

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Ruiz, Montse C.; Watt, Anthony P.; Samper, Natalia Palacios; Morris, Tony

Title: Cross validation of hard-copy and web-based formats of the Sport Imagery Ability Measure

Year: 2019

Version: Accepted version (Final draft)

Copyright: © 2019 Taylor & Francis

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Ruiz, M. C., Watt, A. P., Samper, N. P., & Morris, T. (2019). Cross validation of hard-copy and web-based formats of the Sport Imagery Ability Measure. *International Journal of Sport and Exercise Psychology*, 17(6), 647-658. <https://doi.org/10.1080/1612197X.2018.1428209>

Cross validation of hard-copy and web-based formats of the Sport Imagery Ability Measure

Montse C. Ruiz^{1*}, Anthony P. Watt², Natalia Palacios Samper^{1,3}, & Tony Morris⁴

1 Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

2 College of Arts and Education, Victoria University, Melbourne, Australia

3 Sports Studies, Research and Sport Medicine Centre (CEIMD), Government of Navarra, Pamplona, Spain

4 Institute of Sport, Exercise & Active Living, Victoria University, Melbourne, Australia

Abstract

The purpose of this multi-sample study was to examine the psychometric characteristics, factor structure, and measurement invariance of the hard-copy and web-based versions of a measure of sport imagery ability, termed Sport Imagery Ability Measure (SIAM). In the first sample Spanish athletes ($N = 274$, 161 men, 113 women, $M_{\text{age}} = 21.91$, $SD = 6.67$) completed a hard-copy version of the SIAM. A newly developed web-based version of the SIAM was cross validated in an independent group ($N = 266$, 147 men, 119 women, $M_{\text{age}} = 25.93$, $SD = 9.84$). A small group of participants ($n = 16$) completed both versions. Exploratory structural equation modelling and confirmatory factor analysis of the data from the hard-copy version showed that a 3-factor (i.e., generation, feeling, and single senses) solution of the SIAM reached satisfactory fit indices. The 3-factor solution showed good fit to the data obtained through the web-based version of the SIAM. Multi group comparisons provided support for measurement invariance across gender and competitive level. Evidence of full invariance of factor loadings was obtained for both formats of administration indicating that the factor structure was held across groups, while partial invariance was seen for item intercepts indicating inequality in the intercepts of some indicators.

Keywords: sport imagery; dimensional validity; replicability

The utilization of imagery skills constitutes an important element of the mental preparation of athletes and remains a critical domain of applied sport psychology research (Morris, Spittle, & Perry, 2004). Previous research has outlined that imagery processing constitutes an important element of cognitive processing associated with sports performance enhancement that can be improved through training (see Williams & Cumming, 2012; Morris, Spittle, & Watt, 2005), and incorporated within many facets of the psychological preparation of athletes, such as skill rehearsal, pre-competition routine, problem solving, strategy development, and coping with pain and injury. The development of imagery training programs in sport are best served when implemented with the accurate assessment of the sensorial and generational processing characteristics associated with the construct and the identification of variables that may differentiate athletes on the basis of their imagery skills (e.g., gender, competitive level). It is also crucial that the formats of assessment are sufficiently adaptable to support ease of use for both administrators and respondents and reflect the integration of technology within the development of measures employed in sport psychology. Current trends in psychological testing in sport highlight an increase in the use of online assessments because of these approach can support procedural improvements such as rapid feedback, accessibility to broader possibilities of participant recruitment, and creation of secure feedback reports to an athlete (Dillman, Smyth, & Christian, 2014).

Recent literature reinforces that the capacity for self-supervision of imagery processing is necessary for athletes to effectively engage in an imagery-training program (Budnik-Pyzybylska, Karasiewicz, Morris, & Watt, 2014). An essential early phase in this process is determining athletes' imagery ability. Morris et al. (2004) described imagery ability as "the capacity of the individual to create images, and is typically evaluated in terms of generational, sensorial and emotional qualities" (p. 21).

The process of athletes applying imagery as an element to support overall performance enhancement requires an adequate level of the dimensional attributes of proficiency to create, sustain, and control clear images, and the capacity to utilise a range of sensory modalities (Budnik-Pyzybylska, et al., 2014). The measurement of a range of dimensional and sensorial imagery abilities, therefore, is important to both research and practice in sport psychology.

