Social Science Research Grant (2014-2016)

Designing and Validating Measures of Doping Moral Disengagement and Self-Regulatory Efficacy, and Assessing a Model of Doping Behaviour

Report to WADA submitted on November 29 2016 by:

DR IAN D. BOARDLEY (PRINCIPAL INVESTIGATOR), PROFESSOR ALAN L. SMITH (CO-INVESTIGATOR), DR JONATHAN GRIX (CO-INVESTIGATOR), MR CERI WYNNE (RESEARCH ASSOCIATE), MR ANTON ARO (RESEARCH ASSOCIATE)

Table of Contents

List of Tables

List of Figures

Executive Summary

Evidence associating doping behavior with Moral Disengagement (MD) has accumulated over recent years (e.g., Boardley & Grix, 2014; Boardley, Grix, & Dewar, 2014; Boardley, Grix, & Harkin, 2015; Lucidi, Grano, Leone, Lombardo, & Pesce, 2004; Lucidi, Zelli, Mallia, Grano, Russo, & Violani, 2008; Zelli, Mallia, & Lucidi, 2010). MD involves the conditional endorsement of transgressive conduct, and can occur through any of eight methods as described by Bandura (1991), six of which have been consistently linked with doping (Boardley & Grix, 2014; Boardley et al., 2014, 2015). However, research examining links between MD and doping has either been qualitative (Boardley & Grix, 2014; Boardley et al., 2014, 2015), or has been conducted on convenience samples and hadn't considered MD alongside other key variables from Bandura's (1991) theory (e.g., Lucidi et al., 2004, 2008; Zelli et al., 2010). As such, there is a need to quantitatively examine the predictive abilities of MD on doping alongside other key elements of Bandura's (1991) theoretical framework in samples drawn from key sport and exercise populations. One of these variables is Self-Regulatory Efficacy (SRE), which when applied to doping represents a person's capacity to withstand personal and social influences that encourage the use of Performance Enhancing Drugs. However, to conduct such work valid and reliable measures of doping MD and doping SRE need to be developed. To this end, this report presents the findings from a line of research that had two key aims. First, we aimed to develop valid and reliable instruments to measure doping MD and doping SRE in sport and exercise contexts. Second, we sought to test a model of doping behavior grounded in Bandura's (1991) theory with team- and individual-sport participants and corporate- and hardcore-gym exercisers.

To achieve the aims of this project, data were collected from three samples. The first two were utilized in both the scale development and model-testing aspects of the project, whereas the third was specifically used to examine the test-retest reliability of the new measures. Participants in the first two samples were specifically recruited to represent male and female athletes and exercisers from the four key sport and exercise contexts outlined above. Across samples 1 and 2 as a whole, 610 athletes and exercisers participated and overall self-reported lifetime prevalence of doping was 13.6%. Following a complete and robust set of scale-development procedures, we developed and validated two measures of doping MD (i.e., the Doping Moral Disengagement Scale [DMDS] and Doping Moral Disengagement Scale – Short [DMDS-S]) and one of doping SRE (i.e., the Doping Self-Regulatory Efficacy Scale [DSRES]). Supporting our hypotheses, in developing and validating the three scales we identified six sub-dimensions of doping MD and one overall dimension of doping SRE; this dimensionality was consistent across gender and sport/exercise context. Then, during the testing of our hypothetical model of doping behavior, we again found support for all our study hypotheses. Specifically, we established: (a) empathy and doping SRE negatively predicted reported doping; (b) the predictive effects of empathy and doping SRE on reported doping were mediated in part by doping MD and anticipated guilt; (c) doping MD positively predicted reported doping; (d) the predictive effects of doping MD on reported doping were partially mediated by anticipated guilt; and (e) these predictive effects were largely invariant across gender and sport/exercise contexts. Overall, this project has made major contributions to the improvement of anti-doping research by designing and validating three psychometric tools that can be applied in future research on doping, and has contributed important knowledge to our understanding of the psychosocial factors that may lead to doping by testing a behavioral model of doping behavior that can inform future research as well as the design and delivery of anti-doping education programmes and intervention strategies.

Introduction

Sport can bring joy and fulfillment to both participants and spectators. However, such joy and fulfillment can be undermined by performers who seek to gain a performance advantage over their competitors through use of prohibited performance enhancing substances or methods, often referred to as doping. Although accurate prevalence rates are difficult to determine, a recent article summarizing the best available evidence estimated prevalence of doping in adult elite sport to be between 14 and 39% (de Hon, Kuipers, & van Bottenburg, 2015). An important aim for researchers investigating doping is to identify and understand psychosocial factors that influence the likelihood of athletes and exercisers using illicit performance enhancing substances. The current project sought to extend previous research on doping in sport and exercise and is underpinned by the social cognitive theory of moral thought and action (Bandura, 1991).

Bandura (1991) proposed transgressive activities – such as doping – are deterred when people anticipate experiencing negative emotional reactions (e.g., guilt) as a result of engaging in them. As doping constitutes cheating, athletes and exercisers should be deterred from engaging in it if they anticipate feeling guilty as a result. However, Bandura (1991) also explained how people can reduce or eliminate anticipation of such negative emotional reactions through use of any of eight psychosocial mechanisms collectively referred to as mechanisms of Moral Disengagement (MD). Representing the conditional endorsement of transgressive acts, MD may facilitate doping by allowing athletes and exercisers to use prohibited substances or methods without experiencing negative emotional reactions such as guilt or shame.

Consistent with this possibility, qualitative research has shown how athletes with experience of doping morally disengage when asked to explain their reasons for doping. First, Boardley and Grix (2014) conducted semi-structured interviews with nine (eight male, one female) PED-using bodybuilders. The interviews centered on psychosocial processes facilitating doping and deductive content analysis revealed evidence of six of the eight MD mechanisms. One weakness of this study was that participants all originated from a single gym. To address this, Boardley, Grix and Dewar (2014) conducted a WADA-funded follow-up study with 64 male bodybuilders with experience of doping from across England. Consistent with the initial study, deductive content analysis of the study data revealed evidence of the same six MD mechanisms. Although presenting consistent evidence for the relevance of six MD mechanisms for doping, these first two studies were limited by the fact they only involved bodybuilders. To address this limitation, Boardley, Grix and Harkin (2015) interviewed twelve male athletes with experience of doping from a variety of team and individual sports. Subsequent data analyses again showed evidence of the same six MD mechanisms. Therefore, across three studies Boardley and colleagues provided consistent evidence that doping athletes from a range of sport and exercise contexts use the same six mechanisms to morally disengage when explaining their PED use.

The first of the six relevant mechanisms – moral justification – occurs when harmful activities are made personally and socially acceptable through their portrayal as achieving commendable social or moral purposes. The second – euphemistic labelling – diminishes the damaging nature of actions through palliative and/or convoluted language. The third – advantageous comparison – makes detrimental conduct appear less damaging by comparing the act to more heinous acts. The fourth – displacement of responsibility – diminishes personal responsibility for transgressive action and/or its consequences by proffering it resulted from implicit or explicit social pressures. The fifth – diffusion of responsibility – also acts by diminishing personal accountability for harmful acts and/or their outcomes, but this time though group decision making (i.e., a group collectively taking the decision to engage in injurious conduct) or group action (i.e., a group collectively engaging in harmful conduct). The final mechanism – distortion of consequences – occurs when the perpetrator of a transgressive act avoids information on the harm caused, and/or downplays its significance.

Quantitative research has also established links between MD and doping. More specifically, researchers have identified positive links between MD, intention to dope, and reported doping across a small number of quantitative studies (e.g., Lucidi, Grano, Leone, Lombardo, & Pesce, 2004; Lucidi, Zelli, Mallia, Grano, Russo, & Violani, 2008); Zelli, Mallia, & Lucidi, 2010). However, it is important to note these studies were conducted with convenience samples of high-school students, a significant proportion of whom (43.0 - 45.2%) did not partake in extracurricular sport. As such these findings may not extend to participants from key sport and exercise contexts, including those in which prevalence rates for doping are likely to be higher. Also, the predictive effects of MD on doping outcomes were examined alongside those of other variables, primarily drawn from the Theory of Planned Behavior (Ajzen, 1991). As such, the predictive effects of MD on these outcomes may have been impacted by the incorporation of these variables during model testing. Further, key variables (e.g., anticipated guilt) from Bandura's (1991) theory were not included in model testing. Thus, although these studies have been helpful in establishing statistical associations between MD and doping outcomes, research is needed that specifically targets participants from key sport and exercise contexts and examines the effects of MD on doping within models that include all of the key aspects of Bandura's (1991) theory.

As alluded to earlier, Bandura (1991) suggests MD impacts upon transgressive behaviors through its effect on regulatory emotions such as guilt. Guilt represents a distasteful emotional state experienced as tension and regret resulting from empathic feelings for someone suffering anguish, combined with recognizing one's personal responsibility for the anguish (Hoffman, 2000). Due to its unpleasant connotations, guilt can be adaptive in regulating harmful conduct, as people are deterred from engaging in behaviors they anticipate will result in guilt (Bandura, 1991). Support for the adaptive role of guilt is evidenced by negative relationships between proneness to experience guilt and aggression (Stuewig, Tangney, Heigel, Harty, & McCloskey, 2010). Therefore, athletes and exercisers anticipating guilt as a result of PED use may be less like to adopt doping practices than those who do not. Importantly, anticipation of guilt is thought to be diminished by MD, which involves portraying transgressions favorably, reducing personal accountability for them and downplaying their detrimental consequences (Bandura, 1991; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). Anticipation of guilt is unlikely when one views detrimental behavior in such ways. Importantly, work in and out of sport (Bandura et al., 1996; Stanger, Kavussanu, & Ring, 2012) supports the notion that anticipation of guilt may be reduced by MD. Thus, any effect of MD on doping may be mediated – at least in part – by reductions in anticipated guilt.

As well as potential outcomes, it is also important to consider potential antecedents of MD such as empathy. Empathy represents the tendency to vicariously experience other individuals' emotional states, and is thought to incorporate both emotional and cognitive components (Davis, 1983; 1994). A lack of empathy implies an inability to view the world from other individuals' perspective or to feel sympathy toward them (Davis, 1994). Conversely therefore, increases in empathy may inhibit transgressions because they heighten peoples' ability to understand and experience the distress that may be caused by such actions (Bandura, 1986, Feshbach, 1975; Hoffman, 2000). Such effects of empathy may occur through changes in MD, as empathy is thought to impair MD as endorsement of deleterious conduct is more difficult when one can anticipate and experience the consequences of one's actions for others (Bandura, 1991; Paciello, Fida, Cerniglia, Tramontano, & Cole, 2013). Consistent with these propositions, negative relationships between empathy and transgressive conduct in sport have been demonstrated (Kavussanu, Stamp, Slade, & Ring, 2009; Stanger et al., 2012), and research out of sport has negatively linked empathy and MD (Paciello et al., 2013). Thus, higher levels of empathy in athletes and exercisers should be associated with lower levels of MD, which may in turn be linked with a reduced likelihood to dope.

Another variable that may influence athletes' and exercisers' MD is self-regulatory efficacy (SRE), which represents one's ability to resist personal and social pressures to engage in detrimental conduct (Bandura, Caprara, Barbaranelli, Pastorelli, & Regalia, 2001). Importantly, Bandura et al. (2001) proposed increased SRE should lead to lower levels of MD, because those who have strong beliefs in their ability to resist incentives to engage in harmful conduct have no need to develop the abilities to justify and rationalize them. When specifically applied to doping, SRE represents one's capacity to withstand personal and social influences encouraging the use of performance enhancing substances/methods. In accord with Bandura's theorizing, research by Lucidi and colleagues – introduced earlier – has shown that doping SRE negatively predicts intention to dope in Italian adolescents (Lucidi et al., 2008; Zelli et al., 2010). However, as mentioned earlier the findings of these studies are limited due to their use of convenience samples of high-school students, many of whom (43-45.2%) were not partaking in extracurricular sport. Additionally, Lucidi and colleagues considered the predictive abilities of MD alongside SRE rather than as a mediator of its prediction of doping, which is inconsistent with the causal pathway hypothesized and empirically supported by Bandura et al. (2001). As such, research testing relationships accurately grounded in theory and specifically sampling participants from key sport and exercise contexts is needed.

One of the major objectives of the current project was to test a model of doping behavior grounded in Bandura's (1991) theory with team- and individual-sport participants and corporate- and hardcore-gym exercisers. Based on the arguments presented to this point, we aimed to test a process model (see Figure 1) whereby empathy and SRE negatively predicted MD, MD subsequently negatively predicted anticipated guilt, and anticipated guilt then negatively predicted reported doping (Bandura, 1986, 1991; Bandura et al., 1996; Bandura et al., 2001; Lucidi et al., 2008; Paciello et al., 2013; Stanger et al, 2012; Zelli et al., 2010). Further we anticipated that SRE and empathy would negatively predict doping indirectly via changes in MD and anticipated guilt. Finally, MD was expected to positively predict doping both directly (i.e., to account for any predictive effects of MD on doping that don't operate through guilt) and through a mediated effect via anticipated guilt (Bandura, 1991; Bandura et al., 1996). Unfortunately, existing psychometric measures currently available to assess some of the constructs in our hypothesized model have questionable validity.

