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

Submitted: 18th July 2017

Abstract

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

Design: Cross-sectional, correlational

Methods: Data were collected from male and female team- and individual-sport athletes and corporate- and hardcore-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. 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 and test-retest reliability 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

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

Understanding factors that may influence use of Performance and Image Enhancing Drugs (PIED) is important for both sport and exercise populations due to the potential detrimental health and legal consequences stemming from their use (McVeigh & Begley, in press; Pope, Wood, Rogol, Nyberg, Bowers, & Bhasin, 2013). Although accurate prevalence rates are difficult to obtain, the estimated prevalence of doping in athletes ranges between 5 and 31% (Momaya, Fawal, & Estes, 2015). An important aim for researchers investigating doping is to identify and understand psychosocial factors that influence the likelihood of athletes and exercisers using PIED (i.e., doping). To conduct such research instruments assessing key variables with scores shown to be reliable and valid are required. Accordingly, we sought to develop and validate scores for psychometric instruments assessing two psychological variables relevant to doping.

The theoretical framework for the current work was Bandura's (1991) social cognitive theory of moral thought and action. Bandura proposed that harmful activities – such as doping – are deterred when people anticipate experiencing negative emotional reactions (e.g., guilt) because of engaging in them. However, people can reduce or eliminate anticipation of negative emotional reactions through use of psychosocial mechanisms collectively termed Moral Disengagement (MD). Representing the conditional endorsement of harmful acts, MD may facilitate doping by allowing sport and exercise participants to use PIED without experiencing negative emotional reactions.

Qualitative research has shown that sport and exercise participants who dope evidence MD when explaining their reasons for doping. For example, Boardley and Grix (2014) conducted semi-structured interviews with nine bodybuilders who had doped; content analysis of the interview data revealed evidence of six MD mechanisms. Boardley, Grix and Dewar (2014) expanded this work in a follow-up study with 64 male bodybuilders from across England, all with experience of doping. Consistent with the initial study, content analysis revealed evidence of the same six MD mechanisms. As these first two studies only involved bodybuilders, Boardley, Grix and Harkin (2015) further extended this line of research by interviewing twelve male team- and individual-sport athletes who had doped. Data analysis again revealed the same six MD mechanisms. Importantly, this third study illustrated considerable consistency in the way in which sport and exercise participants morally disengage with respect to doping. Therefore, qualitative research with sport and exercise participants has provided consistent evidence for the use of the same six MD mechanisms. Further, there is considerable consistency in the way in which sport and exercise participants use these six mechanisms, supporting the potential benefits of developing a single measure of doping MD appropriate for use in both contexts.

The first of these six mechanisms – *moral justification* – occurs when harmful activities are made personally and socially acceptable by portraying how they achieve commendable social or moral purposes. The second – *euphemistic labelling* – diminishes the damaging nature of actions through palliative or convoluted language. The third – *advantageous comparison* – makes detrimental conduct appear less damaging by comparing the act to more heinous ones. The fourth – *displacement of responsibility* – diminishes personal accountability for harmful behavior or its consequences by proffering the act resulted from social pressures. The fifth – *diffusion of responsibility* – also diminishes personal accountability for harmful acts or their outcomes, but instead through group decision making or group action. The final mechanism –

distortion of consequences – occurs when perpetrators of harmful acts avoid information relating to the harm caused or downplay its significance.

These six mechanisms of MD have also been linked with doping in quantitative research. More specifically, researchers have identified positive links between MD, intention to dope, and reported doping across a small number of studies (e.g., Lucidi, Grano, Leone, Lombardo, & Pesce, 2004; Lucidi, Zelli, Mallia, Grano, Russo, & Violani, 2008). However, these studies were conducted with Italian high-school students, a significant proportion of whom (43.0 - 45.2%) did not partake in any extracurricular sport. As such these findings may not extend to participants from all sport and exercise contexts, including those in which prevalence rates for doping are likely to be much higher.

Another variable from Bandura’s (1991) theory that has been empirically linked with doping is self-regulatory efficacy (SRE; Lucidi et al., 2008). Self-regulatory efficacy reflects belief in one’s capabilities to resist personal and social pressures to engage in harmful conduct (Bandura, Caprara, Barbaranelli, Pastorelli, & Regalia, 2001), and increases in SRE are theoretically linked with less frequent engagement in such behavior (Bandura, 1991). When applied to doping, SRE represents a person’s belief in his/her ability to forbear personal and social pressures to dope. In accord with this theorizing, Lucidi et al. (2008) found a negative association between doping SRE and intention to dope in Italian high-school students.

Although Lucidi and colleagues (2008) developed psychometric instruments to assess doping MD and doping SRE, several concerns exist regarding the development and validation of scores obtained with these instruments. One concern relates to the item-development process. Specifically, items were developed based on interviews with 35 high-school students who played sport. However, no information was provided as to whether any of these students had any experience with doping. During these interviews, participants were asked to list situations in which (a) doping would or should not be completely condemned (i.e., to inform doping MD items) and (b) doping would be more likely (i.e., to inform doping SRE items). For the MD measure, the frequencies of common situations were then summed and categorized into the MD mechanisms evoked, leading to the selection of 21 items. Six of these items were then selected for use in the doping MD measure. However, no information was provided on the process through which items were selected. For the SRE measure, the researchers only described how 10 items were developed based on situations described during the interviews. Thus, items for the two instruments were: (a) developed based on interviews with sport participants with little apparent experience of doping when psychometric instruments should be developed using samples representative of intended end users (Clark & Watson, 1995), (b) not appraised for content validity when this should be a key aspect of the scale-development process (Haynes, Richard, & Kubany, 1995), and (c) selected based on criteria not made clear by the researchers. Beyond these concerns, two of the six MD items (i.e., ‘The use of illicit substances is a way to “maximize one’s potential”’ and ‘It is okay to use illicit substances if this can help one to overcome one’s own limits’) are not representative of the definitions for their proposed mechanisms and the doping SRE measure has an imbalance in items representing personal (eight items) versus social (two items) challenges to self-regulation. Finally, the factor structure of neither instrument was appropriately examined nor confirmed in a second sample (see Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Kavussanu, Hatzigeorgiadis, Elbe and Ring (2016) also developed a measure of doping MD. Although the development process was more rigorous than that of Lucidi et al. (2008), several concerns are still apparent. First, one item (i.e., ‘Doping is just a way to “maximize your potential"') lacks content validity because it doesn't reflect its proposed mechanism (i.e., euphemistic labeling). Specifically, it is not clear how this item masks or confers a respectable status on doping through use of sanitized or convoluted verbiage (see Bandura, 1991). Then, a further item (i.e., 'Doping is alright because it helps your team') lacks face validity as it does not resonate with data from qualitative research with PED users in sport that specifically investigated *how* sport participants *actually* morally disengage (Boardley et al., 2015). Finally, two versions (i.e., team-sport and individual-sport) of the instrument were administered during development and validation. Having two versions of the same instrument is problematic because any comparisons made between datasets collected using the different versions may be confounded by differences in item content. A preferable approach would be to develop items suitable across relevant contexts. Thus, although this instrument was largely developed through a more rigorous process than that of Lucidi et al. (2008), issues with aspects of score validity are still apparent.