Imagery ability is typically evaluated using scores derived from mental-ability tasks or responses to measures that consider behavioural or emotive imagery characteristics (Sheehan, Ashton, & White, 1983). Subjective approaches utilising self-report questionnaires constitute the typical procedure to determine the key elements of an individual's imagery abilities (Morris et al., 2005). The format has been consistently employed as the platform for measures designed for use within the sport and motor performance domains. A range of imagery questionnaires are currently employed for use in the motor skills domain, including the Movement Imagery Questionnaire-Revised (MIQ-R; Hall & Martin, 1997), the Movement Imagery Questionnaire-3 (MIQ-3, Williams, Cumming, Ntoumanis, Nordin-Bates, Ramsey, & Hall, 2012), and revised Vividness of Movement Imagery Questionnaire-2 (VMIQ-2; Roberts, Callow, Hardy, Markland, & Bringer, 2008). These questionnaires typically assess imagery ability associated with general motor movements, are limited in relation to dimension, perspective, and modality, and were not designed to examine images related to sport (Morris et al., 2005). As a consequence, concerns regarding the ecological validity of such measures for the assessment of the sport oriented imagery ability of athletes have been raised (Bhasavanija, Vongjaturapat, Morris, & Muangnapo, 2011).

A number of imagery ability measures have been developed specifically for use in sport. These include the Sport Imagery Questionnaire (SIQ; Martens, 1982), the

Motivational Imagery Ability Measure for Sport (MIAMS; Gregg & Hall, 2006), the Sport Imagery Ability Questionnaire (SIAQ; Williams & Cumming, 2011), and the Sport Imagery Ability Measure (SIAM; Watt, Morris, & Andersen, 2004). The SIAQ and SIAM, however, are the two measures that assess multiple characteristics related to sport imagery ability that have been more widely examined psychometrically.

The SIAQ was developed to assess sport-specific imagery abilities related to the ease in generating images associated with the five functions of athlete cognitive and motivational imagery use (i.e., skill, strategy, goal, affect, and mastery). The SIAQ has good factorial validity for the five-subscale model confirmed by fit indices of CFI = .96, TLI = .95, and RMSEA .06. Internal reliability was established based on composite reliability ranging from .76 to .86, and average variance extracted values ranging from .51 to .68. Temporal reliability assessed over a 3-month period resulted in the following coefficients for skill (.83), strategy (.86), goal (.86), affect (.75), and mastery (.85). Good factorial validity was demonstrated for a four-factor model with strong fit indices reported, including RMSEA of .05 and CFI of .96 (Williams & Cumming, 2011). Subsequent results provide further support of the reliability of the measure with alpha values ranging from $\alpha = .66$ for affect to $\alpha = .82$ for skill (Simonsmeier & Buecker, 2017). The current version of the SIAQ is limited to a single generational component rather than considering a broad range of generational and sensorial imagery characteristics (Budnik-Pyzybylska, et al., 2014).

Watt et al, (2004) developed the SIAM questionnaire to assess imagery ability directly related to sport. The measure expanded on Martens' (1982) SIQ by revising and extending the content of the imagery scenarios to encompass a broader variety of sports, and incorporate items related to all sense modalities and multiple dimensions. This approach represented an important strategy in addressing the limitations of previous

measures (e.g., SIQ, MIQ) that considered only one or two senses and a small set of dimensions. The final version of the SIAM consists of four generic sport-related scenarios from which the instrument can assess five imagery dimensions, namely vividness, control, ease of generation, speed of generation, and duration, six senses, namely visual, auditory, kinaesthetic, olfactory, gustatory, and tactile sense, and the experience of emotions (Morris et al., 2005).

The original validation study of the SIAM recruited Australian athletes and utilised exploratory factor analysis (EFA), confirmatory factor analysis (CFA), convergent, discriminant, and construct validity analyses. Further reliability and validity evaluation of the SIAM has produced consistent positive results from translations of the original English version. These studies have used versions of the SIAM developed in Finnish (Watt & Jaakkola, 2003), Thai (Bhasavanija et al., 2011), Swedish (Weibull, Johnson, & Watt, 2009), and Polish (Budnik-Przybylska et al., 2014). Psychometric data of the SIAM in those languages indicated strong internal consistency for the 12 subscales, with alphas rarely below .7 for specific subscales. Bhasavanija et al., (2011) reported adequate test-retest reliability for the Thai version of the measure. In the development of the measure a CFA demonstrated adequate fit for a three-factor structure (i.e., generation, feeling, senses) based on subscale scores with fit indices close to or above the accepted values of good fit (>0.8 for AGFI; >0.9 for NFI, CLI, TLI) for all but the RMSEA (< 0.1) (Watt et al., 2004). Similar fit indices were reported for subsequent CFA of a three-factor structure of the Polish version (Budnik-Pyzybylska et al., 2014).

