Comparing sport motivation scales: A response to Pelletier et al.

Chris Lonsdale [a,b,]*, Ken Hodge [c], Elaine A. Hargreaves [c], Johan Y.Y. Ng [d]

[a] Australian Catholic University, Australia
[b] University of Western Sydney, Australia
[c] University of Otago, New Zealand
[d] The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

ABSTRACT

Objectives: Pelletier, Rocchi, Vallerand, Deci, and Ryan (2013) proposed a new version of the Sport Motivation Scale (SMS; Pelletier, Fortier, Vallerand, Tuson, & Blais, 1995) as a measure of different types of behavioral regulations in sport, as outlined in self-determination theory (Ryan & Deci, 2000). They examined various aspects of reliability and validity of scale scores, and concluded that the new scale performs better than the original version. They also claimed that the SMS-II is superior to other measures of motivation in sport, including the Behavioral Regulation in Sport Questionnaire (BRSQ) developed by Lonsdale, Hodge, and Rose (2008). By comparing the evidence presented in papers by Pelletier et al. and Lonsdale et al., our objective was to examine the relative merits and shortcomings of the two measures and suggest directions for future research into sport motivation measurement.
Conclusions: Both the SMS-II and BRSQ have shown relative strengths and weaknesses. Overall, the construct validity evidence of scores derived from the two measures was similar. There is insufficient information to support the claim that one scale is superior to the other. Researchers are encouraged to make direct comparisons by administering both measures to the same group of participants in future studies.

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

Pelletier et al. (2013) examined the validity and reliability of scores derived from the Sport Motivation Scale II (SMS-II). The SMS-II, a revision to the Sport Motivation Scale (SMS; Pelletier, Fortier, Vallerand, Tuson, & Blais, 1995), is based on self-determination theory (SDT; Ryan & Deci, 2000), and was designed to measure different types of behavioral regulations in sport. Pelletier et al. conducted a two-staged study and stated they had overcome the problems with the original SMS, concluding that the SMS-II "performs as well as or better than the original scale" (p. 338). Furthermore, Pelletier et al. concluded that compared with the "BRSQ [Behavioral Regulation in Sport Questionnaire] developed by Lonsdale, Hodge, and Rose (2008), we think that the SMS-II represents a scale that better addressed the limitations observed with the original SMS and that showed more consistent results with SDT" (p. 339). In the current paper we examine the evidence related to this claim and suggest future directions for research that may improve sport motivation measurement in the future.[1]

In 2008, we published "The Behavioral Regulation in Sport Questionnaire: Instrument Development and Initial Validity Evidence" (Lonsdale et al., 2008). The measure described in that report was designed to assess sport motivation, as outlined in SDT (see Pelletier et al. (2013) or Lonsdale et al. (2008) for a description of relevant aspects of the theory, including definitions of motivational constructs). One issue with the BRSQ that we acknowledged was that although the majority of evidence showed that there was a distinction between subscales measuring intrinsic motivation, autonomous extrinsic motivation (integrated and identified regulation), controlled motivation (external and introjected regulation), and amotivation, some evidence suggested that scores of integrated and identified regulation were not distinctive. Similarly, external and introjected regulation scores were correlated and showed similar correlations with some constructs (e.g., other types of behavioral regulations). Recently, Pelletier et al. (2013) drew conclusions somewhat similar to our own regarding the nomological validity of BRSQ scores. They wrote,

"Overall, although [Lonsdale et al.'s (2008)] results showed support for the distinctions between the [BRSQ] self-determined subscales (intrinsic motivation and identified and integrated regulation) and the non-self-determined subscales (external and introjected regulation), the finer discrimination

within each type of category appears to be lacking ... there was a lack of discrimination between external and introjected regulation scores in terms of their relationships with amotivation; identified and integrated regulation subscales both had similar high correlations with intrinsic motivation; and there was a lack of discrimination between the self-determined subscales (intrinsic motivation and identified and integrated regulation) and the concepts of flow and burnout." (p. 331).

After acknowledging the limitations of the original SMS and voicing their concerns regarding nomological validity related to some of the BRSQ scores, Pelletier et al. (2013) decided to revise the SMS. They conducted two studies to evaluate the reliability and validity of scores derived from this new measure, entitled the SMS-II. They concluded that compared with the BRSQ, the SMS-II addressed the shortcomings of the SMS better. However, the design of their study allowed them to only directly compare the original SMS and the SMS-II, but not the BRSQ, as it was not included in their data collection. In the absence of direct evidence collected from the same sample of participants, it is our opinion that firm conclusions regarding the relative merits of the SMS-II and the BRSQ are premature.

In this paper we examine the evidence supporting Pelletier et al.'s (2013) claim that the SMS-II is superior to the BRSQ. Currently, comparisons between the SMS-II and the BRSQ and other sport motivation measures (e.g., SMS-6; Mallett, Kawabata, Newcombe, Otero-Forero, & Jackson, 2007) can only be achieved by comparing data collected from different samples. This procedure is clearly not ideal, but at this stage we feel it is necessary to debate the merits of Pelletier et al.'s conclusions. We aim to create an opportunity to enter a respectful, constructive discussion of ways to improve the measurement of SDT constructs in the sport setting.