Beyond addressing the issues outlined above, there are further ways in which assessment of doping MD and doping SRE could be furthered. A key consideration when developing psychometric instruments is that they are validated with samples representative of the target population/s (Clark & Watson, 1995). As such, doping-specific instruments should be developed and validated with samples that include participants with actual experience of doping. To ensure this, doping prevalence should be assessed in the samples used to develop instruments. To date, this has not been the case (Kavussanu et al., 2016; Lucidi et al., 2008).

A further issue relates to the assessment of doping in the exercise context. Specifically, although doping is an issue in exercise as well as sport (see Sjöqvist, Garle, & Rane, 2008), there are currently no instruments available to assess doping MD and doping SRE in exercise populations. This is particularly concerning when one considers doping in exercise populations is now considered to be a public health issue (see McVeigh & Begley, in press; Pope et al., 2013). Given this, there is a need for the development of instruments assessing doping MD and doping SRE suitable for use in exercise populations.

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

Finally, Bandura (1991) proposed MD to be multidimensional, describing eight mechanisms of MD. However, there is currently no instrument capable of measuring individual mechanisms of doping MD in either sport or exercise contexts. This is particularly important when one considers six of the eight mechanisms of MD have been shown to be relevant to doping in both contexts (see Boardley & Grix, 2014; Boardley et al., 2014, 2015). A multidimensional measure of doping MD capable of measuring individual mechanisms of doping MD would make a significant contribution to doping research as it would allow researchers to investigate whether levels of individual MD mechanisms differ and if certain mechanisms are of greater importance to doping than others.

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

When developing these new instruments and validating their scores, we considered five of the six aspects of construct validity identified by Messick (1995). These were content, substantive, structural, generalizability, and external. The content aspect relates to the relevance, representativeness and technical quality of item content and was assessed presently through expert opinion. We considered the substantive aspect – relating to the theoretical rationale for the observed test responses – by ensuring item content was consistent with qualitative research studying the process being assessed (e.g., Boardley & Grix, 2015; Boardley et al., 2014, 2015) and examining theory-based relations between scores generated with the new measures and conceptually associated variables. The structural aspect pertains to the fidelity of the scoring structure to the structure of the construct domains being assessed and was examined through factor analysis. We addressed the generalizability aspect – the extent to which score properties and interpretations generalize to and across groups and settings – through multisample analyses. The external aspect relates to evidence for convergent and discriminant validity as well as criterion relevance, and was considered currently through relations with theoretically relevant instrument scores. Evidence for the final aspect of construct validity – consequential – is represented through positive and negative consequences stemming from application of new measures. As such, this final aspect was more relevant to future application of the measures rather than initial instrument development and score validation.

Method

We sought to develop three psychometric instruments: (1) the Doping Moral Disengagement Scale (DMDS), (2) the Doping Moral Disengagement Scale – Short (DMDS-S), and (3) the Doping Self-Regulatory Efficacy Scale (DSRES). We followed appropriate procedures for developing psychometric instruments (i.e., Clark & Watson, 1995; Fabrigar et al., 1999; Haynes et al., 1995), collecting data from two primary samples (i.e., Samples 1 and 2) for the main analyses plus another sample for examining test-retest reliability (i.e., Sample 3).

Participants

Sample 1. Participants were team- (n = 181) or individual- (n = 70) sport or hardcore[1]- (n = 44) or corporate[2]- (n = 23) gym participants (n_{male} = 203; n_{female} = 115); ages ranged from 16 to 70 years (M = 23.3, SD = 8.2). Participants had been training/competing for an average of 7.3 years (SD = 5.4), spent an average of 8.6 hours (SD = 3.9) per week training, and had trained in their current gym/with their current team for an average of 4.1 years (SD = 4.5). Self-reported lifetime prevalence of doping was 14.5%[3].

Sample 2. Participants were team- (n = 14) or individual- (n = 99) sport or hardcore- (n = 89) or corporate- (n = 98) gym participants (n_{male} = 172; n_{female} = 128); ages ranged from 17 to 73 years (M = 29.6, SD = 12.4). Participants had been training/competing for an average of 9.1 years (SD = 8.6), spent an average of 7.9 hours (SD = 5.0) per week training, and had trained in their current gym/with their current team for an average of 3.8 years (SD = 4.9). Self-reported lifetime prevalence of doping was 15.3%.

Sample 3. Participants were team- (n = 9) or individual- (n = 78) sport or hardcore- (n = 5) or corporate- (n = 9) gym participants (n_{male} = 60; n_{female} = 41); ages ranged from 16 to 70 years (M = 35.2, SD = 13.5). Participants had been training/competing for an average of 9.4 years (SD = 7.4), spent an average of 7.9 hours (SD = 4.8) per week training, and had trained in their current gym/with their current team for an average of 4.4 years (SD = 5.0).[4]

Measures

---

[1] Hardcore gymnasia are those targeting and designed for experienced/competitive bodybuilders.
[2] Corporate gymnasia are those targeting and designed for users training for general health and fitness.
[3] Prevalence rates by gender and sport/exercise context for the combined sample for the multisample analyses 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%.
[4] 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 once they had provided this information. 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.

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

Sport moral disengagement. The eight-item moral disengagement in sport scale-short (MDSS-S; Boardley & Kavussanu, 2008) was used to measure sport MD. Participants were asked to read statements representing MD (e.g., 'Insults among players do not really hurt anyone') and asked to indicate their level of agreement with each using a Likert scale anchored by 1 (*strongly disagree*) and 7 (*strongly agree*). Scores on this instrument have shown very good levels of internal consistency and evidence for factorial and convergent validity (Boardley & Kavussanu, 2008).

Peer pressure SRE. The five-item SRE scale (Bandura et al., 2001) was used to assess peer-pressure SRE. Items (e.g., 'Resist peer pressure to drink beer, wine or alcohol') assess peoples' beliefs regarding their ability to resist peer pressure to engage in high-risk activities involving alcohol, drugs and transgressive behavior. For each item, participants rated their confidence using a Likert scale anchored by 1 (*no confidence*) and 5 (*complete confidence*). Scores on this instrument have shown good levels of internal consistency and evidence for construct validity (Bandura et al., 2001).

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

> *Having returned to training following a period of injury, you are feeling very out of shape. As such, you feel the need to get back in shape as soon as possible. A friend who you train with has been taking a training supplement that he/she says really helped him/her get back in shape quickly following a similar injury. He/she offers to give you some and you decide*

to take it. Subsequently you get back in shape much quicker than expected, but then discover the supplement you have been taking is a banned performance-enhancing substance. However, due to the improvements you have experienced, you decide to continue taking the substance.