Although hard-copy questionnaires are commonly adopted in survey research to collect data, the development of computer technology has facilitated the increased use and popularity of web-based measures as mediums for psychological assessment (Tu,

Tseng, Chang, & Lin 2017; Wang, Liu, Cheng, & Cheng, 2013). Advantages of web-based surveys have been reported as less costly, offering greater accessibility to larger and geographically diverse participant cohorts, reducing coding errors, increasing efficiency of scoring (Dillman et al., 2014), the capacity to embed audio and video, and incorporate response formats not possible in hard-copy measures such as drag and drop or sliding scales (Krille, Wuttke, & Seifried, 2017). Web-based approaches also support the use of programming languages that allow for sophisticated and engaging questionnaire design and offer direct transfer of information, which can expedite collection, analysis, and reporting of data (Wyrick & Bond, 2011). Reips (2006) stated that, with web-based administration, ethical and procedural limitations, including truly voluntary participation and the confounding of data due to extrinsically-motivated measure completion, are less likely to occur. Web-based questionnaires also have disadvantages such as limited participant control, inability to manage testing conditions such as noise or light, and management of personal characteristics such as typing skills or familiarity with technology (Krille, Wuttke, & Seifried, 2017).

Previous research, however, indicates that acceptable psychometric properties of hard-copy questionnaires do not necessarily translate into acceptable properties of web-based versions (Hirsch, Hauschild, Schmidt, Baum, & Christiansen, 2013). Thus, comparative analysis of web-based instruments of which there is a hard-copy version is recommended (Buchanan, 2003). In addition to hard-copy and web-based comparisons, in instrument development, multi-group comparisons are important to examine the extent to which an instrument operates in the same way or has the same theoretical structure across groups (i.e., gender, competitive levels). The analysis of the equivalence of the relationship between the observed variables and the relationships with the latent or unobserved variables (e.g., factor loadings) is labelled measurement

invariance. The examination of structural invariance, only tested when measurement invariance exists, focuses on the equivalence of unobserved variables (e.g., factor means) across groups, is critical to assure valid group comparisons made on a specific latent variable (Brown 2015; Farmer & Farmer, 2014). The equivalence of different administration modes or across gender, and different competitive levels in the assessment of imagery ability however, continues to remain unexplored.

The main purpose of the current study was to examine the psychometric properties of a Spanish language version of the SIAM, to further substantiate its validity as a multi-modal, multidimensional measure of imagery specifically operationalised within the sport context. This was specifically achieved through the examination of the dimensional structure, reliability, and equivalence of hard-copy and web-based Spanish versions of the SIAM. Measurement invariance was evaluated across the forms of administration (hard-copy vs. web-based), gender (male vs. female), and competitive level (regional, national, international). In addition, the stability of responses over time was examined in a separate group of participants.

Method

Participants

Two separate samples, in total 540 participants, were involved in this investigation. The first cohort, who completed the hard-copy SIAM, consisted of 274 athletes (161 men, 113 female), aged 15 to 57 years ($M = 21.91$; $SD = 6.67$), drawn from main sport organizations in the northern region of Spain. The athletes represented a range of individual sports ($n = 91$; e.g., track & field, taekwondo) and team sports ($n = 183$; e.g., handball, rugby). On average, participants had trained 11 h per week ($SD = 5.50$). Athletes competed at regional level (10.9%), national level (60.2%), and international

level (28.8%). The second independent sample who completed the web-based SIAM, included 266 athletes (147 men, 119 female), aged 15 to 61 years ($M = 25.93$; $SD = 9.84$). These athletes, drawn from sport organizations across Spain, had been involved in individual sports ($n = 164$) and team sports ($n = 102$) for an average of 12.60 years ($SD = 8.84$). They practised their sport an average of 9.98 h per week ($SD = 6.90$) and competed at regional level (28.6%), national level (44.7%), and international level (26.7%). A separate group of 7 athletes (6 male, 1 female) aged 22 to 29 years ($M = 24.29$; $SD = 2.56$) each representing a different sport modality, participated in a comprehensibility study.