For instance, there are numerous concerns regarding the validity of the scales developed by Lucidi et al. (2008) to assess doping MD and doping SRE. First, items were developed based on interviews with 35 high-school students who regularly practiced sport. To develop items, during the interviews participants were asked to spontaneously list situations in which (a) doping would or should not be completely condemned (MD) and (b) doping would be more likely (SRE). For the MD measure, the frequencies of common situations were then summed and categorized into the MD mechanisms evoked, leading to the selection of a total of 21 items. The researchers then selected six of these 21 items for inclusion in their measure, although no information was provided on what process was used to select these items. For the SRE measure, even less information was provided on item development, with the researchers only describing how ten items were developed based on situations described during the interviews. Thus, the items for these measures 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 the intended population/s (Clark & Watson, 1995), (b) not appraised for their content validity when this should be a key aspect of the scale-development process (Haynes, Richard, & Kubany, 1995) and (c) selected using a process not made clear by the researchers. It should therefore perhaps not be that surprising that two of the six 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 their proposed mechanisms. Further, in the doping SRE measure there is an imbalance in items pertaining to personal influences (eight items) compared to social influences (two items). In addition, the factor structure of new scales should be confirmed in at least two samples (Fabrigar, Wegener, MacCallum, & Strahan, 1999), which Lucidi and colleagues (2008) did not do when developing these scales. Finally, information on how the factor structure of the scales was determined was not provided. Although factor loadings were provided, no actual assessment of the scales' dimensionalities was reported. Currently it is therefore unclear whether the two scales were unidimensional (as apparently assumed by the researchers), or if they contained more than one lower-order factor subsumed under higher-order factors.

Recently, Kavussanu, Hatzigeorgiadis, Elbe and Ring (2016) developed a further measure of doping MD. Although this measure was developed through a more rigorous process than that created by Lucidi et al. (2008), issues remain regarding the content/face validity of some items, as well as its applicability across all contexts relevant to doping. With respect to content validity, one item (i.e., 'Doping is just a way to "maximize your potential"') – similar in content to a problematic item in the scale of Lucidi et al. (2008) – does not represent its proposed mechanism. Also, doping is an issue not just in sport, but also in exercise contexts (Sjöqvist, Garle, & Rane, 2008). As such, any measure of doping MD should ideally be appropriate for use in both sport and exercise contexts so comparisons between the two domains can be made. However, two of the items in the scale developed by Kavussanu et al. (2016) make direct reference to 'sport' and 'teammates' (i.e., 'Players cannot be blamed for doping if their teammates pressure them to do it' and 'A player should not be blamed for doping if everyone on the team is doing it'), rendering the measure less relevant to non-team-sport athletes. The researchers attempted to address this by administering an adapted (i.e., replacing 'team' with 'club') version of the scale to individual-sport athletes, but having two versions of the same scale is clearly problematic when it comes to comparing findings across studies. A better approach would have been to develop items that were suitable for use in both contexts from the outset. Also, none of the three items above resonate with data from qualitative studies (Boardley & Grix, 2014; Boardley et al., 2014, 2015) that have specifically investigated *how* sport and exercise participants *actually* morally disengage, suggesting these items also have questionable face validity. Finally, a key consideration when developing psychometric instruments is that the samples used for validation are representative of the target population (Clark & Watson, 1995). As such, researchers should ensure samples used for validating doping-specific measures include participants with actual experience of doping. Unfortunately, Kavussanu and colleagues failed to report prevalence of doping in any of the samples they used so it is unknown how many – if any – of the athletes recruited had actual experience with doping. Given the concerns outlined above regarding the validity and applicability of the currently available measures of doping MD and doping SRE, a second major objective of the current project was to develop valid and reliable instruments to measure these constructs across sport and exercise contexts.

The Current Project

The current project was specifically designed to achieve the two primary objectives outlined above, and in doing so answer a number of research questions and test several hypotheses. The research questions we sought to answer were: (a) what is the dimensionality of doping MD and doping SRE across a range of key sport and exercise populations? (b) do empathy and SRE predict reported doping across key sport and exercise populations, and if so, do doping MD and anticipated guilt mediate these predictions? and (c) does doping MD predict reported doping across key sport and exercise populations and, if so, does anticipated guilt mediate this prediction? Based on these research questions, we sought to test the following a priori hypotheses: (a) doping MD will be multidimensional (Bandura, 1991), (b) doping SRE will be unidimensional (e.g., Bandura et al., 2001), (c) the dimensionality of doping MD and doping SRE will be invariant across key sport and exercise contexts (Boardley & Kavussanu, 2008; Pastorelli, Caprara, Barbaranelli, Rola, Rozsa, & Bandura, 2001), (d) empathy and doping SRE will negatively predict reported doping across key sport and exercise populations, (e) the predictive effects of empathy and doping SRE on reported doping will be partially mediated by doping MD and anticipated guilt (Bandura, 1986; Bandura et al., 2001; Lucidi et al., 2008; Paciello et al., 2013; Zelli et al., 2010), (f) doping MD will positively predict reported doping across key sport and exercise populations, and (g) the effect of doping MD on reported doping will be partially mediated by anticipated guilt (Bandura, 1991; Bandura et al., 1996; Stanger et al., 2012).

Method

To address the aims of the current project, data were collected from two primary samples for the main analyses plus one further sample for examining test-retest reliability. The data collected from the primary samples were grouped differently when addressing each of the two main project aims. As such, both the method and results sections are separated into two sections, one addressing the first main objective of the project (i.e., questionnaire development) and another addressing the second (i.e., structural model testing).

Objective 1 - Questionnaire Development

To allow assessment of some of the key constructs of interest, in the current research we sought to develop three psychometric scales. Specifically, we developed long and short measures of doping moral disengagement termed the Doping Moral Disengagement Scale (DMDS) and Doping Moral Disengagement Scale – Short (DMDS-S), respectively, as well as a measure of doping SRE termed the Doping Self-Regulatory Efficacy Scale (DSRES). The development of these measures followed established procedures for the development of psychometric instruments (i.e., Clark & Watson, 1995; Fabrigar et al., 1999; Haynes et al., 1995). The specific application of these procedures within the current project can be categorized into the following distinct phases: (a) item development, (b) construct validity and (c) test-retest reliability.

Item development. The first stage of scale development for both measures involved the development of a large pool of items (see Clark & Watson, 1995). This process was informed by a review of existing measures of the target constructs, qualitative papers that have investigated the constructs in physical-activity contexts and consultation with relevant experts. The resultant items were then piloted with a convenience sample of sport and exercise participants. Item analyses were conducted on these pilot data to ensure the bivariate correlations between items were in the target range (i.e., r = .15-.50); items that did not perform as expected were eliminated or adapted.

These item pools were then subjected to content validity assessment to determine whether items characterize their intended domain (Haynes et al., 1995). Content validity is typically examined through expert opinion (Kline, 2015), and as such both item pools were evaluated by 12 sport-psychology professionals experienced in scale development. Items with poor content validity were either eliminated or adjusted. The final item pools were then pilot-tested with a second convenience sample of sport and exercise participants to (a) ensure items were still correlated in the target range (i.e., r = .15-.50) with other items measuring the same construct, (b) determine the difficulty of the items and (c) obtain qualitative feedback on item wording (Clark & Watson, 1995). Any further necessary adjustments to item wording were then made.

Construct validity. Construct validity is a broad concept representing whether scores collected with an instrument represent those of the target construct (Kline, 2015). Construct validity has numerous aspects, and evidence for a measure's construct validity accumulates as evidence is established for each of these aspects. To provide initial evidence for the construct validity of the new measures, we sought to establish evidence for the factorial, convergent and discriminant validity and internal reliability of both measures in two separate samples. 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 then confirming it in the second. Evidence for convergent validity is established if a scale correlates at least moderately with established measures of variables within the target construct's nomological network (i.e., variables with which theoretically it should be related with; Vaughn & Daniel, 2012). As such, DMDS and DSRES associations with empathy and anticipated guilt were computed to provide evidence for the scales' convergent validities. In contrast, discriminant validity represents the degree to which a construct is distinguishable from closely related but conceptually distinct constructs (Vaughn & Daniel, 2012). To establish evidence for the DMDS's discriminant validity, we first analyzed its correlation with sport MD, before then computing the interrelationships amongst the different subscales of the DMDS. Similarly, for the DSRES we computed its correlation with peer pressure SRE. Finally, internal reliability is another important aspect of construct validity. The internal reliabilities of the two instruments plus the subscales of the DMDS were estimated using Cronbach's alpha.

***Sample 1 participants*.** Participants were team (n = 181) and individual (n = 70) sport or hardcore (n = 44) or corporate (n = 23) gym attendees representing both genders (n_{male} = 203; n_{female} = 115); ages ranged from 16 to 70 years (M = 23.33, SD = 8.20). Participants had been training/competing for an average of 7.27 years (SD = 5.41), spent an average of 8.64 hours (SD = 3.91) per week training, and had trained in their current gym/with their current team for an average of 4.14 years (SD = 4.51).

Sample 2 participants. Participants were team (n = 14) and individual (n = 99) sport or hardcore (n = 81) or corporate (n = 98) gym attendees representing both genders (n_{male} = 168; n_{female} = 124); ages ranged from 17 to 73 years (M = 29.46, SD = 12.37). Participants had been training/competing for an average of 9.02 years (SD = 8.54), spent an average of 7.90 hours (SD = 5.01) per week training, and had trained in their current gym/with their current team for an average of 3.78 years (SD = 4.97).

Measures. In addition to those being developed, a series of further instruments were administered. This was to facilitate the aforementioned analyses pertaining to construct validity, as well as for use in the structural model testing relating to the second major project objective. Specifically, measures of sport MD, peer pressure SRE, guilt, empathy, doping susceptibility and reported doping were administered.

Sport moral disengagement. The moral disengagement in sport scale-short (MDSS-S; Boardley & Kavussanu, 2008) was used to measure participants' sport MD. This scale consists of eight items (e.g., "Insults among players do not really hurt anyone") that assess MD in sport. Participants were asked to read a number of statements describing thoughts and feelings that athletes may have and indicate their level of agreement with each statement using a Likert scale anchored by 1 (*strongly disagree*) and 7 (*strongly agree*). The scale has shown very good levels of internal consistency and evidence for its factorial and convergent validity has been reported (Boardley & Kavussanu, 2008).

Peer pressure SRE. The SRE scale (Bandura et al., 2001) was used to assess peer pressure SRE. This measure consists of five items (e.g., "Resist peer pressure to drink beer, wine or alcohol") that 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 in their ability to engage in relevant behaviors using a Likert scale anchored by 1 (*no confidence*) and 5 (*complete confidence*). The scale has shown good levels of internal consistency and evidence for its construct validity has been reported (Bandura et al., 2001).

Guilt. To assess participants' anticipated guilt responses 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 were then asked to indicate how they would anticipate feeling about continuing to take the substance by responding to the five items (e.g., "I would feel remorse, regret") that form the guilt scale in 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 construct validity and internal reliability of this sub-scale.

Empathy. Scores on 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 were asked to indicate how well the statements described them and responded on a scale with anchors of 1 (*does not describe me well*) and 7 (*describes me very well*). This scale has been used in past research, and has shown very good internal consistency (Carlo, Raffaelli, Laible, & Meyer, 1999).

Reported doping. Our approach to the assessment of reported doping was based on the method used by Lucidi et al. (2008). More specifically, participants were provided with a list of doping substances and asked to indicate which ones they currently used, had used in the past three months, had used prior to the past three months or never used. The list of doping substances was based on those banned in sport by WADA, and included anabolic-androgenic steroids, peptide hormones and stimulants. Participants' responses were used to form a score from one to four, with participants being assigned a one if they indicated never using any of the substances, two if they had used any one of them but only prior to the past three months, three if they had used any one of them in the past three months and four if they currently used any of the substances.

Susceptibility to doping. Participants' susceptibility to doping was assessed using the approach of Gucciardi, Jalleh, & Donovan, 2010). This first involved presenting the following scenario: "If you were offered a banned performance enhancing substance under medical supervision at low or no financial cost and the banned performance-enhancing substance could make a significant difference to your performance and was currently not detectable". Then, participants were asked to report how much consideration they would you give to this offer on a scale from 1 (*none at all*) to 7 (*a lot of consideration*). Previous research has validated this method of assessing susceptibility to doping (Gucciardi et al., 2010).

Procedure. Recruitment for the project commenced once ethical clearance was provided by the ethics committee of the PI's institution. Our approach to recruitment differed for sport versus exercise participants. For sport participants, we contacted the team- and individual-sport coaches regarding participation of the athletes they coached in the study. For coaches who agreed to allow access to their athletes we arranged a designated training session during which we introduced the project to athletes and invited them to participate. In contrast, exercise participants were recruited through managers of hardcore and corporate gymnasia who were contacted and asked whether it would be possible to invite exercisers at their gymnasia to participate in the study. Once access was agreed through gymnasia managers, potential participants were approached in the reception area of gymnasia and invited to participate. Before completing the questionnaire, all respondents were informed that the survey examined sporting attitudes and that honesty in responses was vital to the success of the study. It was also explained that all responses would be kept strictly confidential and would be used only for research purposes. Participants signed an informed consent form prior to completing the questionnaire, which took approximately 10–15 minutes to complete for each participant. All recruitment and data collection was conducted by one of two research associates. Data were collected across two phases, with approximately equal sample sizes across the two phases. The data from the first phase (i.e., Sample 1) was analyzed before the second phase of data collection (i.e., Sample 2) commenced; this allowed for any necessary adjustments to item content and/or item generation between the two data-collection phases.