Participants then expressed how they anticipated feeling about continuing to dope by responding to the five items (e.g., 'I would feel remorse, regret') of the guilt scale from the State Shame and Guilt Scale (SSGS; Marschall, Saftner, & Tangney, 1994). Participants responded on a 5-point scale ranging from 1 (*not at all*) to 5 (*extremely*). Marschall et al. (1994) provided evidence supporting the internal consistency and validity of scores from this subscale.

Empathy. The 7-item perspective taking (e.g., 'before criticizing somebody, I try to imagine how I would feel if I were in their place') and 7-item empathic concern (e.g., 'I am often quite touched by things that I see happen') subscales of the Interpersonal Reactivity Index (Davis, 1983) were used to measure empathy. Participants indicated how well each statement described them by responding on a scale from 1 (*does not describe me well*) to 7 (*describes me very well*). Scores from these respective subscales have shown internal consistency reliability and validity (Davis, 1983). In line with previous research (Kavussanu, Stanger, & Boardley, 2013), we combined these subscales to produce one empathy score for use in our primary analyses.

Procedures

Instrument development and score validation. Development of the new instruments and examination of the validity of instrument scores occurred over several distinct phases. The procedures for each of these phases follow.

Recruitment and data collection. Recruitment for the primary samples commenced once approved by the ethics committee of the first author's institution. Our approach to recruitment differed for sport versus exercise participants. For sport participants, we contacted sport coaches and sought permission to visit a designated training session to introduce the project to athletes and invite them to participate. For exercise participants, we contacted gym managers and sought permission to visit their gymnasia to introduce the project to exercisers and invite them to participate. Before completing the questionnaire, all volunteers were informed the survey examined sporting attitudes and honesty in responses was vital to the study. We also explained that all responses would be kept confidential and would be used only for research purposes. Participants signed an informed consent form prior to participating and the questionnaire pack took approximately 10–15 minutes to complete. Data were collected across two phases. Data from the first phase (i.e., Sample 1) was analyzed before the second phase (i.e., Sample 2) commenced. This allowed for adjustments to item content or item generation between phases.

Item development. First, informed by a review of existing instruments assessing the target constructs, qualitative papers investigating the constructs in physical-activity contexts and consultation with relevant experts (i.e., sport psychologists, sport coaches, exercise leaders, sport and exercise participants), we developed large pools of items for doping MD and doping SRE (see Clark & Watson, 1995). These items were then piloted with a sample (N = 280) of sport and exercise participants. Item analyses were conducted on these pilot data to ensure the bivariate correlations between items were consistent with a target mean interitem correlation of .15 to .50 for each subscale (see Clark & Watson, 1995); items with interitem correlations inconsistent with this target range were eliminated or adapted.

The two item pools were then subjected to content-validity assessment to determine whether they characterized their intended domain (Dunn, Bouffard, & Rogers, 1999; Haynes et al., 1995). Content validity was examined through expert opinion, and the size and composition of the expert panel was informed by relevant guidelines (Dunn et al., 1999). As such, items were evaluated by 12 academics[5] who had experience in instrument development but were not involved in the item-writing phase of the project. Each of the experts had a PhD in sport psychology or psychology, had published in peer-reviewed international sport psychology journals, and were at the time of assessment employed in psychology, kinesiology or sport science departments in Europe, North America, or Australia.

Content-validity assessment packs were emailed to the 12 experts, and completed packs were sent via return email; all experts returned their completed packs in less than eight weeks from the date of dispatch. Assessment packs consisted of ten sections: (a) introduction and instructions, (b) moral justification, (c) euphemistic labelling, (d) advantageous comparison, (e) displacement of responsibility, (f) diffusion of responsibility, (g) distortion of consequences, (h) response format, scale, and general comments (moral disengagement), (i) doping self-regulatory efficacy, and (j) response format, scale, and general comments (doping self-regulatory efficacy). Sections b-g and i consisted of a definition of the relevant construct followed by the items developed to assess the relevant construct. Experts were asked to rate each item using a Likert-type scale ranging from -3 (*Not at all Representative*) to 3 (*Very Representative*), and were provided with the opportunity to comment on the relevance of each item to the sport/exercise context and construct definition. Sections h and j presented the proposed response format and scale for the new instruments, and asked the experts to comment on the format and scale, the match of the instrument's attributes to its function, and suggested additions, deletions,

[5] 16 potential experts were contacted, 12 of whom agreed to act as content-validity assessors.

modifications, or any other matter they felt appropriate. Items with poor content validity (see results section for details on how this was determined) were eliminated or adjusted.

The final item pools were tested with a second pilot sample (N = 122) of sport and exercise participants to: (a) ensure the inter-item correlations of theoretically related items remained consistent with the target mean interitem correlation for subscales (i.e., 15-.50) and (b) obtain qualitative feedback on the difficulty and wording of items (Clark & Watson, 1995). Any further necessary adjustments to item wording were then made.

Factorial, convergent and discriminant validity, and internal consistency. To provide further evidence for construct validity, we sought to establish evidence for the factorial, convergent and discriminant validity of scores from the new instruments, as well as for internal consistency. Factorial validity relates to the number of separate dimensions represented in a measure and was analyzed by identifying the factor structure in the first sample before confirming it in the second. Evidence for convergent validity is established if scores correlate at least moderately with scores from instruments assessing variables within the target construct's nomological network (Vaughn & Daniel, 2012). As such, associations of DMDS, DMDS-S and DSRES scores with empathy and anticipated guilt were computed to provide evidence for convergent validity. In contrast, discriminant validity represents the degree to which scores are empirically distinguishable (i.e., $r \leq .90$) from those of closely related but conceptually distinct constructs (Vaughn & Daniel, 2012). To establish discriminant validity evidence for DMDS and DMDS-S scores, we analyzed their respective correlations with sport MD. For DSRES scores, we computed the correlation with peer pressure SRE. Here and elsewhere, the magnitudes of correlation coefficients were interpreted based on the guidelines provided by Cohen (1992), with coefficients of 0.10, 0.30, and 0.50 representing small, moderate, and large effect sizes, respectively. Finally, the internal consistency of scores for the overall scales plus the subscales of the DMDS were estimated using Cronbach's alpha.