Measures

The Sport Imagery Ability Measure (SIAM; Watt et al., 2004) is a 48-item self-report questionnaire that uses four sport-related scenarios. Athletes image each scenario and rate their imagery on five dimensions (i.e., vividness, control, duration, ease, and speed of generation), six senses (i.e., visual, auditory, olfactory, gustatory, tactile, and kinaesthetic) and the experience of emotion. Athletes are asked to listen to a description of each of the sport-related scenarios, after which they are allowed one minute to imagine the situation. At the end of the imagery interval, they are requested to rate each scenario by placing a cross (X) on a 100 mm visual analogue scale. No time limits are specified to answer the 12 items for each scenario. In previous research, the SIAM has demonstrated adequate internal reliability (i.e., Cronbach's ranging from $\alpha = .73$ for ease to $\alpha = .90$ for visual) and factorial validity (Budnik-Przybylska et al., 2014).

To develop the hard-copy of the Spanish version of the SIAM, back translation procedures and expert review were utilized. A bilingual person with knowledge about sport psychology conducted a direct translation into Spanish. A panel of four academics in sport psychology, whose first language was Spanish and who were competent in both

written and spoken English, examined the translated version. Each expert examined this version ranking each item as 1 (no change), 2 (minor change required), 3 (major change required), or 4 (reject and retranslate). Discrepancies between items were discussed and efforts made to ensure that the underlying meaning remained unchanged. The modified Spanish version was back-translated into English, and compared to the original to ensure that meaning and intent of the original items were maintained. Prior to conducting the main study, the translated version of the questionnaire was administered to a small group of athletes in a comprehensibility study. Minor changes were made to clarify Spanish language items.

The web-based version of the Spanish SIAM was constructed using the Qualtrics programme. A requirement of the hard-copy SIAM is the need for an administrator to manage the protocols associated with participant completion. The web-based version incorporates an electronic administration process to simulate the verbal description, timing, and sequencing requirements of the original measure.

Procedure

The protocols were submitted to, and approved by the institutional review board where the first author is based. Coaches were contacted and informed about the purpose of the study to gain access to their athletes. Prior to data collection the study purpose and procedures were explained to the participants, and written consent was obtained. Athletes under 18 years of age gave their assent and a guardian provided consent. American Psychological Association (APA) guidelines regarding anonymity and confidentiality were followed. The questionnaire was administered individually or in small groups, in a quiet place at a convenient time agreed with the participants. To allow an uninterrupted administration of the questionnaire, the participants were instructed on the procedures (e.g., limited time to imagine each situation), and were

given an example. In addition, to ensure the protocol of administration was understood, one situation was provided to give participants the opportunity to practice, and to ask questions. Once the procedure was clear, the administration was started. The first and third authors provided instructions about the standard procedures and administered the hard-copy of the questionnaire, which took approximately 30 minutes. The web-based questionnaire was distributed via sport organizations and Federations across Spain by sending a survey link. Completion of the web-based questionnaire took approximately 20 minutes.

Data Analysis

Data were screened for missing values, distribution, and potential outliers. Factor structure of the SIAM was examined separately for hard-copy and web-based versions using robust maximum likelihood estimator (MLR) to adjust for non-normality in Mplus 7.31 (Muthén & Muthén, 2012). We examined the factorial validity of SIAM using exploratory structural equation modelling (ESEM) and CFA, following suggestions from previous research (Marsh et al., 2009). In ESEM (Asparouhov & Muthén, 2009), which integrates exploratory and confirmatory factor analysis, factor loadings and cross-loadings are estimated for each item. Factor loading matrices were rotated using Geomin rotation facilitating the interpretation of the extracted factors. CFA, however, restricts factor loadings to align with specified latent factors, constraining all cross-loadings to zero. The competing models examined included one-factor (i.e., all items load to one factor), and three-factor (i.e., generation, feeling, and senses) solutions. Goodness of fit was examined using multiple indices that included Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and chi-square (χ^2). Good model fit was inferred based on the following