Test-retest reliability. Test-retest reliability represents the temporal reproducibility of scores obtained using a scale, or an instrument's ability to provide consistent scores across time in a population with stable scores on the construct being assessed (Aaronson et al., 2002). It is usually examined by administering a measure to the same people twice and correlating the two sets of scores (Pedhazur & Schmelkin, 1991). Given both doping MD and doping SRE should be quite stable over the short term, good levels of test-retest reliability (i.e., ≥ 0.70) for the two measures would provide evidence for their temporal reproducibility.

Participants. Participants were team- (n = 9) and individual- (n = 78) sport athletes and hardcore- (n = 5) or corporate- (n = 9) gym exercisers representing both genders (n_{male} = 60; n_{female} = 41); ages ranged from 16 to 70 years (M = 35.19, SD = 13.46). Participants had been training/competing for an average of 9.35 years (SD = 7.44), spent an average of 7.87 hours (SD = 4.76) per week training, and had trained in their current gym/with their current team for an average of 4.38 years (SD = 5.02).

Procedure. For the purposes of estimating the test-retest reliability of the two new measures, we collected data from a separate sample on two occasions. To facilitate the collection of data on two occasions, we collected data using both paper and online versions of the two instruments; the online questionnaires were hosted by Lime Survey. Face-to-face recruitment of participants and subsequent data collection at Time 1 matched that used during data collection for the construct-validity samples. Recruitment of online participants involved advertising the study to individual- and team-sport athletes and hardcore- and corporate-gym exercisers through sport club and gymnasia websites and discussion groups, social media, and personal contacts. These adverts included the main details on what participation involved and a link to the survey webpage. Potential participants who clicked on this link were then provided with full details of the study, including their rights as participants as well as the purpose and procedure of the study. The front page of the survey also emphasized that participation was voluntary and all information provided would be fully confidential. Prior to starting the 10-minute online questionnaire, participants were informed that by clicking on the link to start the questionnaire they were providing their informed consent to participate.

In terms of the recommended interval between administrations when estimating the test-retest reliability of a measure, relatively short intervals of approximately one to two weeks are recommended (Pedhazur & Schmelkin, 1991). This is to ensure that any differences in scores between the two administrations are largely due to random measurement error rather than actual changes in participants' scores on the constructs being assessed. As such, we aimed for an inter-administration interval of between nine and 16 days. For face-to-face participants, arrangements were made to collect data again within this window, and for online participants, reminder emails were sent nine days following initial completion, and continued until 16 days afterwards; the online questionnaire closed for participants who had not completed their second administration by 16 days after their first completion.

Objective 2 - Structural Model Testing

The analyses conducted to achieve the second objective of the project utilized a number of different sample combinations. First, for the purposes of testing and confirming the hypothesized model in Figure 1 and testing mediation, we utilized Sample 1 and Sample 2 as described previously. However, to maximize sample size and balance sub-groups as much as possible we combined these two samples when conducting the multigroup analyses. These analyses involved separating this combined sample into males and females for the gender invariance analyses, and into individual-sport, team-sport, hardcore-gym and corporate-gym subgroups for the sport/exercise group invariance testing. The characteristics of the combined sample and these six sub-divisions of it are described below.

***Combined sample participants*.** Participants were team (n = 195) or individual (n = 169) sport or hardcore (n = 125) or corporate (n = 121) gym attendees representing both genders (n_{male} = 371; n_{female} = 239); ages ranged from 16 to 73 years (M = 26.27, SD = 10.84). Participants had been training/competing for an average of 8.10 years (SD = 7.12), spent an average of 8.29 hours (SD = 4.48) per week training, and had trained in their current gym/with their current team for an average of 3.97 years (SD = 4.73). 527 (86.4%) participants reported never having used performance enhancing drugs, 46 (7.5%) had used them prior to the past three months, 20 (3.3%) had used them in the past three months and 17 (2.8%) were currently using them.

***Combined sample male participants*.** Participants were team (n = 135) or individual (n = 88) sport or hardcore (n = 102) or corporate (n = 46) gym attendees; ages ranged from 16 to 73 years (M = 25.91, SD = 10.49). Participants had been training/competing for an average of 7.81 years (SD = 7.56), spent an average of 8.68 hours (SD = 4.53) per week training, and had trained in their current gym/with their current team for an average of 3.90 years (SD = 4.65). 304 (81.9%) participants reported never having used performance enhancing drugs, 36 (9.7%) had used them prior to the past three months, 17 (4.6%) had used them in the past three months and 14 (3.8%) were currently using them.

Combined sample female participants. Participants were team (*n* = 60) or individual (*n* = 81) sport or hardcore (*n* = 23) or corporate (*n* = 75) gym attendees; ages ranged from 18 to 65 years (*M* = 26.82, *SD* = 11.35). Participants had been training/competing for an average of 8.54 years (*SD* = 6.38), spent an average of 7.67 hours (*SD* = 4.34) per week training, and had trained in their current gym/with their current team for an average of 4.07 years (*SD* = 4.87). 223 (93.3%) participants reported never having used performance enhancing drugs, 10 (4.2%) had used them prior to the past three months, three (1.3%) had used them in the past three months and three (1.3%) were currently using them.

Combined sample individual-sport participants. Participants were male (*n* = 88) or female (*n* = 81) individual-sport athletes; ages ranged from 18 to 63 years (*M* = 26.68, *SD* = 11.30). Participants had been training/competing for an average of 8.26 years (*SD* = 7.51), spent an average of 10.25 hours (*SD* = 4.77) per week training, and had trained with their current club for an average of 4.41 years (*SD* = 5.00). 159 (94.1%) participants reported never having used performance enhancing drugs, seven (4.1%) had used them prior to the past three months, one (0.6%) had used them in the past three months and two (1.2%) were currently using them.

Combined sample team-sport participants. Participants were male (*n* = 135) or female (*n* = 60) team-sport athletes; ages ranged from 17 to 41 years (*M* = 20.51, *SD* = 2.61). Participants had been training/competing for an average of 6.99 years (*SD* = 5.14), spent an average of 7.73 hours (*SD* = 3.16) per week training, and had trained with their current team for an average of 3.03 years (*SD* = 3.05). 175 (89.7%) participants reported never having used performance enhancing drugs, 13 (6.7%) had used them prior to the past three months, four (2.1%) had used them in the past three months and three (1.5%) were currently using them.

Combined sample hardcore-gym participants. Participants were male (*n* = 102) or female (*n* = 23) hardcore-gym exercisers; ages ranged from 17 to 70 years (*M* = 27.97, *SD* = 9.17). Participants had been training/competing for an average of 6.90 years (*SD* = 6.16), spent an average of 9.89 hours (*SD* = 4.52) per week training, and had trained in their current gym for an average of 4.19 years (*SD* = 4.53). 76 (60.8%) participants reported never having used performance enhancing drugs, 23 (18.4%) had used them prior to the past three months, 15 (12.0%) had used them in the past three months and 11 (8.8%) were currently using them.

Combined sample corporate-gym participants. Participants were male (*n* = 46) or female (*n* = 75) corporate-gym exercisers; ages ranged from 18 to 73 years (*M* = 33.20, *SD* = 14.61). Participants had been training/competing for an average of 10.92 years (*SD* = 9.23), spent an average of 4.78 hours (*SD* = 3.44) per week training, and had trained in their current gym for an average of 4.72 years (*SD* = 6.51). 117

(96.7%) participants reported never having used performance enhancing drugs, three (2.5%) had used them prior to the past three months, none had used them in the past three months and one (0.8%) were currently using them.

Results

Objective 1 - Questionnaire Development

Item development. First, a large pool of items for both measures (i.e., doping MD and doping SRE) was generated by (a) reviewing existing measures of MD and SRE, (b) adapting relevant data from qualitative studies of doping behavior (i.e., Boardley & Grix, 2014; Boardley et al., 2014, 2015), and (c) consulting with sport psychologists, sport coaches, exercise leaders, and sport and exercise participants. This resulted in an initial pool of 38 items for doping MD and 13 items for doping SRE. Item analyses were conducted on these items using data from an initial pilot study including sport and exercise participants (N = 280). These analyses demonstrated the bivariate correlations between items were largely in the target range (i.e., r = .15 - .50). However, any items that did not perform as expected were eliminated or adapted, and a small number of further items were generated based on feedback and data from the pilot analyses.

Following the initial item generation and pilot testing, the content validity of 43 doping MD items and 13 doping SRE items were examined through expert opinion. A total of 12 sport-psychology professionals with relevant experience in scale development rated how representative each item was of the definition for the relevant construct using a scale ranging from -3 (*not at all representative*) to +3 (*very representative*) with a mid-point of 0 (*uncertain*). The experts were also asked to provide qualitative comments on the degree to which items characterized the relevant mechanism of doping MD or doping SRE, as well as their relevance to the context of doping in sport and exercise. For each item, we took the mean of the experts' ratings once the lowest outlier had been removed to prevent extreme low responses skewing the mean. Items with mean scores of 2.0 or above (i.e., representative of their definitions) were retained, whereas those below 2.0 were either revised based on the qualitative comments of the experts or removed altogether; some items with means > 2.0 also underwent minor revisions based on experts' comments. Overall, of the 56 items assessed, 28 were retained without change, 18 underwent revisions to item content and 10 were removed altogether as a result of the content validity assessments. Three new doping MD items were also created, giving a total of 36 doping MD and 13 doping SRE items to take forward into pilot testing.

The two item pools were then pilot-tested with a sample of 122 sport and exercise participants to (a) ensure the items were correlated (r > .15) with other items measuring the same construct (i.e., relevant MD mechanism or doping SRE), (b) examine the internal consistency reliability of all subscales in the two measures, and (c) obtain qualitative feedback on item difficulty and wording (Clark & Watson, 1995). For the DMDS 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 was to be consistent with existing MD measures in sport (e.g., Boardley & Kavussanu, 2007, 2008) and because this approach provides the best compromise between reliability, validity, discriminating power, and respondent preference (Preston & Colman, 2000). Also consistent with existing measures (e.g., Bandura et al., 2001), for the DSRES a 5-point Likert scale anchored by 1 (*no confidence*) and 5 (*complete confidence*) was used during pilot testing and all subsequent data collections. Overall the two item sets performed well in pilot testing, with only minor adjustments to wording needed for a small number of items based on these analyses.

Construct validity. Data from three samples were used in the analyses aimed at developing and testing the construct validity of the measures. 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 investigation. Analyses examining the convergent and discriminant validity as well as internal reliabilities of the measures used the data from Sample 1 and Sample 2. Finally, analyses testing the test-retest reliability of the measures employed 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 testing the factorial validity of the measures, we used a two-stage process to identify the most effective items for measuring the respective construct; selected items were retained for use in subsequent testing. First, inter-item correlations were examined within each construct; any item not correlated ≥ .15 with all other items was removed (Clark & Watson, 1995). All items were interrelated ≥ .15, although 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 for this MD mechanism. Second, exploratory factor analysis (EFA) was conducted on each of the seven constructs (i.e., six MD mechanisms plus doping SRE) using principal axis extraction, with extraction based on an eigenvalue ≥ 1.00. Each of these EFAs produced a unidimensional factor structure, and with the exception of one item, all items had factor loadings ≥ 0.61. The item with the weaker factor loading (i.e., 0.32) was the displacement of responsibility item with weaker interrelations with other items mentioned previously. Given the issues with weaker interrelations and lower factor loading for this item, this item was replaced 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 commenced for Sample 2. Due to their consistently strong factor loadings, all other items were retained for use in subsequent analyses. This resulted in a total of 35 doping MD and 13 doping SRE items being retained for the main factorial validity analyses.

Factorial validity. The next step involved using confirmatory factor analysis (CFA) to identify the best items for use in the final measures. Confirmatory factor analysis was employed because (a) it offers a rigorous test of the plausibility of the factor structure, and (b) is the most appropriate method for confirming hypothesized factor structures (Fabrigar et al., 1999). The EQS 6.1 (Bentler & Wu, 2002) statistical package with the Maximum Likelihood method was used to conduct all CFAs; this package was also used to conduct the structural equation modelling (SEM) analyses reported later. As discussed previously, the DMDS encompassed six mechanisms of MD, and therefore was hypothesized to be multidimensional with potentially six first-order factors. Throughout the item-selection process a first-order six factor model (i.e., M1) was therefore specified to represent the maximum dimensionality of the new measure. In contrast, existing measures of SRE have consistently been unidimensional (e.g., Bandura et al., 2001; Lucidi et al., 2008), so we specified a single-factor model from the outset when examining the factor structure of our doping SRE measure.