Test-retest reliability. Test-retest reliability represents consistency of scores across time in a population with stable scores on the construct being assessed (Lohr, 2002). It can be examined by administering a measure to the same sample twice and correlating the two sets of scores (Pedhazur & Schmelkin, 1991). Given both doping MD and doping SRE should be stable over the short term, good levels of test-retest reliability (i.e., ≥ 0.70) for scores obtained with the two instruments would provide evidence for their temporal reproducibility. To determine this, once ethical approval from the first author's institutional ethical review committee had been obtained, doping MD and doping SRE data were collected on two occasions using paper and online versions of the new instruments. Recruitment procedures for face-to-face data collections at Time 1 matched procedures used with Samples 1 and 2. Recruitment of online participants involved advertising the study through sport-club and gymnasia websites and discussion groups, social media, and personal contacts. Adverts included the basic information on the study, as well as a link to the survey webpage. The webpage provided full details of the study, equivalent to that normally presented at face-to-face collections. Prior to starting the online questionnaire, participants were informed that by clicking on the link to start the questionnaire they were providing informed consent to participate. For both face-to-face and online collections, participants were informed participation involved providing data on two occasions, and that data provided on the two occasions would be linked.

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

Results

Item Development

Thirty-eight items representing doping MD and 13 for doping SRE were generated through the processes described earlier and underwent initial pilot testing with a sample of sport and exercise participants (*N* = 280). For DMDS/DMDS-S items a 7-point Likert scale anchored by 1 (*strongly disagree*) and 7 (*strongly agree*) was used during pilot testing and all subsequent data collections. This is consistent with existing instruments assessing MD in sport (e.g., Boardley & Kavussanu, 2007, 2008) and supported by research that suggests this format offers the best compromise between reliability, validity, discriminating power, and respondent preferences (Preston & Colman, 2000). However, to be consistent with existing instruments assessing SRE (e.g., Bandura et al., 2001) the DSRES used a 5-point Likert scale anchored by 1 (*no confidence*) and 5 (*complete confidence*) throughout pilot testing and all subsequent data collections. Analyses of the pilot data[7] demonstrated bivariate correlations among conceptually related items to be largely in the target range. As a result of these analyses, 28 doping MD items

[6] 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.
[7] Both pilot samples were distinct from the three primary samples described earlier.

were retained and 10 were adapted. Also, five new items were generated based on participant feedback and data from the pilot analyses. For doping SRE, 11 items were retained and two adapted; no new items were developed. Following these changes, the item pool consisted of 43 doping MD (i.e., 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) and 13 doping SRE items.

The content validity of the items was then examined. For each item, we took the mean of the experts' ratings once any discrepant ratings had been removed. Discrepant ratings were those that deviated so much from the other judges that they negatively impacted the validity of the statistics generated (see Hambleton, 1980), and such ratings were defined as those that were two or more response options lower the next lowest option (e.g., scoring -2 when the next lowest score was a 0). Once any such scores had been removed, mean scores for each item were calculated, and those with mean scores of 2.0 or above were retained; those with scores below 2.0 were revised (based on the experts' qualitative comments) or removed. Of the 56 items assessed, 26 were retained without change, 17 underwent revisions to item content and 13 were removed altogether. Six new doping MD items were also created based on expert comments.

These 49 items (i.e., 36 doping MD; 13 doping SRE) were then pilot tested with a sample of 122 sport and exercise participants to: (a) ensure theoretically related items were correlated ($r \geq .15$), (b) examine the internal consistency of scores from all subscales and (c) obtain qualitative feedback on item difficulty and wording (Clark & Watson, 1995). Correlation analyses demonstrated the majority (i.e., 86%) of bivariate correlations for doping MD items to be in the target range; items ($n = 3$) not correlated as expected with conceptually related items were adapted to improve the wording; all doping SRE items were correlated in the target range.

Cronbach alpha coefficients ranged from .70 to .85 (*M* = .78) for scores from the doping MD subscales; alpha for scores from the 13 doping SRE items was .94. Qualitative feedback on item difficulty and wording was largely positive, with only minor adjustments to wording needed for three items. Following initial item development, 49 items (i.e., 36 doping MD; 13 doping SRE) were taken forward into the main construct-validity phase.

Factorial, Convergent and Discriminant Validity and Internal Consistency

Data from the three samples were used at different stages of the instrument-development process. Data from Sample 1 were used in the preliminary analyses and the exploratory stages of factorial validity examination. Sample 2 data were used in the confirmatory stages of factorial validity examination. Analyses examining the convergent and discriminant validity as well as internal reliabilities of instrument scores were performed on the Sample 1 and Sample 2 data separately. Finally, analyses testing the test-retest reliability of scores from the instruments employed the Sample 3 data collected solely for this purpose.

Preliminary analyses. Only 0.83% of data points were missing, and missing data were assumed missing at random such that the probability of a missing value on a variable was assumed to be unrelated to values of that variable (see Enders, 2006). The expectation maximization algorithm was used to impute missing values. Before seeking to establish evidence for the factorial validity of instrument scores, we used a two-stage process to identify the most effective items for measuring each construct; selected items were retained for use in subsequent testing. First, inter-item correlations were examined within each construct and all item scores were intercorrelated as anticipated (see Clark & Watson, 1995). However, one *displacement of responsibility* item (i.e., “Seeing athletes achieve goals through doping encourages others to dope too") had considerably weaker intercorrelations (*M* = .26) than all other items (*M* = .52) for this MD mechanism.

Exploratory factor analysis (EFA) was then conducted on each of the seven constructs individually (i.e., six MD mechanisms plus doping SRE) using principal axis extraction, with extraction based on an eigenvalue ≥ 1.00. Prior to conducting these analyses, we determined the appropriateness of the matrices using the following criteria as suggested by Dziuban and Shirkey (1974): (a) a significant Bartlett's test of sphericity and (b) a Kaiser-Meyer-Olkin measure of sampling adequacy of > .80; all seven matrices satisfied these criteria. Each EFA produced a unidimensional factor structure, and except for one item, all items had factor loadings ≥ 0.61. The item with the weaker factor loading (0.32) was the aforementioned displacement of responsibility item with weaker interrelations with other items. Given the issues with this item, we removed it from further analyses and replaced it with a newly developed item (i.e., "Athletes shouldn't be held responsible for doping if they feel pressured to do it to keep up with others") before data collection for Sample 2 commenced. All other items were retained.

Factorial validity. Confirmatory factor analysis was used in the next step to identify the best items and because it offers a rigorous and appropriate method for confirming hypothesized factor structures (Fabrigar et al., 1999). EQS 6.1 (Bentler & Wu, 2002) software and the Maximum Likelihood method were used. As discussed previously, the DMDS was hypothesized to consist of six first-order factors and the DSRES was hypothesized to be unidimensional.

In initial analyses, the normalized estimate of Mardia's coefficient indicated deviation from multivariate normality. Thus, the Robust Maximum Likelihood estimation method was used for all analyses. This method provides more accurate standard errors, chi-squared values, and fit indices when data are non-normally distributed (Bentler & Wu, 2002). The case numbers with the largest contribution to normalized multivariate kurtosis suggested minimal impact of outliers and as a result no cases were deleted. Standard indices and criteria were used to estimate model fit (Hu & Bentler, 1999). Fit indices for all models appear in Table 1.