guidelines proposed by Hu and Bentler (1999): RMSEA less than or equal to .06, SRMR less than or equal to .08, and CFI and TLI close to .95. Akaike's Information Criterion (AIC) values were included for comparisons between competing models. Lower AIC values are representative of better fit. To test invariance across groups we conducted multi-group CFAs increasing parameter constraints and comparing equality of parameters in a sequential manner (Byrne, 2008). Specifically, an initial or configural model is tested. This model is the least restrictive and provides a baseline against which the following subsequently nested models were compared. In the configural model no equality constraints are imposed and the extent to which the same factor structure or pattern of fixed and freely estimated parameters holds across groups is tested. Metric invariance is determined next, by constraining factor loadings to be equal across groups, thus examining the extent to which the content of each item is interpreted in the same manner across groups. Strong metric invariance is then tested, constraining factor loading and item intercepts (predicted values of the indicator when the factor is zero) to be equal across groups. After having established measurement invariance, structural invariance was then examined to test the equivalence of the unobserved variables. Specifically, factor variance was determined by constraining the variance of the factor scores to be equal across groups. Finally, latent factor means were constrained to be equal across groups. The overall fit of the nested models was assessed using the selected goodness of fit indices. In addition, differences in overall χ^2 values and related degrees of freedom for the nested models were calculated using the Satorra Bentler chi-square difference ($\Delta\chi^2$), adjusting for scaling correction factor for MLR. Statistically non-significant $\Delta\chi^2$ values in comparison of the two nested models are indicative of invariance. A difference of CFI (ΔCFI) between the models was computed. A ΔCFI value of less than or equal to .01 is indicative of invariance. Full invariance denotes that

equivalence of parameters can be determined, whereas partial invariance indicates that some of the parameters may be specified differently for any of the groups.

Results

The descriptive analysis of the hard-copy administration identified no missing data. Participants reported higher scores for ease and speed (Table 1). The lowest scores were reported for the olfactory and gustatory subscales. Low Cronbach's alpha coefficients were found for ease ($\alpha = .56$) and speed ($\alpha = .61$) subscales. Composite reliability (CR) scores for the modified three-factor model were above .50 for SIAM subscales (generation, CR = .93; feeling, CR= .79; senses, CR= .54).

[Insert Table 1]

Strong positive correlations ($r > .78$) were found between the visual and vividness subscales, control and duration, ease and speed (Table 2) and no correlation was found between control and gustatory subscales.

[Insert Table 2]

Factor structure was examined through ESEM for two competing models with one and three factors, respectively. As Table 3 shows, better fit was achieved for a 3-factor ESEM model. Further, CFA analyses for one and three factors were conducted. A single factor CFA did not fit data well. A 3-factor CFA including a generation factor (i.e., vividness, control, ease, speed, duration and visual), a feeling factor (kinaesthetic, tactile, and emotion), and single sense factor (auditory, olfactory, gustatory) did not fit data well. An inspection of modification indices detailed the correlation of item residual errors. As the items loaded onto the same factor, these parameters were added to the model in a step-by-step fashion.

[Insert Table 3]

A model including these re-specifications fitted the data well. Standardized factor loadings were all statistically significant ($P < .001$; see Figure 1).

[Insert Figure 1]

For the web-based SIAM data, 45.54% out of 584 who opened the survey link completed the entire questionnaire. Empty responses were not allowed; thus, there was no missing data. Similar descriptive statistics (Table 1) and correlation (Table 2) patterns were reported for web-based SIAM subscales. Cronbach's alpha coefficients were above .62 for ease and speed, and above .71 for all other subscales. Composite reliability (CR) scores for the modified three-factor model were above .50 for web-based SIAM subscales (generation, CR = .92; feeling, CR= .87; senses, CR= .76). Best fit for web-based SIAM data was obtained for a 3-factor CFA (i.e., generation, feeling, single sense) model, allowing the correlation of the same residuals specified in the hard-copy version (Table 3).

Measurement invariance

Multi-group CFA results of invariance across hard-copy and web-based versions of SIAM are reported in Table 4. Good fit to the data was obtained for the configural multiple-group CFA confirming that SIAM had the same factor structure for both forms of administration. Model comparisons did not reveal differences between metric and configural models supporting the invariance of factor loadings. However, the difference between scale and metric models was statistically significant, thereby failing to support equality of item intercepts. Partial invariance was explored by releasing equality constraints of the intercepts of parameters with large difference between groups. The

constraints of the intercepts of the kinaesthetic item were released, after which, non-significant differences were obtained, providing support for partial invariance $\chi^2(119) = 180.816$, $p < .001$, CFI = .980, TLI = .976, RMSEA = .049, SRMR = .051. Model comparisons revealed that the difference between the metric model and the partial scalar invariance model was non-significant, $\Delta\chi^2(109) = 180.816$, $p < .001$, $\Delta df = 8$, $\Delta CFI = .001$. Uniqueness of partial invariance was tested excluding the kinaesthetic residual variance. Model comparisons revealed statistically significant differences. However, partial invariance was supported after releasing duration and auditory residual variances, $\Delta\chi^2(118) = 196.055$, $p < .086$, $\Delta df = 9$, $\Delta CFI = .001$. In regards to invariance across gender, multi-group model comparisons revealed a non-significant difference between metric and configural models. The difference between the scalar and metric models was also non-significant, thus supporting invariance of factor loadings and item intercepts.