In initial analyses the normalized estimate of Mardia's coefficient indicated substantial deviation from multivariate normality. Thus, the Robust Maximum Likelihood estimation method was used for all analyses, as 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. Indices used to estimate model fit for each model were the Satorra–Bentler scaled robust chi-square ($R\chi^2$), the robust comparative fit index (RCFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). Good model fit is attained when RCFI values are close to or above .95, the RMSEA is less than .06, and the SRMR is less than .08 (Hu & Bentler, 1999). To compare models, the robust consistent Akaike information criterion (RCAIC) was used. When making comparisons between nested models, the model with the lowest value is preferred (Hair, Anderson, Tatham, & Black, 1998).

In the first analysis testing the structure of the DMDS, six items were specified for each of the six MD mechanisms, except for displacement of responsibility which was represented by five items (M1a). The results showed an inadequate fit of the model to the data (Table 1, row 1) suggesting the need for respecification. Subsequently, 17 items that had 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 a final scale that would contain three items representing each of the six MD mechanisms. A final model (M1b) with 18 items produced a six-factor solution with excellent fit indices (Table 1, row 2).

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

Although the 18-item scale provides the flexibility of measuring each of the six MD mechanisms individually, researchers are sometimes primarily interested in measuring overall MD. For such instances, short versions of MD scales have been developed to reduce the time needed for scale completion during data collection (e.g., Boardley & Kavussanu, 2008) and in accord which such processes we presently also developed the DMDS-S. Our aim was to develop a six-item scale with one item representing each of the six mechanisms of MD that are represented in the DMDS. Two main steps were involved in selecting the best 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. The next step involved examining the factor structure of the remaining items. For these analyses, we performed EFA and CFA using the data from sample 1. The EFA was performed using principal axis extraction, with just a single factor extracted. The purpose of this analysis was to identify the best indicator of an overall doping MD factor for each of the six mechanisms of MD based on the strength of their factor loadings. Once these six items had been identified, they were then specified in a single-factor CFA (i.e., M5a). As can be seen in Table 1 (i.e., 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 parameter specified resulted in an excellent model fit. Although it is important to specify correlated errors when they are present to prevent the possibility of inaccurate parameter estimates (see Kline, 2015), such associations between items can be sample specific. As such, we accepted this model under the proviso that its presence would be confirmed in Sample 2.

When developing psychometric measures, it is important to confirm any identified factor structures using a separate sample to ensure the results obtained from the first sample are not sample specific (Fabrigar et al., 1999). As such, we then used the data from Sample 2 to confirm the factor structures identified in sample 1 for both the DMDS and DMDS-S. As shown in Table 1 (Rows 10 & 11), testing of the final models from the Sample 1 analyses again resulted in excellent model fit with Sample 2. Also, the significant correlated error between the advantageous comparison and distortion of consequences items identified in Sample 1 (i.e., $r = .30$) was again present in Sample 2 (i.e., $r = .25$). As such, models M1b and M5b were accepted as the final models for the DMDS and DMDS-S, respectively.

Similar procedures were then followed for the development of the DSRE. In the first analysis testing the structure of the DSRES with the data from Sample 1, all 13 items were specified to load on a single factor (M6a). The results showed an inadequate fit of the model to the data (Table 1, row 8) suggesting the need for respecification. Subsequently, 7 items that had large modification indices as indicated by the 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 a final scale 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 (Table 1, row 9). Subsequent testing of this model using the data from Sample 2 also resulted in excellent model fit (Table 1, row 12) and therefore confirmed the unidimensional structure of the DSRES. Items, factor loadings and error variances for the DMDS-S and DSRES can be found in Table 4. The final versions of the DMDS, DMDS-S and DSRES can be found in the appendices.

Multisample Analyses. When developing measures for use in diverse populations, it is important to determine the measurement invariance of such measures across the various sub-groups within the population. As such, in the current analyses we tested for measurement invariance between males and females and across different sport and exercise groups 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 scale was appropriate for making comparisons between groups, we tested three aspects of invariance relevant to these issues (Byrne, 2006): (a) configural invariance, which exists when the items of a scale are indicators of the same factors in different groups; (b) metric invariance, which is present when all factor loadings are equal across groups; and (c) equivalence of construct variance and covariance across the two genders, which determines whether the variances and covariances of the latent variables are equivalent across groups.

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	Rχ^2/*df*	RCFI	SRMR	RMSEA	RCAIC
			Sample 1				
1. M1a, 35 items	545	1183.39	2.17	.873	.050	.061	-2501.93
2. M1b, 18 items	120	222.54	1.85	.955	.034	.052	-588.91
3. M2, 18 items	132	667.24	2.88	.765	.079	.113	-225.35
4. M3, 18 items	135	779.76	5.78	.717	.088	.123	-133.12
5. M4, 18 items	129	303.92	2.36	.923	.043	.065	-568.38
6. M5a, 6 items	9	21.35	2.37	.968	.031	.066	-39.51
7. M5b, 6 items	8	7.38 (ns)	0.92	1.000	.020	.000	-46.72
8. M6a, 13 items	65	210.48	3.24	.871	.060	.084	-229.05
9. M6b, 6 items	9	12.17 (ns)	1.35	.991	.028	.033	-48.69
			Sample 2				
10. M1b, 18 items	120	248.14	2.07	.974	.032	.061	-553.07
11. M5b, 6 items	8	15.83	1.98	.987	.023	.058	-37.59
12. M6b, 6 items	9	12.44 (ns)	1.38	.991	.020	.036	-47.65

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; RCAIC = robust consistent Akaike information criterion; ns = $p > .05$. M1 = 8-factor DMDS model; 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* = 292)

Factor	1	2	3	4	5	6
1. Moral Justification	.86/.93	.55	.77	.87	.86	.86
2. Euphemistic Labelling	.51	.86/.90	.54	.57	.51	.48
3. Advantageous Comparison	.80	.54	.82/.86	.75	.81	.93
4. Displacement of Responsibility	.88	.48	.72	.91/.95	.90	.79
5. Diffusion of Responsibility	.84	.46	.79	.91	.88/.90	.85
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)

Item	Factor Loading	Error Variance	Factor
1. It is okay to dope if it helps an athlete to provide for his/her family.	.76/.82	.65/.57	MJ
2. Doping is okay if it helps an athlete advise others on how to do it right.	.87/.95	.49/.32	MJ
3. It is acceptable to dope if knowledge gained helps an athlete advise others on safe doping.	.87/.95	.49/.30	MJ
4. Saying you "take steroids" feels worse than saying you "use some gear".	.67/.75	.75/.67	EL
5. Using words like "roids", "gear" and "pinning" makes doping feel more acceptable.	.93/.93	.37/.37	EL
6. Using terms such as "gear" or "juice" makes doping sound less harmful.	.91/.94	.41/.35	EL
7. Compared to most lifestyles in the general public, doping isn't that bad.	.82/.86	.57/.50	AC
8. Compared to smoking, doping is pretty safe.	.79/.86	.62/.51	AC
9. Compared to physical violence, doping isn't that serious.	.75/.64	.66/.77	AC
10. Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	.87/.92	.50/.39	DisR
11. An athlete shouldn't be blamed for doping if a member of his/her training group has encouraged it.	.92/.94	.39/.34	DisR
12. An athlete shouldn't be held responsible for doping if his/her coach encouraged him/her to do it.	.88/.92	.47/.39	DisR
13. If most athletes in a sport dope, no one athlete should be held responsible for doing it.	.77/.81	.64/.59	DifR
14. It's not right to condemn individuals who dope when many in their sport are doing the same.	.87/.92	.50/.39	DifR
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	DifR
16. Risks associated with doping are exaggerated.	.85/.83	.53/.55	DC
17. Doping doesn't really harm anyone else.	.70/.78	.72/.63	DC
18. The negative aspects of doping are exaggerated by the media.	.82/.67	.58/.75	DC

Note. MJ = Moral Justification; EL = Euphemistic Labelling; AC = Advantageous Comparison; DisR = Displacement of Responsibility; DifR = Diffusion of Responsibility; DC = Distortion of Consequences. 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	Factor Loading	Error Variance	Mechanism
DMDS-S			
1. Doping is okay if it helps an athlete advise others on how to do it right.	.61/.86	.79/.52	MJ
2. Using terms such as "gear" or "juice" makes doping sound less harmful.	.44/.51	.90/.86	EL
3. Compared to most lifestyles in the general public, doping isn't that bad.	.70/.76	.72/.65	AC
4. Athletes shouldn't be blamed for doping if training partners/teammates pressure them to do it.	.76/.87	.65/.49	DisR
5. It's not right to condemn individuals who dope when many in their sport are doing the same.	.83/.88	.56/.48	DifR
6. Risks associated with doping are exaggerated.	.67/.73	.74/.68	DC
DSRES			
1. …resist doping even if your training group encouraged you to do it?	.84/.83	.55/.56	DSRE
2. …resist doping even if you knew you could get away with it?	.82/.88	.57/.48	DSRE
3. …ignore the temptation to dope even if you knew it would improve your performance?	.86/.86	.51/.50	DSRE
4. …resist peer pressure to dope?	.80/.87	.60/.49	DSRE
5. …reject doping even if most of your training partners did it?	.84/.87	.54/.49	DSRE
6. …ignore the temptation to dope when feeling down physically	.81/.77	.59/.63	DSRE

Note. MJ = Moral Justification; EL = Euphemistic Labelling; AC = Advantageous Comparison; DisR = Displacement of Responsibility; DifR = Diffusion of Responsibility; DC = Distortion of Consequences; DSRE = Doping Self-Regulatory Efficacy. Factor Loadings and Error Variances are presented as follows: Sample 1 / Sample 2.

We tested these aspects of invariance across sex and sport/exercise group using multisample CFA. Prior to invariance testing, we estimated baseline model fit separately for each group (see Byrne, 2006). Then, we tested for configural invariance by examining the fit of a model in which only the pattern of free and fixed parameters was constrained to be the same across groups. Next, we tested for metric invariance by comparing the fit of the metric invariance model to the fit of the configural invariance model (Byrne, 2006). Finally, we tested for ECVC by comparing the fit of the ECVC invariance model to the fit of the configural invariance model. To compare fit between more- and less-constrained models we examined the invariance of imposed constraints using the Lagrange multiplier (LM) test results. Imposed constraints were considered variant if they led to an increase in Rχ^2 of ≥ 5.0/df (Byrne, 2006). The results from these analyses are presented in Table 5.

Gender invariance. For the DMDS, model fit for the baseline models was very good for males and acceptable-to-good for females, and configural invariance was demonstrated by the good fit of the relevant model. Further, metric invariance was established, as shown by good fit for the metric invariance model as well as the absence of any constraints that resulted in an increase in χ^2 of ≥ 5.0/df. The equivalence of construct variance and covariance was not established though, as imposing such constraints had a significant impact on model fit. Inspection of the LM test results indicated several variances and covariances diverged between males and females. As such, no attempt was made to release specific constraints to achieve acceptable model fit.

For the DMDS-S, model fit for the baseline models was very good for males and excellent for females, and configural invariance was demonstrated by the very good fit of the relevant model. However, complete metric invariance was not established, as the LM test results for this model indicated constraining the equivalence of the euphemistic labelling item led to an increase in χ^2 of 5.18. Respecification of this model with this constraint released led to a model with no further variant constraints, indicated partial metric invariance (see Byrne, Shavelson, & Mûthen, 1989). Then, similar to the DMDS, the equivalence of construct variance and covariance was again not established. Inspection of the LM test results indicated both the variance of the doping MD factor and the covariance between the errors of the advantageous comparison and distortion of consequences items diverged between males and females.

Finally, for the DSRES, model fit for the baseline models was excellent for males and females, and configural invariance was demonstrated by the excellent model fit. Complete metric invariance was again not established though, as the LM test results for this model indicated constraining the equivalence of one of the six items led to an increase in χ^2 of 5.09. Respecification of this model with this constraint released led to a model with no further variant constraints, indicating partial metric invariance. Then, like the DMDS and DMDS-S, the equivalence of construct variance and covariance was not established as constraining the variance of the doping SRE factor to equivalence led to an increase in χ^2 of 15.99.