Comparison of BRSQ and SMS-II reliability and validity evidence

In the absence of data pertaining to a direct comparison of scores derived using the BRSQ and SMS-II, we compared results presented in the Pelletier et al. (2013; Study 2) and Lonsdale et al. (2008; Study 3) papers. Specifically, we compared the results of reliability and validity testing performed on data gathered from two samples. These two samples were chosen because of their relatively similar mean age (approximately two years older in the New Zealand sample). We decided not to present detailed information from the adult athletes from Pelletier et al.'s Study 1 (mean age = 40.44 years) or the elite athletes from the New Zealand Academy of Sport in our Study 2 (mean age = 25.9 years). While the evidence gathered from these samples was largely similar to that gathered from the adolescent and young adult samples, we felt it was appropriate to limit the sample variation as much as possible. Also, limiting the samples allowed us to present a reasonable amount of data for the reader to digest. Indeed, presenting the results from all analyses in our paper and Pelletier et al.'s would not have added greatly to the discussion and may have made an already nuanced comparison practically inscrutable.

Internal consistency

Both studies examined internal consistency using Cronbach alpha coefficients (α). The SMS-II subscale scores had alpha coefficients that ranged from .73 to .86. The BRSQ subscale scores ranged from .76 to .91. Thus, both questionnaires produced scores that would typically be considered internally consistent when measuring these constructs (Nunnally, 1978).

Factorial validity

A comparison of fit indices is limited by the fact that SMS-II and BRSQ have different numbers of items per factor and thus non-equivalent models. Nonetheless, an overview of model fit from confirmatory factor analyses of scale scores from the two studies is presented below with reference to commonly accepted cut-off criteria used to judge model fit. Both models had a significant χ^2 at $p < .001$, suggesting a lack of model fit. However, as the χ^2 is sensitive to sample sizes, model fit based on other fit indices were considered. Hu and Bentler (1999) suggested cut-off criteria for the RMSEA ($\leq.06$), CFI ($\geq.95$) and TLI ($\geq.95$) that have been widely adopted in recent years. Previously, less stringent criteria (RMSEA $\leq .08$, CFI $\geq .90$, and TLI $\geq .90$) were often employed. BRSQ fit statistics were generally strong and met Hu and Bentler's criteria (RMSEA = .07, 90% CI [.06, .08], CFI = .97, TLI = .97). Approximate fit statistics from the SMS-II were also strong, but in some instances fell below Hu and Bentler's suggested cut-off scores (RMSEA = .07, 90% CI [.05, .08], CFI = .94 and TLI = .92). Pelletier et al. (2013) also reported NFI = .90, an index that we did not examine with the BRSQ data (Lonsdale et al., 2008). For the readers' interest, NFI for the analysis using BRSQ scores was .96. It is noteworthy that neither report (Lonsdale et al., 2008; Pelletier et al., 2013) included the SRMR index that Hu and Bentler (1999) showed is particularly important when assessing model fit. For the reader's interest, the SRMR result from our 2008 study was .08, which meets the criterion for good fit ($\leq.08$). Overall, the results generally supported the factorial validity of scores derived from both the BRSQ and the SMS-II, with the BRSQ model fit surpassing Hu and Bentler's cut-off criteria.

Nomological validity

Both studies included examinations of nomological validity that focused on inspection of inter-factor correlations (simplex structure) and correlations between motivation scores and theoretically related constructs. We consider correlations among autonomous motivation scores, then inspect the relations between autonomous motivation scores and controlled motivation, and then scrutinize the correlations among controlled motivation scores. According to SDT, scores representing constructs that are closer together on the self-determination continuum (see Pelletier et al., 2013) should be more strongly and positively correlated (i.e., simplex structure). More distal constructs would be expected to show weaker positive or stronger negative correlations. Finally, we conclude our examination of nomological validity by examining relations between motivation scores and measures of theoretically related constructs. According to SDT, more autonomous motives should be more strongly and positively related with adaptive outcomes, compared with more controlled motives that are expected to be more strongly and positively correlated with maladaptive outcomes. In all instances where we compared correlations across samples, we employed a z-test: $(r_1' - r_2')/\sqrt{(1/(n_1 - 1)) + (1/(n_2 - 1))}$ (Clark-Carter, 2005), where r' represents a Fisher transformation of the Pearson correlation. When comparing correlations within a sample (e.g., comparing the correlation between intrinsic motivation and identified regulation with the correlation between intrinsic motivation and integrated regulation) we followed procedures outlined by Steiger (1980). Details of all analyses can be obtained from the first author.

Relations among autonomous motivation scores

Pelletier et al. (2013) criticized the simplex structure of the scores derived from the BRSQ subscales intended to represent autonomous forms of extrinsic motivation. Specifically, they