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

Although our hypothesized model was supported, it was important to rule out alternative models. For instance, the development of MD instruments for other contexts has shown that pairs of MD mechanisms sometimes converge to form single factors (e.g., Boardley & Kavussanu, 2007) or can all converge to form a unidimensional measure (e.g., Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). Thus, once the final item content for the DMDS was confirmed, we compared the fit of the six-factor model with two other possible structures based on those seen in existing instruments. Specifically, these were a three-factor model in which the six mechanisms were grouped according to the aspect/s of detrimental conduct they operate upon (M2; Osofsky, Bandura, & Zimbardo, 2005) and a unidimensional model in which all items loaded on a single factor (M3; Bandura et al., 1996). Also, based on very strong factor correlations between advantageous comparison/distortion of consequences (i.e., .93) and displacement/diffusion of responsibility (i.e., .91), we also tested a four-factor model (M4) in which these respective mechanism pairs were combined into single factors. As shown in Table 1, the fit of the six-factor model (M1b) was superior to any of these alternative models. Thus, the six-factor model was accepted as the best model for the DMDS. Factor correlations for this model can be found in Table 2, and items, factor loadings and error variances are shown in Table 3.

Although the 18-item instrument provides the capability of measuring the six MD mechanisms individually, some users may only need to measure overall doping MD. For such instances, a short version would reduce the time needed for completion. For these reasons, we also developed the DMDS-S. Our aim was to develop a six-item instrument with one item for each of the six relevant mechanisms of MD. Two main steps were involved in selecting items for the DMDS-S. Potential items were first selected based on item content, with the 12 (i.e., two for each MD mechanism) shortest and simplest items retained for further analysis. Then, the factor structure underlying these 12 items was examined using EFA and CFA with the data from Sample 1; the suitability of the matrix for EFA was determined using the same approach as described earlier. The EFA was conducted using principal axis extraction, and extraction was based on an eigenvalue ≥ 1.00; a single factor was extracted. Then, for each of the six MD mechanisms the item with the stronger factor loading was retained. Once these six items had been identified, a single-factor CFA was specified (i.e., M5a). As seen in Table 1 (Row 7), this model had a very good fit. However, the LM Test results indicated the presence of a correlated error between the advantageous comparison and distortion of consequences items. Testing of a subsequent model (M5b) with this specified resulted in an excellent model fit. Specifying correlated errors when present is important to prevent possible inaccurate parameter estimates (see Kline, 2015). However, such associations can be sample specific so this model was accepted under the proviso that the presence of this correlated error would be confirmed in Sample 2.

When developing psychometric instruments, it is important to confirm initial factor structures using a separate sample to ensure initial results are not sample specific (Fabrigar et al., 1999). As such, we used the data from Sample 2 to confirm the factor structures identified in Sample 1 for the DMDS and DMDS-S. As shown in Table 1, the final DMDS model from the Sample 1 analyses showed excellent model fit (Row 11) and again supported the presence of a second-order doping MD factor (Row 12). This was also the case for the DMDS-S model (Row 13), where the significant correlated error identified in Sample 1 (i.e., $r = .30$, $p < .05$) was again present (i.e., $r = .25$, $p < .05$).

Similar procedures were followed for the development of the DSRES. Using the Sample 1 data, initially all 13 items were specified to load on a single factor (M6a). This model had an inadequate fit (Row 9). Guided by modification indices and/or standardized residuals, 7 items were removed in a series of CFAs. This iterative process was also guided by an aim to develop a final instrument that would contain items representing all the main personal and social influences on doping use as identified in past research (e.g., Boardley & Grix, 2014; Boardley et al., 2014, 2015). A final six-item model (M6b) demonstrated an excellent model fit (Row 10). Subsequent testing of this model using the data from Sample 2 also resulted in excellent model fit (Row 14), therefore confirming the unidimensional structure of the DSRES. Items, factor loadings and error variances for the DMDS-S and DSRES appear in Table 4. The final versions of the DMDS, DMDS-S and DSRES can be found in the appendices.

Multisample analyses. When developing instruments for use in diverse populations, it is important to determine their measurement invariance across sub-groups within the population. As such, we tested for measurement invariance by sex and across the four sport and exercise groups represented in our sample using multisample analyses. Different aspects of invariance can be tested depending on the research question (Cheung & Rensvold, 2002). As we were interested in construct validity and whether the instruments were appropriate for making comparisons between groups, we tested three aspects of invariance relevant to these issues (Byrne, 2006): a) configural invariance (i.e., when items are indicators of the same factors in different groups), b) metric invariance (i.e., when all factor loadings are specified as equal across groups) and (c) equivalence of construct variance and covariance (ECVC; i.e., when the variances and covariances of the latent variables are equivalent across groups). Prior to invariance testing, we estimated baseline model fit separately for each group (see Byrne, 2006). We then tested for configural invariance, metric invariance and ECVC, respectively, by progressively imposing the appropriate constraints. We examined ΔCFI as the constraints were imposed, with values of less than 0.01 indicating no significant difference between models (Cheung & Rensvold, 2002). Results are presented in Table 5.

Sex invariance. For the DMDS, model fit for the baseline models was very good for male participants and acceptable-to-good for female participants, and configural invariance was demonstrated by the good fit of the relevant model. Metric invariance was also established by a ΔCFI of .00. The ECVC was not established though, as imposing such constraints resulted in a ΔCFI > .01. For the DMDS-S, model fit for the baseline models was very good for male participants and excellent for female participants, and configural invariance was demonstrated by the very good fit of the relevant model. Metric invariance was also established, as shown by a ΔCFI < .01. However, like the DMDS, the ECVC was not established as ΔCFI > .01. Finally, for the DSRES, model fit for the baseline models was excellent for male and female participants, respectively, and configural invariance was demonstrated by the excellent model fit. Metric invariance was also established, as shown by a ΔCFI < .01. However, like the DMDS and DMDS-S, the ECVC was not established as ΔCFI > .01.

Sport/exercise group invariance. For the DMDS, model fit for the baseline models ranged from acceptable-to-good for corporate-gym exercisers to excellent for hardcore-gym exercisers and individual-sport athletes. Configural invariance was established through the good fit of the relevant model. Metric invariance was also established, as shown by a ΔCFI < .01. However, the ECVC was not established as ΔCFI > .01. For the DMDS-S, model fit for the baseline models was excellent for all four groups, and configural invariance was demonstrated by the excellent fit of the relevant model. Complete metric invariance was not established though, as the LM test results for this model indicated two of the specified constraints (i.e., constraining the factor loadings of the advantageous comparison and diffusion of responsibility items for equivalence between corporate gym exercisers and individual-sport athletes) each led to an increase in χ^2 of ≥ 5.0/df. Respecification of this model with these constraints released led to an excellent model fit and a ΔCFI < .01. This supports partial metric invariance (see Byrne, Shavelson, & Muthén, 1989). However, ECVC was again not established. Finally, for the DSRES, model fit for the baseline models was excellent for all four groups, and configural invariance was demonstrated, with an excellent fit for the relevant model. Metric invariance was also established, as shown by a ΔCFI < .01. However, the ECVC was again not established.

