scaled chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; ns = $p > .05$.

Table 6

Correlations of the Doping Moral Disengagement Scale (DMDS), Doping Moral Disengagement Scale-Short (DMDS-S), and Doping Self-Regulatory Efficacy Scale (DSRES) Scores and DMDS Subscale Scores with Empathy and Anticipated Guilt

	Sample 1 (<i>N</i> = 318)		Sample 2 (<i>N</i> = 300)	
	Empathy	Guilt	Empathy	Guilt
DMDS	-.28	-.68	-.32	-.60
DMDS-S	-.26	-.66	-.33	-.59
DSRES	.15	.38	.36	.48
DMDS Subscales				
Moral Justification	-.24	-.67	-.34	-.58
Euphemistic Labeling	-.18	-.30	-.10 (ns)	-.24
Advantageous Comparison	-.25	-.56	-.25	-.54
Displacement of Responsibility	-.21	-.60	-.33	-.54
Diffusion of Responsibility	-.20	-.61	-.31	-.57
Distortion of Consequences	-.32	-.65	-.32	-.61

Note. Correlations significant at $p < .01$ unless indicated by ns, where $p > .05$.

Table 7
Results of item stability analyses in Study 2 ($N = 101$)

Item No.	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)	≥ 1	0 diff	≤ -1
Doping Moral Disengagement Scale / Doping Moral Disengagement Scale – Short										
1	1.57	1.03	1.61	0.95	-2	3	98.0	15	77	9
<u>2</u>	<u>1.52</u>	<u>0.86</u>	<u>1.43</u>	<u>0.74</u>	<u>-1</u>	<u>3</u>	<u>97.0</u>	<u>9</u>	<u>79</u>	<u>13</u>
3	1.56	0.89	1.48	0.83	-2	3	95.0	12	74	15
4	2.70	1.73	2.35	1.58	-4	6	81.2	11	61	29*
5	2.19	1.44	2.16	1.48	-3	3	88.1	11	73	17
<u>6</u>	<u>2.27</u>	<u>1.52</u>	<u>2.30</u>	<u>1.60</u>	<u>-3</u>	<u>3</u>	<u>92.1</u>	<u>12</u>	<u>73</u>	<u>16</u>
<u>7</u>	<u>2.34</u>	<u>1.34</u>	<u>2.12</u>	<u>1.35</u>	<u>-4</u>	<u>4</u>	<u>89.1</u>	<u>13</u>	<u>61</u>	<u>27*</u>
8	2.52	1.48	2.31	1.42	-3	3	90.1	16	56	29
9	3.25	1.75	2.76	1.63	-3	4	78.2	15	42	44*
<u>10</u>	<u>1.74</u>	<u>1.02</u>	<u>1.70</u>	<u>1.02</u>	<u>-3</u>	<u>3</u>	<u>94.1</u>	<u>13</u>	<u>71</u>	<u>17</u>
11	1.64	0.94	1.65	1.06	-2	3	95.0	12	78	11
12	1.72	0.96	1.69	1.07	-3	2	93.1	15	65	21
13	1.87	1.31	1.74	1.02	-4	6	89.1	19	59	23
<u>14</u>	<u>1.82</u>	<u>1.13</u>	<u>1.73</u>	<u>1.00</u>	<u>-3</u>	<u>3</u>	<u>93.1</u>	<u>17</u>	<u>60</u>	<u>24</u>
15	1.62	0.90	1.67	0.99	-2	2	97.0	16	73	12
<u>16</u>	<u>2.43</u>	<u>1.39</u>	<u>2.07</u>	<u>1.27</u>	<u>-2</u>	<u>3</u>	<u>87.1</u>	<u>10</u>	<u>59</u>	<u>32*</u>
17	1.54	0.82	1.60	0.87	-3	2	97.0	17	71	13
18	2.29	1.44	1.94	1.20	-2	4	87.1	12	58	31*
Doping Self-Regulatory Efficacy Scale										
1	4.77	0.51	4.68	0.63	-1	2	99.0	4	85	12
2	4.69	0.64	4.50	0.82	-1	2	93.1	5	78	18*
3	4.66	0.67	4.54	0.77	-2	2	95.0	8	76	17
4	4.78	0.52	4.71	0.61	-1	1	100.0	5	84	12
5	4.67	0.74	4.60	0.74	-3	2	98.0	5	83	13
6	4.64	0.72	4.56	0.75	-2	2	94.1	8	77	16

Note. Item numbers relate to those presented in Tables 3 and 4. PA = Proportion of Agreement. Underlined items are those used in the Doping Moral Disengagement Scale – Short.