[Insert Table 4]

Comparisons between a model constraining residual variances to be equal and the scalar model were non-significant, thus, providing support for the full uniqueness invariance across gender. Comparisons between metric and configural models across the three different competitive level groups (i.e., regional, national and international) did not reveal significant differences, thus, invariance of factor loadings was supported. The difference between scalar and metric models was also non-significant, supporting the invariance of item intercepts. Regarding uniqueness, a non-significant difference was obtained when comparing a model constraining residual variances and the scalar model; thus, full uniqueness invariance was supported across competitive level.

SIAM scores replicability

A small sample of participants drawn from five teams completed both SIAM versions (hard-copy and web-based). Sixteen athletes completed both versions one day to one week after the first administration. A group of ten participants completed the hard-copy version first, while the remaining six completed the versions in the opposite order.

Significant correlations were found for emotion, visual and tactile indicators (Table 1).

Non-significant differences were found between the subscales.

Discussion

The purpose of this study was to evaluate the validity and reliability of hard-copy and web-based formats of the Spanish version of the SIAM. Overall, the results extend previous psychometric findings regarding the assessment of sport imagery and provide support for the measurement properties, measure invariance and test-scores replicability over one-week of the Spanish version of the SIAM.

Scale descriptives for the Spanish SIAM were in line with those reported for hard-copies of the instrument developed in the Finnish (Watt & Jaakkola, 2003), Polish (Budnik-Przybylska et al., 2014), and Swedish (Weibull et al., 2009) languages.

Specifically, higher mean values were reported for ease, speed and visual dimensions obtained from both the hard-copy and web-based versions. Much lower means were reported for the gustatory and olfactory sub-scales compared to those for all other sub-scales in both hard-copy and web-based versions of the SIAM. The highest intercorrelations were found for the visual and vividness subscales for both hard-copy and web-based versions of the SIAM. Lowest intercorrelations were observed between the gustatory and olfactory senses and the other imagery ability characteristics (e.g., vividness). These findings are in line with the intercorrelation patterns observed for

other language versions (e.g., Budnik-Przybylska et al., 2014). The intercorrelations between the emotion subscale and generational elements of imagery ability were similar to those reported for the Polish version. Reliability results indicated acceptable to good levels of internal consistency for all scores except the ease subscale (values of .56 for hard-copy and .62 for web-based versions). Composite reliability scores, however, which are based on factor loadings and measurement errors, were higher than the recommended level of .50.

Factor structure of the hard-copy and web-based versions of the SIAM was examined for several competing models. ESEM for the 3-factor (i.e., generation, single sense, and feeling) model previously proposed by Watt et al. (2003) demonstrated slightly better fit to the data than the 3-factor CFA model for both versions. This was due to the fact that in ESEM models cross-loadings between factors are allowed, while in CFA all cross-loadings are constrained to zero. However, the consideration of modification indices prompted in CFA, allowing residuals associated with some items to correlate demonstrated superior fit for the proposed 3-factor CFA model. Our results indicated that all fit indices for the 3-factor CFA model, in particular for the web-based version were above the recommended cut off values (Hu & Bentler, 1999) and slightly superior to those reported in other international samples. These findings are comparable to samples that have used English (Spittle et al., 2005), Finnish (Watt & Jaakkola, 2003), Polish (Budnik-Przybylska et al., 2014) and Swedish (Weibull et al., 2009) language versions of the SIAM. Moreover, our findings demonstrate the reproducibility of the factor structure for both hard-copy and web-based Spanish SIAM supporting the notion that the instrument is able to substantiate the multisensory and multidimensional nature of imagery ability in a sport context.

As an extension of previous research examining the psychometric properties of the SIAM, we tested measurement invariance across hard-copy and web-based forms of administration, gender, and athletes' competitive level. Non-significant differences were found between metric and configural models supporting the invariance of factor loadings. Support for partial invariance of the scalar model was found after releasing the constraint of the kinaesthetic intercepts. Invariance of uniqueness was achieved after excluding kinaesthetic, duration and auditory dimensions of imagery, providing support for the notion that the rest of items are measured with the same degree of reliability. Thus, it seems that the environmental conditions for the evaluation of imagery ability (e.g., presence of a researcher or other auditory background noises, possibility of enhanced focus on internal aspects of the performance) may have had an effect of the reported specific imagery dimensions. Full measurement invariance including metric, scalar and uniqueness was found across gender and athletes' competitive levels. Taken together the findings indicate that the Spanish version of the SIAM can be used in the assessment of athletes' imagery ability, allowing for comparisons across male and female athletes, and participants involved in sports at regional, national or international level.