Convergent and discriminant validity. We examined convergent validity by computing associations of DMDS, DMDS-S and DSRES scores with scores for empathy and anticipated guilt. According to Bandura's (1991) theory, doping MD should correlate negatively and doping SRE should relate positively with these variables. In both samples, DMDS (Sample 1 $r = -.28$, $p < .01$; Sample 2 $r = -.32$, $p < .01$) and DMDS-S (Sample 1 $r = -.26$, $p < .01$; Sample 2 $r = -.33$, $p < .01$) scores were negatively related to empathy, whereas DSRES scores were positively associated with it (Sample 1 $r = .15$, $p < .01$; Sample 2 $r = .36$, $p < .01$). Further, DMDS (Sample 1 $r = -.68$, $p < .01$; Sample 2 $r = -.60$, $p < .01$) and DMDS-S (Sample 1 $r = -.66$, $p < .01$; Sample 2 $r = -.59$, $p < .01$) scores were negatively related to anticipated guilt, whereas DSRES scores were positively associated with it (Sample 1 $r = .38$, $p < .01$; Sample 2 $r = .48$, $p < .01$). Table 6 contains correlations of DMDS subscale scores with empathy and anticipated guilt in both samples. These correlations provide evidence for the convergent validity of all subscale scores. However, the degree of convergence was generally weaker for the euphemistic labelling subscale in comparison to the other five. Collectively these correlations also provide some evidence of distinct predictive capabilities, supporting some degree of conceptual separation between the subscales despite their largely strong inter-correlations (see Table 2).

We examined discriminant validity by computing the correlation between sport MD (Boardley & Kavussanu, 2008) and DMDS and DMDS-S scores and between peer pressure SRE and DSRES scores. Evidence for the discriminant validity of scores obtained with the three new instruments would be provided if moderately strong positive correlations were obtained. However, if the correlations were too strong (i.e., $r > .90$) this would suggest too much overlap between the instrument scores (Kline, 2015). In both samples, DMDS (Sample 1 $r = .59$, $p < .01$; Sample 2 $r = .58$, $p < .01$) and DMDS-S (Sample 1 $r = .56$, $p < .01$; Sample 2 $r = .58$, $p < .01$)

scores were positively related to sport MD, and DSRES scores were positively related to peer pressure SRE (Sample 1 $r = .41$, $p < .01$; Sample 2 $r = .56$, $p < .01$). Thus, overall the findings support the convergent and discriminant validity of DMDS, DMDS-S and DSRES scores.

Internal consistency. Cronbach alpha values showed internal consistency to be either good or very good for all subscales of the DMDS in both samples (see Table 2). Alpha values for overall doping MD for the DMDS were excellent in both samples (i.e., Sample 1 = .95; Sample 2 = .96). Similarly, alpha values were very good for the DMDS-S in both samples (i.e., Sample 1 = .86; Sample 2 = .89) and excellent for the DSRES in both samples (i.e., Sample 1 = .93; Sample 2 = .94), providing strong support for their internal consistency.

Test-retest reliability

Test–retest reliabilities were assessed for overall scores obtained using the DMDS, DMDS-S and DSRES, as well as for the DMDS subscales. Intraclass correlation coefficients for the DMDS, DMDS-S and DSRES were .94, .93 and .87, respectively, and ranged from .87 to .93 ($M = .91$) for the individual DMDS subscales. Overall, the findings demonstrate very good to excellent levels of score reproducibility for all three instruments.

Discussion

Research has highlighted the potential importance of doping MD and doping SRE to the regulation of doping in sport and exercise (e.g., Boardley & Grix, 2014; Boardley et al., 2014, 2015; Lucidi et al., 2004, 2008). However, valid instruments are needed to measure these constructs and substantively advance the knowledge base. The relevant extant psychological assessments tied to doping have shortcomings, and there is currently no multidimensional measure of doping MD available nor any instrument to assess doping MD in exercise populations. Therefore, we sought to develop psychometrically sound instruments assessing doping MD and doping SRE, pursuing item development and assessment of score validity and reliability using an expert panel, two pilot samples, and three primary samples.

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

Convergent validity of overall DMDS scores was evidenced by the strong negative correlation between doping MD and anticipated guilt, and the moderate negative correlation between doping MD and empathy. These associations are consistent with theory and research (Bandura, 1991; Bandura et al., 1996). Evidence for the convergent validity of scores for most DMDS subscales was also provided, with relationships of the subscales with anticipated guilt and empathy generally consistent with those obtained using all 18 instrument items. However, evidence for the convergent validity of scores obtained for euphemistic labelling was weaker than for scores obtained with the other five subscales. Because use of euphemistic terms regarding doping (e.g., gear, juice, etc.) is a key aspect of doping culture (see Andrews, Sudwell, & Sparkes, 2005), people may at times use such terminology to fit in with this culture and not exclusively for making doping appear less harmful.

Discriminant validity of the DMDS was evidenced through the associations of DMDS and MDSS-S scores, which showed the DMDS scores to be related to – but distinct from –

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

We also developed a short version of the DMDS termed the DMDS-S, enabling a concise measure of overall doping MD. Following procedures used successfully to develop previous short versions of MD instruments (e.g., Boardley & Kavussanu, 2008), we found evidence for the factorial validity of DMDS-S scores, as well as their discriminant and convergent validity. Importantly, the instrument has items representing all six of the relevant mechanisms of MD (see Boardley & Grix, 2014; Boardley et al., 2014, 2015), ensuring that the DMDS-S generates doping MD scores that are equally representative of each of the six mechanisms. Although for us the DMDS is the preferred option for most research given it is representative of underlying theory (i.e., Bandura, 1991), the development of the DMDS-S provides a generalized and brief assessment suitable for use when the DMDS is not practical to administer.

Another important context-specific variable from Bandura’s (1991) theory is doping SRE. Although past research had identified the potential importance of this variable to doping (e.g., Lucidi et al., 2008), prior to the current research scores of no doping SRE measure had been appropriately validated. Based on theory (Bandura, 1991) and the factorial structure of existing instruments assessing SRE (e.g., Bandura et al., 1996), the DSRES was constructed and scores demonstrated the expected unidimensional structure. Moreover, associations with empathy and guilt supported the convergent validity of DSRES scores, while associations with peer-pressure SRE supported the discriminant validity of DSRES scores.

We additionally assessed internal consistency and test-retest reliability of scores obtained using the three new instruments. The internal consistency of all higher- and lower-order scales surpassed the minimum criterion level recommended when developing new instruments (i.e., 0.80; Clark & Watson, 1995). Also, test-retest reliability levels across a nine- to 16-day period, using a separate sample, were good to excellent. Thus, over the short term, scores obtained using the DMDS, DMDS-S and DSRES are stable and can be replicated with a high degree of measurement precision (Widaman, Little, Preacher, & Sawalani, 2011). Overall, across all samples and analyses, the instruments performed well in measuring their target constructs.