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*	$R\chi^2$	$R\chi^2/df$	RCFI	SRMR	RMSEA	RCAIC
Gender DMDS							
Baseline Males	120	243.58	2.03	.968	.040	.053	-586.36
Baseline Females	120	211.91	1.77	.924	.039	.057	-565.27
Configural Invariance	240	456.20	1.90	.951	.040	.054	-1323.03
Metric Invariance	252	466.58	1.85	.951	.044	.053	-1401.63
ECVC	273	560.39	2.05	.934	.219	.059	-1463.49
Gender DMDS-S							
Baseline Males	8	21.13	2.64	.981	.028	.067	-34.20
Baseline Females	8	12.63*	1.58	.980	.023	.049	-39.18
Configural Invariance	16	33.82	2.11	.978	.026	.061	-84.79
Metric Invariance	21	40.15	1.91	.976	.049	.055	-115.53
Metric Invariance Revised	20	36.21	1.81	.980	.033	.052	-112.06
ECVC	22	63.78	2.90	.949	.240	.079	-99.32
Gender DSRES							
Baseline Males	9	13.45*	1.49	.992	.021	.037	-48.79
Baseline Females	9	13.12*	1.46	.983	.035	.044	-45.16
Configural Invariance	18	26.58*	1.48	.989	.029	.040	-106.87
Metric Invariance	23	33.13*	1.44	.987	.052	.038	-137.38
Metric Invariance Revised	22	30.28*	1.38	.989	.039	.035	-132.82
ECVC	23	42.38	1.84	.975	.337	.053	-128.13
Sport/Exercise Group DMDS							
Baseline Corporate	120	166.83	1.39	.924	.047	.057	-528.66
Baseline Hardcore	120	209.50	1.75	.967	.053	.078	-489.89
Baseline Team	120	171.43	1.43	.950	.038	.047	-581.33
Baseline Individual	120	157.69	1.31	.966	.060	.043	-577.90
Configural Invariance	480	738.14	1.54	.953	.053	.063	-2761.82
Metric Invariance	516	786.06	1.52	.951	.071	.062	-2976.39
Metric Invariance Revised	513	766.86	1.49	.953	.066	.061	-2973.72
ECVC	576	953.97	1.66	.931	.366	.070	-3245.98
Sport/Exercise Group DMDS-S							
Baseline Corporate	8	11.01*	1.38	.978	.028	.056	-35.36
Baseline Hardcore	8	8.12*	1.02	1.000	.024	.011	-38.51
Baseline Team	8	9.98*	1.25	.989	.035	.036	-40.21
Baseline Individual	8	9.25*	1.16	.992	.035	.030	-39.79
Configural Invariance	32	38.68*	1.21	.992	.031	.037	-198.56
Metric Invariance	47	79.18	1.68	.960	.096	.067	-269.25
Metric Invariance Revised	45	57.37*	1.27	.984	.069	.043	-276.23
ECVC	51	133.45	26.62	.897	.314	.103	-244.63
Sport/Exercise Group DSRES							
Baseline Corporate	9	12.39*	1.38	.970	.050	.056	-39.78
Baseline Hardcore	9	14.17*	1.57	.986	.023	.068	-38.28
Baseline Team	9	11.77*	1.31	.989	.042	.040	-44.69
Baseline Individual	9	9.16*	1.02	.998	.028	.010	-46.01
Configural Invariance	36	46.84*	1.30	.983	.037	.045	-220.05
Metric Invariance	51	65.80*	1.29	.977	.087	.044	-312.29
ECVC	54	76.23	1.41	.965	.242	.052	-324.10

Note. $R\chi^2$ = Satorra–Bentler scaled chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; RCAIC = robust consistent Akaike information criterion; * = $p > .05$.

Sport/exercise group invariance. For the DMDS, model fit for the baseline models ranged from acceptable-to-good for corporate-gym users to excellent for hardcore gym users and individual-sport athletes. Configural invariance was established through the good fit of the relevant model. However, complete metric invariance was not established, as the LM test results for this model indicated constraining the equivalence of three items led to an increase in χ^2 of $\geq$ 5.0/df for each. Respecification of this model with these three constraints released led to a model with no variant constraints, indicated partial metric invariance. The equivalence of construct variance and covariance was again not established, as imposing such constraints had a significant impact on model fit. Inspection of the LM test results indicated several variances and covariances diverged between males and females. As such, no attempt was made to release specific constraints to achieve acceptable model fit.

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 again not established though, as the LM test results for this model indicated two of the specified constraints each led to an increase in χ^2 of $\geq$ 5.0/df. Respecification of this model with these two constraints released led to an excellent model fit with no variant constraints, indicated partial metric invariance. Then, similar to the DMDS, the equivalence of construct variance and covariance was again not established. Inspection of the LM test results indicated both the variance of the doping MD factor and the covariance between the errors of the advantageous comparison and distortion of consequences items diverged between some groups.

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. Complete metric invariance was also established, with this model achieving excellent fit with no variant constraints evidenced. However, the equivalence of construct variance and covariance was again not established as constraining the variance of the doping SRE factor to equivalence amongst the four groups led to poor model fit.

Convergent and discriminant validity. To provide further evidence for the construct validity of the DMDS, DMDS-S and DSRES we sought to establish evidence for their convergent and discriminant validities. We examined the convergent validity of the three new instruments by computing associations between scores obtained from them with participants' levels of empathy and guilt. According to Bandura's (1991) theory, doping MD should correlate negatively with both of these variables, whereas doping SRE should relate positively. Convergent validity of the DMDS, DMDS-S and DSRES would be established if such relationships were found in this study. In both samples, DMDS (Sample 1 r = -.28, p <.01; Sample 2 r = -.31, p <.01) and DMDS-S (Sample 1 r = -.26, p <.01; Sample 2 r = -.31, p <.01) scores were negatively related to empathy, whereas DSRES scores were positively associated with empathy (Sample 1 r = .15, p <.01; Sample 2 r = .35, 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 guilt, whereas DSRES scores were positively associated with guilt (Sample 1 r = .38, p <.01; Sample 2 r = .48, p <.01) in both samples.

In addition, the convergent validity of the individual DMDS subscales was examined by evaluating the correlations between the subscales and empathy and guilt in both samples (see Table 6). Examination of these correlations demonstrates evidence for the convergent validity of all subscales. 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). For example, whereas the correlation between euphemistic labelling and empathy were generally weak (Sample 1 r = -.18, p < .01; Sample 2 r = -.09, p > .05), those between distortion of consequences and empathy were consistently moderate in magnitude (Sample 1 r = -.32, p < .01; Sample 2 r = -.31, p < .01). Similarly, while the relationships between euphemistic labelling and guilt were largely moderate (Sample 1 r = -.30, p < .01; Sample 2 r = -.23, p < .01), those between moral justification and guilt were dependably strong (Sample 1 r = -.67, p < .01; Sample 2 r = -.58, p < .01).

Then, to examine discriminant validity, we computed the correlation between sport MD (Boardley & Kavussanu, 2008) and doping MD (using both the DMDS and DMDS-S) and between peer pressure SRE and doping SRE using the DSRES. Evidence for the discriminant validity of the three new scales would be established if moderately strong positive correlations were obtained. However, if the correlations were too strong (i.e., r > .90) this would suggest to much overlap in variance between the measures (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 = .57, 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 our results provide strong evidence for the convergent and discriminant validity of the DMDS, DMDS-S and DSRES, as well as the DMDS subscales.

Internal consistency. To determine the internal consistencies of the three new measures we calculated Cronbach's alpha values for all scale scores and subscale scores. As can be seen in Table 1, internal consistency was either good or very good for all subscales of the DMDS in both Sample 1 and Sample 2. Further, alphas for overall doping MD (i.e., specifying all scale items) for the DMDS were excellent in both samples (i.e., Sample 1 = .95; Sample 2 = .96). Similarly, internal consistency was very good for the DMDS-S in both samples (i.e., Sample 1 = .86; Sample 2 = .89). Finally, alpha coefficients for the DSRES demonstrated excellent internal consistency in both samples (i.e., Sample 1 = .93; Sample 2 = .94).

Table 6

Correlations Between the Doping Moral Disengagement Scale (DMDS) Subscales and Empathy and Guilt in Sample 1 (*N* = 318) and Sample 2 (*N* = 292)

	Sample 1		Sample 2	
DMDS Subscale	Empathy	Guilt	Empathy	Guilt
Moral Justification	-.24	-.67	-.34	-.58
Euphemistic Labeling	-.18	-.30	-.09*	-.23
Advantageous Comparison	-.25	-.56	-.24	-.55
Displacement of Responsibility	-.21	-.60	-.32	-.54
Diffusion of Responsibility	-.20	-.61	-.29	-.57
Distortion of Consequences	-.32	-.65	-.31	-.61

Note. Correlation significant at $p < .01$ unless indicated by *, where $p > .05$.

Test-retest reliability. Test–retest reliabilities were assessed for overall scores obtained using the DMDS, DMDS-S and DSRES, as well for those for the individual subscales of the DMDS. We estimated test-retest reliability using two methods with the data from the two times points. The first method used was to calculate Pearson correlations between the scores obtained in the two assessments, before then using the two scores to compute intra-class correlation coefficients (ICC). As can be seen in Table 6, test-retest reliability coefficients ranged from good to excellent, with most scales and subscales demonstrating very good to excellent levels of score reproducibility between assessments.

Objective 2 - Structural Model Testing

Testing the hypothesized model. To test the hypothesized model, SEM was employed using the two-step approach recommended by Anderson and Gerbing (1988). The first step involves testing the measurement model which includes the postulated relationships of the observed variables to their respective latent constructs, with all latent constructs allowed to intercorrelate. The measurement model specified included the six items of the DMDS-S, the six items of the DSRES, the best six indicators of empathy (three items for empathic concern and three for perspective taking; determined through factor loadings and modification indices during CFA [see Hofmann, 1995]), five items from the SSGS and one item for reported doping. Specification of this model resulted in an excellent model fit for the data from Sample 1, χ^2 (241) = 339.53, p = >.05; CFI = .964; RMSEA = .036; SRMR = .051, and Sample 2, χ^2 (241) = 287.68, p = >.05; CFI = .984; RMSEA = .026; SRMR = .043.

Table 7

Test-Retest Reliabilities for the Doping Moral Disengagement Scale (DMDS), the Doping Moral Disengagement Scale – Short (DMDS-S), the Doping Self-Regulatory Efficacy Scale (DSRES) and DMDS Subscales

	Scale/Subscale								
Test Method	DMDS	DMDS-S	DSRES	MJ	EL	AC	DisR	DifR	DC
Pearson Correlation	.91	.87	.79	.83	.88	.88	.83	.76	.82
Interclass Correlation	.94	.93	.87	.91	.93	.92	.91	.87	.89

Note. MJ = Moral Justification; EL = Euphemistic Labelling; AC = Advantageous Comparison; DisR = Displacement of Responsibility; DifR = Diffusion of Responsibility; DC = Distortion of Consequences. For all correlations $p < .01$.

We then proceeded to the second step in Anderson and Gerbing's approach, which involves testing a model incorporating the hypothesized structural pathways. Specification of the structural model resulted in a very good model fit for the data from Sample 1, χ^2 (244) = 349.61, p = >.05; CFI = .962; RMSEA = .037; SRMR = .053, and Sample 2, χ^2 (244) = 297.79, p = >.05; CFI = .981; RMSEA = .028; SRMR = .049. As shown by the standardized coefficients (see Figure 1), empathy and doping SRE, respectively, had weak-to-moderate and moderate-to-strong negative predictive effects on doping MD, empathy and doping MD, respectively, had weak-to-moderate positive and strong-to-very strong negative predictive effects on anticipated guilt, and doping MD and anticipated guilt, respectively, had weak-to-moderate positive and negative predictive effects on doping use. In Sample 1 the model accounted for 22% of the variance in doping MD, 58% of the variance in anticipated guilt and 20% of the variance in doping use. In Sample 2 the model accounted for 39% of the variance in doping MD, 50% of the variance in anticipated guilt and 16% of the variance in doping use.

Mediational Analyses. To investigate the extent to which predictive effects operated via the mediational paths shown in Figure 1, we requested the decomposition of model effects into direct, indirect, and total effects (Bollen, 1987). In Sample 1, for the effect of empathy on anticipated guilt via doping MD, the total, direct, and indirect effects were .31 ($p < .05$), .17 ($p < .05$), and .14 ($p < .05$), respectively; the percentage of the total effect mediated by doping MD was 45%. Next, for the effect of doping SRE on anticipated guilt via doping MD, the total, direct, and indirect effects were .27 ($p < .05$), .00 ($p > .05$), and .27 ($p < .05$), respectively; the percentage of the total effect mediated by doping MD was 100%. Then, for the effect of doping MD on doping use via anticipated guilt, the total, direct, and indirect effects were .41 ($p < .05$), .24 ($p < .05$), and .17 ($p < .05$), respectively; the percentage of the total effect mediated by anticipated guilt was 41%. Next, for the effect of doping SRE on doping use via doping MD and anticipated guilt, the total, direct, and indirect effects were -.16 ($p < .05$), .00 ($p > .05$), and -.16 ($p < .05$), respectively;

the percentage of the total effect mediated by anticipated guilt was 100%. Finally, for the effect of empathy on doping use via doping MD and anticipated guilt, the total, direct, and indirect effects were -.12 ($p < .05$), .00 ($p > .05$), and -.12 ($p < .05$), respectively; the percentage of the total effect mediated by anticipated guilt was 100%.

In Sample 2, for the effect of empathy on anticipated guilt via doping MD, the total, direct, and indirect effects were .44 ($p < .05$), .33 ($p < .05$), and .11 ($p < .05$), respectively; the percentage of the total effect mediated by doping MD was 25%. Next, for the effect of doping SRE on anticipated guilt via doping MD, the total, direct, and indirect effects were .24 ($p < .05$), .00 ($p > .05$), and .24 ($p < .05$), respectively; the percentage of the total effect mediated by doping MD was 100%. Then, for the effect of doping MD on doping use via anticipated guilt, the total, direct, and indirect effects were .31 ($p < .05$), .18 ($p < .05$), and .13 ($p < .05$), respectively; the percentage of the total effect mediated by anticipated guilt was 42%. Next, for the effect of doping SRE on doping use via doping MD and anticipated guilt, the total, direct, and indirect effects were -.15 ($p < .05$), .00 ($p > .05$), and -.15 ($p < .05$), respectively; the percentage of the total effect mediated by anticipated guilt was 100%. Finally, for the effect of empathy on doping use via doping MD and anticipated guilt, the total, direct, and indirect effects were -.16 ($p < .05$), .00 ($p > .05$), and -.16 ($p < .05$), respectively; the percentage of the total effect mediated by anticipated guilt was 100%.