* $p < .05$

Appendix A
The Doping Moral Disengagement Scale

A number of statements describing *thoughts that athletes might have about doping* are listed below. Please read these statements carefully and indicate your level of agreement with each one by circling the appropriate number. Please respond **honestly**.

What is your level of agreement with the following statements?	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
1. It is okay to dope if it helps an athlete to provide for his/her family.	1	2	3	4	5	6	7
2. Saying you "take steroids" feels worse than saying you "use some gear".	1	2	3	4	5	6	7
3. Compared to most lifestyles in the general public, doping isn't that bad.	1	2	3	4	5	6	7
4. Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	1	2	3	4	5	6	7
5. If most athletes in a sport dope, no one athlete should be held responsible for doing it.	1	2	3	4	5	6	7
6. Risks associated with doping are exaggerated.	1	2	3	4	5	6	7
7. Doping is okay if it helps an athlete advise others on how to do it right.	1	2	3	4	5	6	7
8. Using words like "roids", "gear" and "pinning" makes doping feel more acceptable.	1	2	3	4	5	6	7
9. Compared to smoking, doping is pretty safe.	1	2	3	4	5	6	7
10. An athlete shouldn't be blamed for doping if a member of his/her training group has encouraged it.	1	2	3	4	5	6	7
11. It's not right to condemn individuals who dope when many in their sport are doing the same.	1	2	3	4	5	6	7
12. Doping doesn't really harm anyone else.	1	2	3	4	5	6	7
13. It is acceptable to dope if knowledge gained helps an athlete advise others on safe doping.	1	2	3	4	5	6	7
14. Using terms such as "gear" or "juice" makes doping sound less harmful.	1	2	3	4	5	6	7
15. Compared to physical violence, doping isn't that serious.	1	2	3	4	5	6	7
16. An athlete shouldn't be held responsible for doping if his/her coach encouraged him/her to do it.	1	2	3	4	5	6	7
17. If an athlete trains/competes in an environment in which doping is the norm, he/she shouldn't be held accountable for doing it.	1	2	3	4	5	6	7
18. The negative aspects of doping are exaggerated by the media.	1	2	3	4	5	6	7

Appendix B

The Doping Moral Disengagement Scale – Short

A number of statements describing *thoughts that athletes might have about doping* are listed below. Please read these statements carefully and indicate your level of agreement with each one by circling the appropriate number. Please respond **honestly**.

What is your level of agreement with the following statements?	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree
1. Doping is okay if it helps an athlete advise others on how to do it right.	1	2	3	4	5	6	7
2. Using terms such as "gear" or "juice" makes doping sound less harmful.	1	2	3	4	5	6	7
3. Compared to most lifestyles in the general public, doping isn't that bad.	1	2	3	4	5	6	7
4. Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	1	2	3	4	5	6	7
5. It's not right to condemn individuals who dope when many in their sport are doing the same.	1	2	3	4	5	6	7
6. Risks associated with doping are exaggerated.	1	2	3	4	5	6	7

Appendix C

The Doping Self-Regulatory Efficacy Scale

Here we would like to get a better **understanding** of **experiences** that can be **difficult to manage**. For each of the questions listed below, please **circle the number** that best corresponds to your level of **confidence right now**. Please respond **honestly**.

How confident are you right now in your ability to ...	No Confidence		Moderate Confidence		Complete Confidence
1. ...resist doping even if your training group encouraged you to do it?	1	2	3	4	5
2. ...resist doping even if you knew you could get away with it?	1	2	3	4	5
3. ...ignore the temptation to dope even if you knew it would improve your performance?	1	2	3	4	5
4. ...resist peer pressure to dope?	1	2	3	4	5
5. ...reject doping even if most of your training partners did it?	1	2	3	4	5
6. ...ignore the temptation to dope when feeling down physically?	1	2	3	4	5