Beyond the main validity and reliability analyses, we also utilized multisample analyses to examine the measurement invariance of scores from all three instruments. In each case, we performed two sets of multisample analyses. The first examined measurement invariance by sex to examine whether items performed similarly in males and females. The second set examined measurement invariance across four sport/exercise groups. All three instruments showed configural invariance, meaning across all groups the same subsets of items are associated with the same constructs (Cheung & Rensvold, 2002). In all but one case, complete metric invariance was established. This demonstrates the strength of the relationship between all items and their underlying constructs was the same (Cheung & Rensvold, 2002). The one exception was across the four sport/exercise groups with the DMDS-S, where partial metric invariance was evidenced

(Byrne et al., 1989). This is present when most items for a given latent variable have loadings that are invariant across groups. If this is the case, cross-group comparisons can still safely be made (Reise, Widaman, & Pugh, 1993). Thus, the DMDS, DMDS-S and DSRES are suitable for research testing substantive hypotheses regarding group differences by sex and among the four sport/exercise types tested.

Equivalence of construct variance exists when the range of responses given to each item is the same across groups, whereas equivalence of construct covariance is apparent when structural relationships are equivalent across groups (Cheung & Rensvold, 2002). As with our analyses, these two forms of invariance are often tested simultaneously (see Byrne, 2006). In contrast to the other forms of equivalence tested, our analyses showed ECVC was not evident for any of the instruments by sex or across the four sport/exercise groups. This suggests the range of item responses and/or the strength of the relations between MDMS subscales varied among these groups. Importantly, prevalence of doping differed markedly between genders and across sport/exercise context. Given that levels of doping MD and doping SRE are likely to be more extreme in athletes with experience of doping, these differences in prevalence may have led to greater construct variance in samples with higher prevalence rates. In addition, the degree of association between MD subscales may also be influenced by doping prevalence. Future research should address what may lead to differences in construct variance and covariance across groups.

Limitations and Future Directions

The current project developed three psychological assessments relevant to doping in sport and exercise that each provide scores with good psychometric properties. Nevertheless, limitations relating to certain aspects of the research should be acknowledged. First, whilst we achieved our aim of sampling from populations showing doping prevalence rates in line with existing estimates, the use of self-report to assess doping prevalence has known limitations when applied to socially sensitive behaviors such as PED use. Specifically, assessing doping through self-report is thought to underestimate the true prevalence of doping (de Hon, Kuipers, & van Bottenburg, 2015). Accordingly, the prevalence of doping in our samples may be higher than reported. Next, although the test-retest analyses presented evidence for good levels of short-term stability of DMDS, DMDS-S and DSRES scores, the sample for these analyses largely consisted of individual-sport athletes. Future research should assess test-retest reliability with samples more representative of the other three sub-populations. Also, difficulties in accessing the same participants on two occasions across the stipulated period resulted in a low percentage of athletes completing the measure twice, especially for face-to-face collections. Although we have no evidence of systematic drop out, in future work attempts should be made to increase the retention rate across the two data-collection occasions.

It is important to recognize that validation is a continuous process (Clark & Watson, 1995). Thus, further aspects of validity remain to be examined, such as predictive validity over time and associations with key variables not assessed here. For example, associations of scores obtained from these instruments with socially situated emotions such as shame could be examined. Further, their validity and reliability in other doping populations and alternative cultures require investigation. Finally, the multidimensional nature of the DMDS – and the divergent associations of its subscales with anticipated guilt – present the opportunity for research determining which MD mechanisms may have the greatest potential to facilitate doping.

Conclusion

Through a rigorous set of processes, we developed three psychometric instruments relevant to the psychology of doping and supported the validity of scores obtained with them.

Evidence for the construct validity, internal consistency and test-retest reliability of scores obtained from all three instruments was provided. Items for all three instruments have high levels of face and content validity, making them particularly suitable for use with athletes and exercisers who have experience in environments where doping is a salient aspect of the culture. The DMDS makes a particular contribution to the literature, being the first measure capable of capturing the individual doping MD mechanisms, and for use with exercise populations. The DSRES is also the first doping-contextualized instrument developed for use with exercise populations. We look forward to seeing these instruments employed and further evaluated in future research.

References

Andrews, G. J., Sudwell, M. I., & Sparkes, A. C. (2005). Towards a geography of fitness: An ethnographic case study of the gym in British bodybuilding culture. *Social Science & Medicine, 60*(4), 877–891. DOI: 10.1016/j.socscimed.2004.06.029

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modelling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411-423. DOI: 10.1037/0033-2909.103.3.411

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W.M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development: Theory, research, and applications* (Vol. 1, pp. 71-129). Hillsdale, NJ: Erlbaum.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*(2), 364-374. DOI: 10.1037/0022-3514.71.2.364

Bandura, A., Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Regalia, C. (2001). Sociocognitive self-regulatory mechanisms governing transgressive behavior. *Journal of Personality and Social Psychology*, *80*(1), 125-135. DOI: 10.1037/0022-3514.80.1.125

Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide.* Multivariate Software.

Boardley, I. D., & Grix, J. (2014). Doping in bodybuilders: A qualitative investigation of facilitative psychosocial processes. *Qualitative Research in Sport, Exercise and Health, 6*(3), 422-439. DOI: 10.1080/2159676X.2013.766809

Boardley, I. D., Grix, J., & Dewar, A. J. (2014). Moral disengagement and associated processes in performance-enhancing drug use: a national qualitative investigation. *Journal of Sports Sciences, 32*(9), 836-844. DOI: 10.1080/02640414.2013.862842

Boardley, I. D., Grix, J., & Harkin, J. (2015). Doping in team and individual sports: A qualitative investigation of moral disengagement and associated processes. *Qualitative Research in Sport, Exercise and Health, 7*(5), 698-717. DOI: 10.1080/2159676X.2014.992039

Boardley, I. D., & Kavussanu, M. (2007). Development and validation of the Moral Disengagement in Sport Scale. *Journal of Sport & Exercise Psychology*, *29*(5), 608-628. DOI: 10.1123/jsep.29.5.608

Boardley, I. D., & Kavussanu, M. (2008). The Moral Disengagement in Sport Scale-Short. *Journal of Sports Sciences*, *26*(14), 1507-1517. DOI: 10.1080/02640410802315054

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466. DOI: 10.1037/0033-2909.105.3.456

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. DOI: 10.1207/S15328007SEM0902_5

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309-319. DOI: 10.1037/1040-3590.7.3.309

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126. DOI: 10.1037/0022-3514.44.1.113

Dunn, J. G., Bouffard, M., & Rogers, W. T. (1999). Assessing item content-relevance in sport psychology scale-construction research: Issues and recommendations. *Measurement in Physical Education and Exercise Science, 3*(1), 15-36.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*(6), 358-361.