Multigroup Analyses. When testing structural models in diverse populations, it is also important to determine the equivalence of the final model across different subgroups within the overall population. As such, in the current analysis we tested for measurement and structural invariance of the final model between males and females and across different sport and exercise sub-groups. We followed the same procedures as described for testing the measurement invariance of the SMDS, DMDS-S and DSRE, except here we also tested for *structural invariance* by determining whether model fit was affected when all structural components in the model were constrained to be equal. Due to the extremely low levels of variance in reported doping in some of the sub-samples (see method section), for the purpose of these analyses we used doping susceptibility in place of reported doping. The results of these analyses are shown in Table 8.

Gender invariance. Model fit for the baseline models was very good for both males and females, and configural invariance was demonstrated by the very good fit of the relevant model. However, complete metric invariance was not established, as three of the imposed constraints led to an increase in χ^2 of ≥ 5.0. Respecification of the metric invariance model with these constraints released resulted in very good model fit and no further variant constraints, indicating partial metric invariance. Then, in the equivalence of construct variance and covariance model three of the imposed constraints were variant. Releasing these constraints in a revised model resulted in very good model fit. Finally, constraining the structural components of the model to equivalence between males and females illustrated the path from empathy to doping MD was variant between the two genders. Specification of a model with this constraint released resulted in a model with very good model fit and no further variant constraints. This model also indicated the nature of the divergent path coefficients, with the standardized coefficient for the path from empathy to doping MD being stronger in males (i.e., -.25, $p < .05$) than females (i.e., .03, $p > .05$).

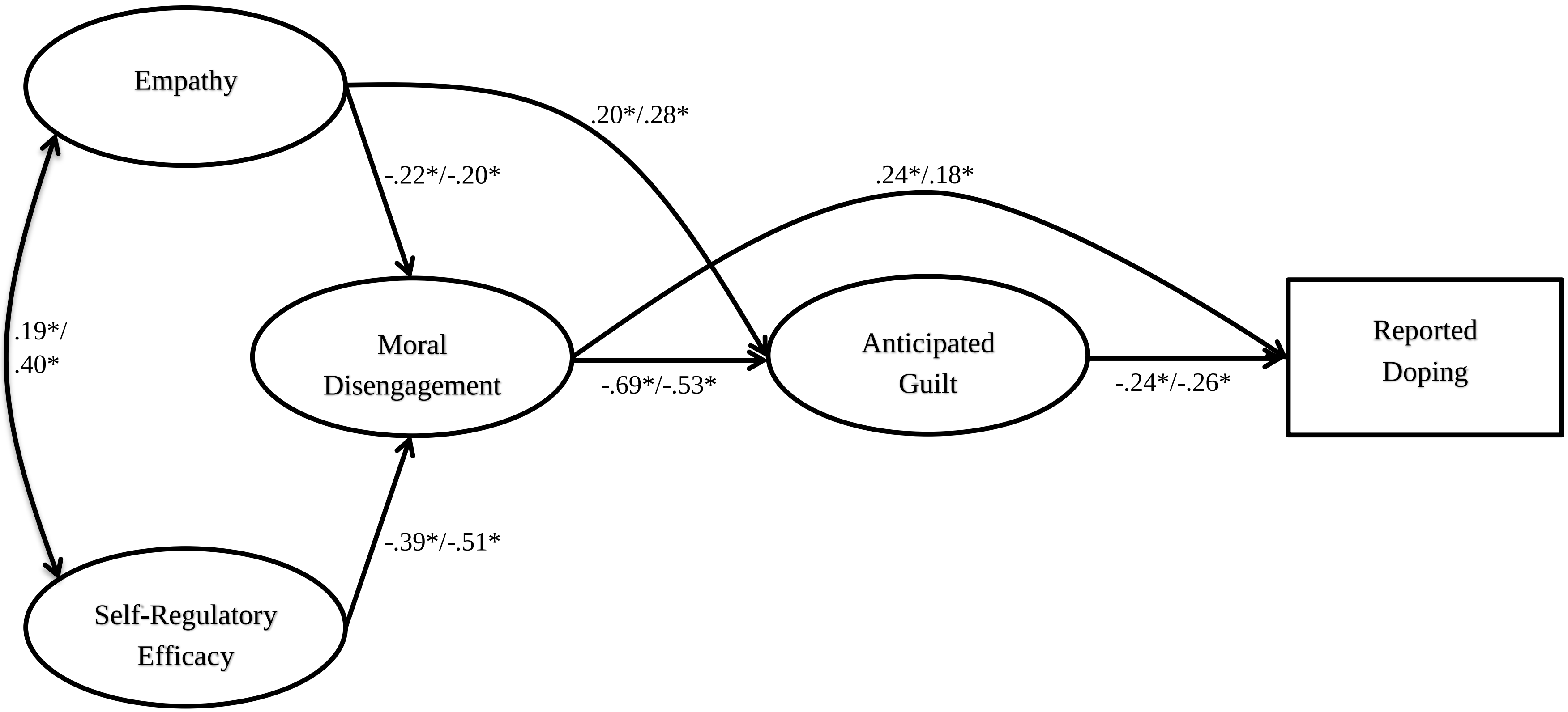
Sport/exercise group invariance. Model fit for the baseline models ranged from acceptable-to-good for corporate-gym exercisers to very good for hardcore-gym attendees. Then, configural invariance was established by the good model fit for this model. Complete metric invariance was not established though, and a number of constraints had to be released to achieve a satisfactory partial metric invariance model. This was also the case for the equivalence of construct variance and covariance model, such that several constraints had to be released to achieve a model without any variant constraints present. Finally, constraining the structural components of the model to equivalence across the four groups illustrated the path from empathy to doping MD was variant between corporate-gym exercisers and hardcore-gym exercisers, team-sport athletes and individual-sport athletes. In addition, the path from doping SRE to doping MD was also variant between hardcore-gym exercisers and corporate-gym exercisers. Specification of a model with these constraints released resulted in a model with acceptable model fit and no further variant constraints indicated. This model also indicated the nature of the divergent path coefficients, with the standardized coefficient for the path from empathy to doping MD being stronger in hardcore-gym exercisers (i.e., -.45, $p < .05$) than corporate-gym exercisers (i.e., -.18, $p < .05$), team-sport athletes (i.e., -.14, $p > .05$) and individual-sport athletes (i.e., .06, $p > .05$). Further, the path from doping SRE to doping MD was stronger in corporate-gym exercisers (i.e., -55, $p < .05$) than hardcore-gym exercisers (i.e., -.39, $p < .05$), individual-sport athletes (i.e., -.43, $p < .05$) and team-sport athletes (i.e., -.30, $p < .05$).

Discussion

Both qualitative and quantitative research has highlighted the potential importance of doping MD and doping SRE to the regulation of doping in sport and exercise contexts (e.g., Boardley & Grix, 2014; Boardley et al., 2014, 2015; Hodge, Hargreaves, Gerrard, & Lonsdale, 2013; Lucidi et al., 2004, 2008; Zelli et al., 2010). However, to date instruments developed to measure these constructs have questions regarding their validity, and have not always been developed through a rigorous scale-development process. Further, researchers have not to date investigated the role of doping SRE and doping MD alongside other key variables (e.g., empathy, guilt) when investigating their predictive effects on doping, nor have they studied them across a range of relevant contexts. The present research sought to address these issues by first developing valid and reliable measures of doping SRE and doping MD, before then using these instruments to test a hypothetical model of doping behavior grounded in Bandura's (1991) theory. Over the coming paragraphs we review and discuss the key findings from the research pertaining to these aims.

Figure 1

Structural Model Including Parameter Estimates for Sample 1 (N = 318) and Sample 2 (N = 292)



Note. Parameter estimates are shown Sample 1 / Sample 2. For all parameter estimates, $p < .05$.

Table 8

Fit Indices for Multisample Analyses on the Structural Process Model

Model	*df*	$R\chi^2$	$R\chi^2/df$	RCFI	SRMR	RMSEA	RCAIC
			Gender				
Baseline Males	244	368.52	1.51	.968	.045	.037	-1319.04
Baseline Females	244	296.23	1.21	.973	.060	.030	-1284.02
Configural Invariance	488	665.44	1.36	.969	.053	.035	-2764.88
Metric Invariance	507	694.22	1.37	.967	.059	.035	-2865.58
Metric Invariance Revised	504	678.74	1.35	.970	.055	.034	-3057.64
ECVC	508	717.58	1.41	.964	.134	.037	-3048.46
ECVC Revised	505	682.20	1.35	.969	.059	.034	-3061.60
Structural Invariance	511	700.32	1.37	.967	.081	.035	-3087.96
Structural Invariance Revised	510	689.07	1.35	.969	.064	.034	-3091.80
			Sport/Exercise Group				
Baseline Corporate	244	317.00	1.30	.917	.070	.050	-1097.17
Baseline Hardcore	244	316.49	1.30	.969	.052	.049	-1105.62
Baseline Team	244	332.79	1.36	.941	.070	.043	-1197.82
Baseline Individual	244	345.29	1.42	.913	.074	.050	-1150.40
Configural Invariance	976	1312.40	1.34	.939	.067	.048	-5923.14
Metric Invariance	1033	1407.41	1.36	.932	.087	.049	-6250.70
Metric Invariance Revised	1006	1338.71	1.33	.939	.073	.047	-6119.24
ECVC	1018	1392.63	1.37	.932	.132	.049	-6154.27
ECVC Revised	1012	1358.12	1.34	.937	.107	.048	-6144.30
Structural Invariance	1030	1408.18	1.37	.931	.145	.049	-6227.68
Structural Invariance Revised	1024	1369.05	1.34	.937	.109	.047	-6222.33

Note. $R\chi^2$ = Satorra–Bentler scaled chi-square; RCFI = robust comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; RCAIC = robust consistent Akaike information criterion; ECVC = equivalence of construct variance and covariance. * = $p > .05$.

Objective 1 - Questionnaire Development

Although Bandura (1991) described eight mechanisms of MD, research in sport and exercise contexts has shown only six of these 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 the final scale to incorporate a six-factor structure. Consistent with this expectation, results from both samples suggested doping MD – as assessed by the DMDS – incorporates six 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 a six-factor structure. However, where the measures diverge is in the number of MD mechanisms represented within them. Whereas the DMDS incorporates six mechanisms of MD, the MDSS incorporates eight. This difference is based on evidence that shows the relevance of all eight MD mechanisms for rationalizing and justifying aggressive and antisocial acts in sport – which the MDSS is designed to assess – contrasting with the six mechanisms that appear to be most relevant to doping (Boardley & Kavussanu, 2007; Boardley & Grix, 2014; Boardley et al., 2014, 2015). These differences between the scales emphasize the context-specific nature of MD, and the need for specific measures of MD developed for specific behaviors and circumstances (see Bandura, 1991).

The construct validity of the DMDS was demonstrated by providing evidence for the scale's convergent and discriminant validity. Convergent validity for the overall scale was evidenced by the very strong negative correlation between doping MD and anticipated guilt, and the moderate negative correlation between doping MD and empathy; these findings were replicated in both samples. These associations are consistent with Bandura's theory (1991) and research (Bandura et al., 1996), and the association between DMDS scores and these theoretically related constructs confirms its usefulness in future research. Further, evidence for the convergent validity of most DMDS subscales was provided. This was indicated through the relationships of the subscales with anticipated guilt and empathy, which for most subscales were generally consistent with those for the overall scale. However, there was some variation in terms of the associations of certain subscales with guilt and empathy. Overall, distortion of consequences tended to have some of the strongest associations with empathy and guilt over the two samples. Given doping MD often centers on downplaying the potential health consequences of doping (see Boardley & Grix, 2014; Boardley et al., 2014, 2015) and that this mechanism primarily focusses on downplaying the consequences of harmful conduct, distortion of consequences may be a particularly important MD mechanism in this context. Opposing this, the associations between euphemistic labelling and guilt and empathy – whilst largely meaningful and in the anticipated direction – were clearly weaker than those for the other five DMDS subscales. These weaker relationships could be explained by evidence that suggests use of euphemistic terms regarding doping (e.g., gear, juice, etc.) is a key aspect of doping culture (see Andrews, Sudwell, & Sparkes, 2005) and as such people may at times use such terminology to fit in with this culture and not just for the purposes of make doping sound and feel less harmful.

Next, evidence for the discriminant validity of the DMDS was evidenced through strong positive associations between DMDS and MDSS scores in both samples. The strength of these associations provided ideal support for the DMDS as a measure of MD that is related to – but distinct from – existing measures of MD developed for use in sport. This suggests it is not just the context but also the behavior that is important when developing measures to assess MD. Evidence for discriminant validity was also provided internally by the strength of the associations amongst the DMDS subscales. Consistent with the associations amongst the MD mechanisms and empathy and guilt discussed above, euphemistic labelling had weaker relationships with the other five mechanisms than the other mechanisms did amongst themselves. This suggests the 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. Clearly, these subscales share a considerable amount of variance with each other which indicates low levels of discriminant validity. However, the strength of these relationships is not out of line with past research in sport, which has shown similar levels of convergence between these mechanisms (Boardley & Kavussanu, 2007). Despite the high inter-correlations, in both samples the fit of the six-factor model was superior to ones in which these factors were merged, which suggests whilst they may have much in common they are still empirically distinct from one another.