Test of replicability calculated for a one-week interval comparison revealed moderate to very good correlations for the specific dimensions in the generation and feeling factors, varying from .36 for ease to .71 for emotion. In contrast, low correlations were found for the different senses ranging from .18 for olfactory to .24 for auditory. One limitation of this study is the fact that while the participants were asked to complete the same procedures, they were not given instructions to imagine the same specific situations. Thus, it may well be that participants were imagining different situations, which may explain these differences. Another possible explanation for the

low correlations may be the fact that during the retest assessment the participants were asked to complete the alternative form of administration, thus, if during the first evaluation they used the hard-copy of the SIAM, during the second evaluation they were asked to complete the web-based version, and vice versa. Furthermore, these results should be interpreted with caution due to the small number of participants.

The focus of the present study was on examining possible differences between both versions this is the reason why we chose this replicability design. However, further research investigating replicability of the scores over time may include results derived from a series of same format administrations. Although imagery ability is not expected to change dramatically, especially when the participants are not undergoing an imagery-training program, longer intervals between re-administration could also be recommended in further research.

Future research replicating the findings of invariance should focus on ascertaining the extent to which the results from either a web-based or hard-copy version of the instrument are equivalent. Continued examination of predictive and concurrent validity of the SIAM, incorporating physiological indices that would complement self-report data is also warranted. Furthermore, a number of applied implications for the use of the SIAM exist. For instance, sport psychology practitioners may use the SIAM as a tool for validating the effectiveness of their interventions. Sport psychology researchers, may also use the SIAM as a screening tool to develop more effective imagery intervention programs targeting specific dimensions, depending on the requirements of the situation.

References

- Arvinen-Barrow, M., Weigand, D. A., Thomas, S., Hemmings, B., & Walley, M. (2007). Elite and novice athletes' imagery use in open and closed sports. *Journal of Applied Sport Psychology, 19*, 93-104. doi:10.1080/10413200601102912
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397-438. doi:10.1080/10705510903008204
- Bhasavanija, T., Vongjaturapat, N., Morris, T., & Muangnapo, P. (2011). Validation of the sport imagery ability measure in Thai for golf. *Thai Journal of Sports Sciences, 8*, 1-18.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guilford Press.
- Buchanan, T. (2003) Internet-based questionnaire assessment: appropriate use in clinical contexts. *Cognitive Behavior Therapy 32*, 100-109.
- Budnik-Przybylska, D., Karasiewicz, K., Morris, T., & Watt, A. (2014). Reliability, factor structure, and construct validity of the Polish version of the sport imagery ability measure. *Current Issues in Personality Psychology, 2*(4). doi:10.5114/cipp.2014.47444
- Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema, 20*, 872-882.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: the tailored design method (4th ed.). Hoboken, NJ: John Wiley.
- Farmer, A. Y., & Farmer, G. L. (2014). *Research with diverse groups: research designs and multivariate latent modeling for equivalence*. New York, NY: Oxford University Press.

- Gregg, M., & Hall, C. (2006). Measurement of motivational imagery abilities in sport. *Journal of Sports Sciences, 24*, 961-971. doi:10.1080/02640410500386167
- Hall, C. R., & Martin, K. A. (1997). Measuring movement imagery abilities: a revision of the movement imagery questionnaire. *Journal of Mental Imagery, 21*, 143-154.
- Hardy, L., & Callow, N. (1999). Efficacy of external and internal visual imagery perspectives for the enhancement of performance on task in which form is important. *Journal of Sport and Exercise Psychology, 21*, 95-112.
- Hirsch, O., Hauschild, F., Schmidt, M. H., Baum, E., & Christiansen, H. (2013). Comparison of web-based and paper-based administration of ADHD questionnaires for adults. *Journal of Medical Internet Research, 15*(3), e47. doi:10.2196/jmir.2225
- 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.
- Krille, C., Wuttke, E. & Seifried, J. (2017). Mode effects in correcting students' errors: a comparison of computer-based and paper-pencil tests. In E. Wuttke, & J. Seifried (Eds). *Professional Error Competence of Preservice Teachers: Evaluation and Support* (pp.47-74). Cham, Switzerland: Springer.
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, R., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439-476.
- Martens, R. (1982). Imagery in sport. In M. L. Howell & A. W. Parker (Eds.), *Proceedings of the Australian sports medicine federation international*