Enders, C.K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine, 68*(3), 427–436. DOI: 10.1097/01.psy.0000221275.75056.d8

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. DOI: 10.1037/1082-989X.4.3.272

Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. A. Berk, (Ed.), *Criterion-referenced measurement: The state of the art* (pp. 80-123). Baltimore, MD: The Johns Hopkins Press.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. DOI: 10.1037/1040-3590.7.3.238

de Hon, O., Kuipers, H., & van Bottenburg, M. (2015). Prevalence of doping use in elite sports: A review of numbers and methods. *Sports Medicine, 45*(1), 57-69. DOI: 10.1007/s40279-014-0247-x

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

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*(1), 111-117.

Kavussanu, M., & Boardley, I. D. (2009). The prosocial and antisocial behavior in sport scale. *Journal of Sport and Exercise Psychology, 31*(1), 97-117. DOI: 10.1123/jsep.31.1.97

Kavussanu, M., Hatzigeorgiadis, A., Elbe, A. M., & Ring, C. (2016). The moral disengagement in doping scale. *Psychology of Sport and Exercise, 24*, 188-198. DOI: 10.1016/j.psychsport.2016.02.003

Kavussanu, M., Stanger, N., & Boardley, I. D. (2013). The Prosocial and Antisocial Behaviour in Sport Scale: Further evidence for construct validity and reliability, *Journal of Sports Sciences, 31, 11*, 1208-1221, DOI: 10.1080/02640414.2013.775473

Kline, R. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: The Guilford Press.

Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*(3), 193-205. DOI: 10.1023/A:1015291021312

Lucidi, F., Grano, C., Leone, L., Lombardo, C., & Pesce, C. (2004). Determinants of the intention to use doping substances. *International Journal of Sport Psychology, 35*(2), 133-148.

Lucidi, F., Zelli, A., Mallia, L., Grano, C., Russo, P. M., & Violani, C. (2008). The social-cognitive mechanisms regulating adolescents' use of doping substances. *Journal of Sports Sciences*, *26*(5), 447-456. DOI: 10.1080/02640410701579370

Marschall, D. E., Saftner, J., & Tangney, J. P. (1994). *The State Shame and Guilt Scale*. Fairfax, VA: George Mason University.

Marsh, H.W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, *24*(1), 17-39. DOI: 10.1111/j.1745-3984.1987.tb00259.x

McVeigh, J. & Begley, E. (in press). Anabolic steroids in the UK: An increasing issue for public health. *Drugs: Education, Prevention and Policy*. DOI: 10.1080/09687637.2016.1245713.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

Momaya, A., Fawal, M., & Estes, R. (2015). Performance-enhancing substances in sports: A review of the literature. *Sports Medicine*, *45*(4), 517-531. DOI: 10.1007/s40279-015-0308-9

Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior*, *29*(4), 371-393. DOI: 10.1007/s10979-005-4930-1

Pedhazur, E. J., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pope, H. G., Kouri, E. M., & Hudson, J. I. (2000). Effects of supraphysiological doses of testosterone on mood and aggression in normal men: A randomized controlled trial. *Archives of General Psychiatry, 57*(2), 133-140. DOI: 10.1001/archpsyc.57.2.133

Pope, H. G., Wood, R. I., Rogol, A., Nyberg, F., Bowers, L., & Bhasin, S. (2013). Adverse health consequences of performance-enhancing drugs: An Endocrine Society scientific statement. *Endocrine Reviews, 35*(3), 341-375. DOI: 10.1210/er.2013-1058.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*, 1-15.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552–566. DOI: 10.1037/0033-2909.114.3.552

Schulte, H. M., Hall, M. J., & Boyer, M. (1993). Domestic violence associated with anabolic steroid abuse. *The American Journal of Psychiatry, 150*(2), 348. DOI: 10.1176/ajp.150.2.348a

Sjöqvist, F., Garle, M., & Rane, A. (2008). Use of doping agents, particularly anabolic steroids, in sports and society. *The Lancet*, *371*(9627), 1872-1882. DOI: 10.1016/S0140-6736(08)60801-6

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

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds), *Secondary data analysis: An introduction for psychologists*, (pp. 39-61). 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	*df*	Rχ^2	RCFI	SRMR	RMSEA
		Sample 1 (N = 318)			
DMDS Models					
1. M1a, 35 items	545	1183.39	.873	.050	.061
2. M1b, 18 items	120	222.54	.955	.034	.052
3. 2nd order, M1b	129	333.48	.910	.052	.071
Alternative DMDS Models					
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
DMDS-S Models					
7. M5a, 6 items	9	21.35	.968	.031	.066

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

Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	.87/.92	.50/.39
An athlete shouldn’t be blamed for doping if a member of his/her training group has encouraged it.	.92/.94	.39/.33
An athlete shouldn't be held responsible for doping if his/her coach encouraged him/her to do it.	.88/.92	.47/.39
Diffusion of Responsibility		
If most athletes in a sport dope, no one athlete should be held responsible for doing it.	.77/.81	.64/.59
It’s not right to condemn individuals who dope when many in their sport are doing the same.	.87/.92	.50/.38
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
Distortion of Consequences		
Risks associated with doping are exaggerated.	.85/.86	.53/.52
Doping doesn’t really harm anyone else.	.70/.77	.72/.64
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 (*mechanism*)	Factor Loading	Error Variance
DMDS-S		
Doping is okay if it helps an athlete advise others on how to do it right (*moral justification*).	.61/.86	.79/.52
Using terms such as "gear" or "juice" makes doping sound less harmful (*euphemistic labelling*).	.44/.52	.90/.86
Compared to most lifestyles in the general public, doping isn't that bad (*advantageous comparison*).	.70/.76	.72/.65
Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it (*displacement of responsibility*).	.76/.88	.65/.48
It’s not right to condemn individuals who dope when many in their sport are doing the same (*diffusion of responsibility*).	.83/.88	.56/.48
Risks associated with doping are exaggerated (*distortion of consequences*).	.67/.74	.74/.68
DSRES		
…resist doping even if your training group encouraged you to do it?	.84/.83	.55/.56

…resist doping even if you knew you could get away with it?	.82/.83	.57/.55
…ignore the temptation to dope even if you knew it would improve your performance?	.86/.88	.51/.48
…resist peer pressure to dope?	.80/.87	.60/.50
…reject doping even if most of your training partners did it?	.84/.87	.54/.50
…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	*df*	Rχ^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 Hardcore	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 Hardcore	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 Hardcore	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) Subscales with Empathy and Anticipated Guilt

	Sample 1 (N = 318)		Sample 2 (N = 300)	
DMDS Subscale	Empathy	Guilt	Empathy	Guilt
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. Correlation significant at $p < .01$ unless indicated by ns, where $p > .05$.

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