As well as developing the DMDS, we also developed a short version of the scale termed the DMDS-S. The development of this measure was based on the premise the availability of a valid and reliable yet concise measure of doping MD would increase the options available to researchers interested in measuring this construct. The development and validation of the DMDS-S followed procedures used successfully to develop previous short versions of MD scales (e.g., Boardley & Kavussanu, 2008). To develop the DMDS-S, we first identified a set of six items from the DMDS-S representing each of the mechanisms of MD represented in the long version of the scale before then establishing evidence for the construct validity of the short version. Following this, CFA was used to confirm the factor structure of the DMDS-S. As expected, across both samples CFA confirmed the unidimensional structure of the DMDS-S. We then analyzed the correlations between scores on the short version with theoretically-related variables. These analyses provided support for the discriminant and convergent validity of the short scale in both samples, with DMDS-S scores being strongly and positively correlated with sport MD, moderately and negatively related to empathy and very strongly and positively related to anticipated guilt. In contrast to the DMDS – which is a multidimensional scale – the DMDS-S is a unidimensional measure of doping MD. Whereas the DMDS was designed to measure the six mechanisms of MD relevant to doping, the DMDS-S was developed solely to measure overall doping MD. Importantly, the scale still has items representing all six of the relevant mechanisms of MD (see Boardley & Grix, 2014; Boardley et al., 2014, 2015). This aspect of the scale was developed intentionally to ensure the DMDS-S generates doping MD scores that are equally representative of each of the six mechanisms. Overall, the evidence presented shows the DMDS-S to be a valid measure of doping MD.

The availability of two measures of doping MD increases the options available to researchers, with each scale having benefits suited to certain research questions. For research questions centered solely on overall doping MD, either the DMDS or DMDS-S would be suitable measures. However, in research that involves administering multiple measures the DMDS-S would be the better choice given the scale's brevity. An example of such research would be that of Hodge et al. (2013) in which a multiple-variable process model of doping susceptibility was tested. As an aside, one of the limitations of this research was the use of the MDSS to assess MD rather than a doping specific-measure. Future work employing either the DMDS or the DMDS-S to test similar process models should explain greater percentages of variance in doping-related outcome variables given the context-specific nature of MD (see Bandura, 1991). However, research questions are sometimes focused on specific mechanisms of MD, such as in the work of Stanger, Kavussanu, Boardley and Ring (2013), in which the specific effects of the attribution of blame mechanism were of primary interest. For similar research with doping-related variables the DMDS would be the ideal choice given the scale's ability to assess each of the six doping MD mechanisms individually. The differing predictive abilities of the individual MD mechanisms in the present work demonstrate the potential importance of such future research. In sum, researchers should choose which of the two measures of doping MD would best suit their research depending on the specific research questions they are looking to answer.

Another important context-specific variable from Bandura's (1991) theory for which a doping-specific measure was needed is doping SRE. Although past research had established the potential importance of this variable to doping (e.g., Lucidi et al., 2008), prior to the current research no fully-validated doping-specific SRE measure had been developed. As such, to address this issue, we also developed and validated a measure of doping SRE termed the DSRES. Doping SRE represents a person's capacity to withstand personal and social influences that encourage doping. As such, we first developed a large pool of items representing all the major personal and social factors that may influence the likelihood of someone doping, before then using the data to identify the best six items to represent these factors. Based on theory (Bandura, 1991) and the factorial structure of existing measures of SRE (e.g., Bandura et al., 1996), we anticipated the DSRES would be unidimensional. Factor analysis on the data from both samples confirmed this hypothesis by demonstrating the unidimensional structure of the DSRES. Once the final item set was selected, we assessed the relationships between DSRES scores and theoretically-related variables. First, moderate-to-strong and strong positive associations with peer-pressure SRE in samples 1 and 2, respectively, provided support for the discriminant validity of the DSRES. Then, weak (sample 1) and moderate (sample 2) positive associations with empathy and moderate-to-strong (sample 1) and strong positive correlations with guilt supported the convergent validity of the scale. Overall, data analyses on the data from samples 1 and 2 strongly supported the validity of the DSRES as a measure of doping SRE.

In addition to establishing evidence for various facets of the new measures' validity, we also assessed their internal consistency and test-retest reliability. First, scores for overall doping MD generated using the DMDS had consistently excellent levels of internal consistency. Further, the individual subscales of the DMDS were found to have good to very good levels of internal consistency across both samples. Next, the internal reliability of the DMDS-S was very good in both datasets. Finally, the DSRES was shown to have consistently excellent levels of internal consistency. Importantly, the internal consistency of all scales and subscales surpassed the minimum criterion level recommended for those developing new psychometric scales (i.e., 0.80; Clark & Watson, 1995). Beyond internal consistency, we also examined the test-retest reliability of the three scales across a nine- to 16-day period using a separate sample. Pearson correlations and ICC values indicated good to excellent levels of reliability across this time frame. These analyses demonstrated that over the short term, scores obtained using the DMDS, DMDS-S and DSRES are relatively stable and can be replicated with a high degree of measurement precision (Widaman, Little, Preacher, & Sawalani, 2011). Overall, in conjunction with other analyses, all three measures were shown to be valid and reliable measures of their target constructs.

Beyond the main validity and reliability analyses, we also utilized multisample analyses to examine the measurement invariance of all three instruments. In each case, we performed two sets of multisample analyses. The first set examined the measurement invariance of the scales across gender to examine whether the items of the scale had the same meaning in male and female participants. The second set of analyses then examined the measurement invariance of the scales across the four sport/exercise groups represented in the sample. First, for all three scales, configural invariance was established. This means that across all groups (i.e., males/females; hardcore gym/corporate gym/team sport/individual sport) the same subsets of items are associated with the same constructs (Cheung & Rensvold, 2002). Next, we examined metric invariance, which can be assessed at both the item- and construct-level. At the item-level, metric invariance was established across the two genders for the DMDS and four sport/exercise group for the DSRES. This demonstrates that strength of the relationship between all items and their underlying constructs was the same in these instances. For the remaining scales and group combinations, partial metric invariance was established (Byrne et al., 1989). This is present when the majority of items for a given latent variable have loadings that are invariant across groups. As long as this is the case, cross-group comparisons can safely be made (Reise, Widaman, & Pugh, 1993). At the construct-level, metric invariance is established when constraining factor loadings to be equivalent across groups does not significantly reduce fit of the overall model. When construct-level metric invariance is demonstrated this means overall the strength of the relationships between items and the constructs they measure are the same across groups (Cheung & Rensvold, 2002). Importantly, construct-level metric invariance was established for all three scales in both the gender and sport/exercise type multisample analyses. Thus, the DMDS, DMDS-S and DSRES are suitable for research testing substantive hypotheses regarding group differences between males and females and among the four sport/exercise types tested.

One further level of measurement invariance was tested, that of ECVC. Equivalence of construct variance exists when the range of responses given to each item is the same across groups (Cheung & Rensvold, 2002). However, this level of equivalence was not established for any of the scales either between genders or across the four sport/exercise groups. This suggests that the range of item responses varies among these groups. Next, equivalence of construct covariance exists when the relationships among constructs (i.e., their covariances) are the same across groups. Given the DMDS was the only multidimensional scale being developed, such constraints were only imposed and tested for this measure. Again, similar to variance constraints, equivalence of covariance was not supported in either the gender or sport-exercise group analyses. This demonstrates that the relationships among dimensions of doping MD differ among these groups. Future researchers should look to determine what may lead to these differences in variance and covariance across groups, although it is quite possible that the greater acceptance of doping in some groups (e.g., males and hardcore-gym exercisers) compared to others (e.g., females and corporate-gym exercisers) is leading to these group differences. These findings highlight the need for researchers using any of these new scales in samples that consistent of different genders and/or participants from different sport/exercise groups to test for any such group differences as part of their analytical strategy.

Objective 2 - Structural Model Testing

The other major aim of the current project was to test a model of doping behavior with team- and individual-sport participants and corporate- and hardcore-gym exercisers. Grounded in Bandura’s (1991) theory, the hypothesized process model (see Figure 1) depicted empathy and SRE would negatively predict doping MD, doping MD would negatively predict anticipated guilt, and anticipated guilt would negatively predict reported doping (Bandura, 1991; Bandura et al., 1996; Bandura et al., 2001; Lucidi et al., 2008; Paciello et al., 2013; Stanger et al, 2012; Zelli et al., 2010). In addition, doping SRE and empathy would negatively predict doping indirectly via changes in doping MD and anticipated guilt. Finally, doping MD was projected to positively predict doping directly and indirectly through anticipated guilt (Bandura, 1991; Bandura et al., 1996). Overall, data analyses strongly supported the efficacy of this model in both samples, and the meaning and implications of the relevant findings are subsequently discussed.

One of the major contributions of this aspect of the current project was the strong support provided for the main tenets of Bandura's (1991) theory. Although researchers have tested some of the more holistic aspects of this theory in other contexts (e.g., Bandura et al., 1996, 2001), doping researchers have instead investigated the predictive effects of MD (e.g., Hodge et al., 2013) and SRE (e.g., Lucidi et al., 2004, 2008) on doping-related outcomes within process models based primarily on other theories. By investigating the predictive effects of these variables in process models that included key aspects of Bandura's (1991) theoretical framework that had not previously been tested in doping, we were able to determine whether the mechanisms these variables may operate through are consistent with Bandura's (1991) theory. The consistent support we found for our hypothesized model provides evidence of the relevance of all major aspects of Bandura's (1991) theory for doping in sport and exercise. Another major strength of the current work was that we purposefully tested our model in a sample that represented a range of relevant populations, including ones in which doping is highly prevalent. This too contrasts with past work investigating doping MD and doping SRE, which has either sampled from populations in which the prevalence of doping was extremely low (e.g., Lucidi et al., 2004, 2008) or failed to assess the prevalence of doping (Hodge et al., 2013). Thus, the current work demonstrates the relevance of Bandura's (1991) theory in athletic populations in which there is a demonstrable need to understand the psychosocial factors that facilitate doping.

Regarding the specific predictive effects shown in model testing, consistent support was found for empathy being a possible antecedent of both doping MD and anticipated guilt. More specifically, weak-to-moderate predictive effects of empathy on these variables were demonstrated in both samples. The predictive effect of empathy on doping MD was negative, supporting Bandura's (1986, 1991) contention that higher levels of empathy lead to lower levels of MD and less frequent transgressive behavior. Endorsement of and engagement in deleterious conduct is more difficult when one can anticipate and experience the consequences of one's actions for others. This is the first study to show this effect in the specific context of doping, or in sport or exercise more generally. However, it is consistent with empirical work investigating unethical business decisions and youth antisocial behavior (Detert, Treviño, & Sweitzer, 2008; Hyde, Shaw, & Moilanen, 2010). Using a sample of business students, Detert et al. (2008) showed empathy to be a negative predictor of MD. Similarly, Hyde et al. (2010) employed a prospective design to show empathy at age 12 negatively predicted MD at age 15 in youth from low-income families. In contrast to those on doping MD, the predictive effects of empathy on anticipated guilt were positive. This is a further novel finding, as researchers to date have not investigated empathy as a precursor of anticipated guilt in doping research. These variables have been empirically linked in children though, with more empathic children showing higher levels of guilt (Roberts, Strayer, & Denham, 2014). Thus, the associations between empathy, MD and guilt in the current study are consistent with theory and empirical work in other contexts, and suggest empathy may be important to our understanding of doping behavior.

The possible role of doping SRE as an antecedent of doping MD was also supported during model testing. In Sample 1 and 2, respectively, moderate-to-strong and strong negative predictive effects of doping SRE on doping MD were shown. Thus, participants who had stronger beliefs in their capacity to withstand personal and social influences encouraging doping reported lower levels of doping MD. This effect is consistent with Bandura et al.'s (2001) assertion that increased SRE leads to lower levels of MD, as those who are confident in their ability to resist incentives to transgress have no need to justify and rationalize detrimental conduct through MD. It is also consistent with Bandura et al.'s (2001) empirical work, which demonstrated a weak negative effect of peer pressure SRE on MD in the context of delinquent behavior in children. Our finding extends this predictive effect to the context of doping across a range of key sport and exercise populations.