- conference. Vol. 8. Sports medicine: medical and scientific aspects of elitism in sport* (pp. 213-230). Brisbane, Australia.
- Mendes, P., Marinho, D., & Petrica, J. (2015). Comparison between genders in imagery ability in Portuguese basketball practitioners. *Journal of Physical Education and Sport, 15*, 391-395.
- Morris, T., Spittle, M., & Perry, C. (2004). Mental imagery in sport. In T. Morris & J. Summers (Eds.), *Sport psychology: theory, applications and issues* (pp. 344-387). Brisbane, Australia: Wiley & Sons.
- Morris, T., Spittle, M., & Watt, A. (2005). *Imagery in Sport*. Champaign, IL: Human Kinetics.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*, 173-184.
- Reips, U. D. (2006). Web-based methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 73-85). Washington, DC: American Psychological Association.
- Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: development and assessment of a revised version of the vividness of movement imagery questionnaire. *Journal of Sport & Exercise Psychology, 30*, 200-221.
- Sheehan, P. W., Ashton, R., & White, K. (1983). Assessment of mental imagery. In A. A. Sheikh (Ed.), *Imagery: current theory, research, and application* (pp. 189-221). New York: Wiley & Sons.

- Simonsmeier, B. A., & Buecker, S. (2017): Interrelations of imagery use, imagery ability, and performance in young athletes. *Journal of Applied Sport Psychology*, 29, 32-43. doi:10.1080/10413200.2016.1187686
- Spittle, M., Watt A.P., & Morris, T. (2005). Psychometric comparison of measures of sport imagery ability and sport imagery use. In T. Morris et al. (Eds.), *Proceedings of the 11th World Congress of Sport Psychology: Promoting health and performance for life*. Sydney, Australia: International Society of Sport Psychology.
- Tu, C. Y., Tseng, M. M., Chang, C. H., & Lin, C. C. (2017). Comparative validity of the Internet and paper-and-pencil versions of the Night Eating Questionnaire. *Comprehensive Psychiatry*, 75, 53-61.
- Wang, C-C., Liu, K-S., Cheng, C-L., & Cheng, Y-Y. (2013). Comparison of web-based versus paper-and-pencil administration of a humor survey. *Computers in Human Behavior* 29, 1007–1011. doi:10.1016/j.chb.2012.12.029
- Watt, A. (2003). Development and validation of the Sport Imagery Ability Measure (Unpublished doctoral dissertation). Victoria University, Melbourne, Australia.
- Watt, A., & Jaakkola, T. (2003). Reliability and factor structure of the Finnish version sport imagery ability measure. In R., Stelter (Ed.), *Proceedings of the 11th European Congress of Sport Psychology. New approaches to exercise & sport psychology* (p.182). Copenhagen, Denmark.
- Watt, A. P., Morris, T., & Andersen, M. B. (2004). Issues in the development of a measure of imagery ability in sport. *Journal of Mental Imagery*, 28, 149-180.
- Weibull, F., Johnson U., & Watt, A. (2009). Psychometric analysis of the Swedish version of the sport imagery ability measure. In A. Baria et al. (Eds.), *Proceedings on CDROM of the ISSP 12th World Congress of Sport Psychology:*

Meeting New Challenges and Bridging Cultural Gaps in Sport and Exercise Psychology. Marrakesh, Morocco: International Society of Sport Psychology.

- Williams, S. E., & Cumming, J. (2011). Measuring athlete imagery ability: the sport imagery ability questionnaire. *Journal of Sport & Exercise Psychology, 33*, 416-440.
- Williams, S. E., & Cumming, J. (2012). Athletes' ease of imaging predicts their imagery and observational learning use. *Psychology of Sport and Exercise 13*, 363-370. doi:10.1016/j.psychsport.2012.01.010
- Williams, S. E., Cumming, J., Ntoumanis, N., Nordin-Bates, S. M., Ramsey, R., & Hall, C. (2012). Further validation and development of the Movement Imagery Questionnaire. *Journal of Sport & Exercise Psychology, 34*, 621-646.
- Wyrick, D. L., & Bond, L. (2009). Reducing sensitive survey response bias in research on adolescents: a comparison of web-based and paper-and-pencil administration. *Applied Research Brief: Measurement Issues, 25*, 5. doi:10.4278/ajhp.080611-ARB-90