In terms of the predictive abilities of doping MD, this variable was found to predict both anticipated guilt and reported doping. Regarding anticipated guilt, doping MD had very strong and strong negative predictive effects, respectively, in Samples 1 and 2. This shows that participants who had a greater tendency to agree with justifications and rationalizations for doping were less likely to anticipate experiencing guilt if they took the decision to dope. This finding provides strong support for this key aspect of Bandura's (1991) theory, and provides the first empirical evidence of its relevance to doping. It is however consistent with research in other contexts, which has shown a negative predictive effect of MD on guilt in research investigating children's' interpersonal aggression and delinquent conduct (Bandura et al., 1996). In addition to its effect on anticipated guilt, doping MD also had a weak-to-moderate positive effect on reported doping in both samples, such that participants with higher levels of doping MD were more likely to report having taken performance enhancing drugs. This finding provides statistical evidence to support qualitative research that links MD with PED use in male bodybuilders and team- and individual-sport athletes (Boardley & Grix, 2014; Boardley et al., 2014, 2015). It is also consistent with Bandura et al.'s (1996) research on children's interpersonal aggression and delinquent conduct in that they too found MD not only predicted guilt, but also had a direct positive predictive effect on such behavior. Similarly, in both the current research and Bandura et al. (1996), guilt was also found to have a negative effect on the target behavior. These findings support another aspect of Bandura's (1991) theory, which suggests increased levels of anticipated guilt should deter transgressive and harmful conduct. As such, collectively these findings provide further support for the potential relevance of Bandura's (1991) theory to our understanding of the psychosocial factors that govern doping behavior.

In addition to the direct effects already discussed, there were also a number of indirect (i.e., mediated) associations identified during model testing. First, empathy had a weak positive predictive effect on anticipated guilt via doping MD in both samples, such that when participants had higher levels of empathy, associated increases in anticipated guilt were explained through lower levels of doping MD. This mediated effect has been identified in past research in other contexts. Detert et al. (2008) showed reduced MD mediated a negative effect of empathy on unethical decision making in business students. Also, Hyde et al. (2010) employed a prospective design to show MD at age 15 meditated an effect of empathy at age 12 on youth antisocial behavior at age 16-17. Integrating this finding with Bandura's (1991) theory, this suggests when athletes and exercisers are more able to understand and experience the impact of their actions on others, they are less able to rationalize and justify doping, which in turn means they anticipate experiencing greater guilt if they did. Similarly, doping SRE had a moderate positive predictive effect on anticipated guilt via doping MD in both samples, meaning participants higher levels of anticipated guilt associated with increased doping SRE could be explained through lower levels of doping MD. Indirect effects such as this have again been reported previously, such as Bandura et al.'s (1996) work showing peer pressure SRE reduced transgressive behavior through changes in MD. Interpreting this finding in the current work suggests athletes and exercisers who are more able to resist internal and external pressures to dope should anticipate feeling more guilt for doping because of their reduced tendency to not rationalize and justify doping.

The third indirect effect was a weak positive effect of doping MD on reported doping via anticipated guilt seen in both samples. As such, the increased levels of reported doping in participants with higher levels of doping MD could be explained through lower levels of anticipated guilt. Consistent with this effect, Bandura et al. (2001) found guilt mediated an effect of MD on children's delinquent behavior. Presently, such a pathway suggests athletes and exercisers who have higher levels of doping MD are more likely to dope because justifying and rationalizing doping allows them to do so without anticipating deterrent emotions such as guilt. Finally, both empathy and doping SRE had weak negative predictive effects on reported doping via their relationships with doping MD and anticipated guilt. As such, both of these variables may influence doping though the combined effects of some of the previously discussed indirect effects. In sum, it is important to consider both the direct and indirect effects operating in Figure 1, and to keep in mind all variables in the hypothesized model had predictive effects on reported doping either directly or via their links with other variables.

Although the multisample analyses largely supported the invariance of the structural model across gender and sport/exercise type, there were a small number of divergent paths. First, in both the gender and sport/exercise type analyses, there were differences in the strength of the path from empathy to doping MD. More specifically, this path was stronger in males than females and in hardcore-gym exercisers than corporate-gym exercisers, and team- and individual-sport athletes. Interestingly, past research has shown related group differences in levels of empathy. For instance, males have been shown to be lower than females (e.g., McGinley & Carlo, 2007; Carlo et al., 1999), and steroid users to be lower than non-users (Porcerelli & Sandler, 1995). It therefore seems possible that in groups in which empathy tends to be lower, any increases in it may have a more powerful deteriorating effect on doping MD than in groups were levels of empathy are higher. Given empathy has been shown to be trainable in groups associated with low levels of empathy (Hepper, Hart, & Sedikides, 2014), future researchers are encouraged to specifically test the causal nature of this path and also identify particular groups in which empathy-based interventions may be most effective at reducing doping MD. The one other divergent path was the one from doping SRE to doping MD, which was stronger in corporate-gym exercisers than in the other three sport/exercise groups. Although it is not clear at this stage what may explain this difference, the differences between groups were not that stark in this case, and the effect was at least moderate in magnitude in all groups. As such, this divergent path doesn't have the possible practical implications that are evident for the one from empathy to doping MD.

Limitations and Future Research Directions

In achieving its first aim, the current project developed three valid and reliable psychometric instruments. However, it is important to consider that validation is a continuing process (Clark & Watson, 1995) and there are further aspects of validity that remain to be assessed in future work. For instance, the predictive validity of the measures over time need to be examined, as do their associations with other theoretically linked variables not assessed in the present work. As an example, their associations with more socially situated emotions such as shame could be examined. Further, their validity and reliability in other sport and exercise populations and alternative cultures also need to be investigated. Finally, the multidimensional nature of the DMDS in particular, and the divergent associations of its subscales with guilt opens up the opportunity for experimental research that determines whether certain MD mechanisms are more potent facilitators of doping than others.

In accomplishing its second aim, this research also revealed a number of interesting findings. However, as with any research these results should be considered alongside certain limitations resulting from the research design. First, given the model-testing aspects of the project were based on cross-sectional data, the causal nature of the predictive effects identified could not be tested and therefore should not be inferred. Future researchers are therefore encouraged to build upon our work by employing experimental or quasi-experimental designs to test the causal nature of the identified associations. Longitudinal research testing the temporal ordering proposed in the models tested would also be a worthwhile direction for future work. A second limitation was our use of self-report measures during model testing, meaning the precision of the reported associations are reliant to some degree on participants' honesty and ability to introspect and provide accurate responses to questionnaire items. Researchers could build on our findings by employing alternative approaches, such as physiological testing for evidence of doping and/or assessment of psychophysiological responses indicative of variables such as empathy.

Recommendations

To conclude this work, we present some practical conclusions stemming from the project. In line with some of the key areas for research activity identified by WADA's Education Committee at the time of its conception, this project has designed and validated three key measures with the potential to improve social science research on doping. As such, our first recommendation would be to consider replacing the existing measure of MD with the DMDS and DMDS-S in the research package provided by WADA to their stakeholders for the purpose of evaluating the effectiveness of their education programs. Bandura (1991) is clear on the context-specific nature of MD, and as such the inclusion of doping-specific measures of MD should improve the face validity of the assessment of MD, strengthen any effects of education programs on MD and improve the ecological validity of the findings stemming from such evaluations. Similarly, given the evidence for its validity and reliability – and links with reported doping in model testing – we also recommend adding the DSRES to the research pack. Also, consistent with key areas of activity identified by WADA's education committee, we robustly examined a hypothesized model of doping behavior across gender and four key sport and exercise contexts. In doing so we have contributed important knowledge on potential causes of doping behavior by identifying four psychosocial factors that potentially influence doping. As such, we recommend incorporating the main findings from model testing into the education materials of National Anti-Doping Organizations and within future iterations of WADA's Athlete Learning Program about Health and Anti-Doping (ALPHA). Finally, we recommend encouraging and supporting work that further examines the causal nature of the identified associations. Such work could further inform intervention strategies aimed at reducing doping in sport, and examine the robustness of some of the divergent associations identified in model testing such that intervention programs could be specifically tailored to particular athlete groups.

References

Aaronson, N., Alonso, J., Burnam, A., Lohr, K. N., Patrick, D. L., Perrin, E., & Stein, R. E. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research, 11*, 193–205.

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.

Ajzen, I. (1991). The theory of planned behaviour: Some unresolved issues. *Organizational Behaviour and Human Decision Processes, 5*, 179-211.

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

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

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*, 364-374.

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*, 125-135.

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.

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.

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.

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

Boardley, I. D., & Kavussanu, M. (2008). The moral disengagement in sport scale-short. *Journal of Sports Sciences*, *26*, 1507-1517.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 17*, 37–69.

Byrne, B. M. (2006). *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming* (2nd edn.). Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M., Shavelson, R. J., & Mûthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Carlo, G., Raffaelli, M., Laible, D. J., & Meyer, K. A. (1999). Why are girls less physically aggressive than boys? Personality and parenting mediators of physical aggression. *Gender Roles, 40*, 711–729.

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

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

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

Davis, M. H. (1994). *Empathy. A Social Psychological Approach*. Madison: Brown and Benchmark.

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.

Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: a study of antecedents and outcomes. *Journal of Applied Psychology, 93*(2), 374-391.

Enders, C.K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine, 68*, 427–436.

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*, 272-299.

Feshbach, N. D. (1975). Empathy in children: Some theoretical and empirical considerations. *The Counseling Psychologist, 5*, 25-30.

Gucciardi, D. F., Jalleh, G., & Donovan, R. J. (2010). Does social desirability influence the relationship between doping attitudes and doping susceptibility in athletes? *Psychology of Sport and Exercise, 11*(6), 479-486.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th Ed.). Upper Saddle River, NJ: Prentice Hall.

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*, 238-247.

Hepper, E. G., Hart, C. M., Sedikides, C. (2014). Moving narcissus: can narcissists be empathic? *Personality and Social Psychology Bulletin, 40*, 1079–1091.

Hodge, K., Hargreaves, E. A., Gerrard, D., & Lonsdale, C. (2013). Psychological mechanisms underlying doping attitudes in sport: Motivation and moral disengagement. *Journal of Sport & Exercise Psychology, 35*(4), 419-432.

Hofmann, R. (1995). Establishing factor validity using variable reduction in confirmatory factor analysis. *Educational and Psychological Measurement, 55*, 572-582.

Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. New York: Cambridge University Press.

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

Hyde, L. W., Shaw, D. S., & Moilanen, K. L. (2010). Developmental precursors of moral disengagement and the role of moral disengagement in the development of antisocial behavior. *Journal of Abnormal Child Psychology, 38*(2), 197-209.

Kavussanu, M., Hatzigeorgiadis, A., Elbe, A. M., & Ring, C. (2016). The moral disengagement in doping scale. *Psychology of Sport and Exercise, 24*, 188–198.

Kavussanu, M., Stamp, R., Slade, G., & Ring, C. (2009). Observed prosocial and antisocial behaviors in male and female soccer players. *Journal of Applied Sport Psychology*, *21*, S62–S76.

Kline, R. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: The Guilford Press.

Lucidi, F., Grano, C., Leone, L., Lombardo, C., & Pesce, C. (2004). Determinants of the intention to use doping substances: An empirical contribution in a sample of Italian adolescents. *International Journal of Sport Psychology, 35*, 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*, 447-456.

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

McGinley, M., & Carlo, G. (2007). Two sides of the same coin? The relations between prosocial and physically aggressive behavior. *Journal of Youth and Adolescence, 36*, 337–349.

Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior, 29*, 371-393.

Paciello, M., Fida, R., Cerniglia, L., Tramontano, C., & Cole, E. (2013). High cost helping scenario: The role of empathy, prosocial reasoning and moral disengagement on helping behavior. *Personality and Individual Differences, 55*, 3-7.

Pastorelli, C., Caprara, G. V., Barbaranelli, C., Rola, J., Rozsa, S., & Bandura, A. (2001). Structure of children's perceived self-efficacy: A cross-national study. *European Journal of Psychological Assessment, 17,* 87-97.

Pedhazur, E. J., & Schmelkin, L. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Porcerelli, J. H., & Sandler, B. A. (1995). Narcissism and empathy in steroid users. *American Journal of Psychiatry*, *152*(11), 1672-1674.

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,* 552–566.

Roberts, W., Strayer, J., & Denham, S. (2014). Empathy, anger, guilt: Emotions and prosocial behaviour. *Canadian Journal of Behavioural Science, 46*(4), 465-474.

Sjöqvist, F., Garle, M., & Rane, A. (2008). Use of doping agents, particularly anabolic steroids, in sports and society. *The Lancet*, *371*(9627), 1872-1882.

Stanger, N., Kavussanu, M., Boardley, I. D., & Ring, C. (2013). The influence of moral disengagement and negative emotion on antisocial sport behavior. *Sport, Exercise, and Performance Psychology, 2*(2), 117.

Stanger, N., Kavussanu, M., & Ring, C. (2012). Put yourself in their boots: Effects of empathy on emotion and aggression. *Journal of Sport & Exercise Psychology, 34*, 208-222.

Stuewig, J., Tangney, J. P., Heigel, C., Harty, L., & McCloskey, L. (2010). Shaming, blaming and maiming: Functional links among the moral emotions, externalization of blame, and aggression. *Journal of Research in Personality, 44*, 91-102.

Vaughn, B. K., & Daniel, S. R. (2012). Conceptualising validity. In G, Tenenbaum, R. C. Eklund, & A. Kamata (Eds.), *Measurement in Sport & 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. http://dx.doi.org/10.1037/12350-003

Zelli, A., Mallia, L., & Lucidi, F. (2010). The contribution of interpersonal appraisals to a social-cognitive analysis of adolescents' doping use. *Psychology of Sport and Exercise*, *11*, 304-311.

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