Original Article

Validation of a Computer-Based Version of the Vividness of Movement Imagery Questionnaire

Stephan F. Dahm

Department for Psychology and Medical Sciences, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

Abstract. A computer-based version of the Vividness of Movement Imagery Questionnaire (VMIQ-2) was tested and compared with the print version. Psychometric characteristics of the print version were replicated with the computer-based version using a between-subject design in Study 1 (n_1 = 240; n_2 = 254) and using a within-subject design in Study 2 (N = 285). Confirmatory factor analyses replicated the three-dimensional structure of the VMIQ-2. Model fits, test–retest reliability, interfactor correlations, and correlations with other questionnaires did not differ significantly between the computer-based and print versions. Convergent validity was shown by significant correlations with the Movement Imagery Questionnaire 3. The computer-based version of the VMIQ-2 is considered a valid instrument to measure vividness of action imagery.

Keywords: vividness of action imagery, questionnaire, assessment medium, paper–pencil, computer-based, imagery ability

Originally, the Vividness of Movement Imagery Questionnaire (VMIQ-2; Roberts et al., 2008) has been set up in a print version. However, computer-based questionnaires offer several advantages which will be outlined in the following. It was the aim of the present study to investigate whether a computer-based version of the VMIQ-2 measures the same construct and reveals similar item characteristics as the print version. A computer-based version of the Sport Imagery Ability Measure (Watt et al., 2004) has already been validated and showed similar item characteristics as the print version (Ruiz et al., 2018). However, a computer-based version of the more frequently used VMIQ-2 (Roberts et al., 2008) has not been validated yet.

By assessing the vividness of imagined movements, the questionnaire measures an important dimension of action imagery (Cumming & Eaves, 2018). Action imagery refers to the representation of an action without actually performing it (Jeannerod, 1994). Instead of the more common wording *motor imagery*, the term *action imagery* is used in the present studies to emphasize that the imagination of movements goes beyond kinesthetic (motor) representations involving also visual, tactile, acoustic, olfactory, and gustatory representations (Cumming & Eaves, 2018). Particularly, visual and kinesthetic aspects have received more attention than other modalities (Jeannerod, 1994; Roberts et al., 2008; Williams et al., 2012). This may be since, in contrast to the other modalities, visual and kinesthetic aspects provide relevant information in all kinds of actions. For the visual modality, it is important to take perspectives into account. External visual imagery (EVI) involves imagining oneself from a third-person perspective like watching oneself on a video. Internal visual imagery (IVI) involves imagining from a first-person perspective like seeing with one's own eyes. Kinesthetic imagery (KIN) involves imagining the feeling of the movement. The preferences for modalities and perspective differ across imagers (Dana & Gozalzadeh, 2017; Moran et al., 2012). The VMIQ-2 aims to assess an individual's ability to create vivid images in EVI, IVI, and KIN of 12 daily actions such as kicking a stone or riding a bike (Roberts et al., 2008). The three subscales are expected to correlate with each other, as they all contribute to vividness of action imagery (Dahm et al., 2019; Roberts et al., 2008). Therefore, in Study 1, the expected three-dimensional structure (Dahm et al., 2019; Roberts et al., 2008) will be tested against a general-factor model (Eid et al., 2018; Gignac & Kretzschmar, 2017).

Being part of the same construct (action imagery ability), vividness of action imagery, ease of action imagery generation, controllability of action imagery, and maintenance of action imagery are expected to converge (Dahm, 2020). Indeed, strong correlations (.30 < r < .70) have been observed between vividness of action imagery (measured with the VMIQ-2) and the ease of action imagery generation (measured with the Movement Imagery Questionnaire

https://doi.org/10.1027/2698-1866/a000022

Revised [MIQ-R]; Roberts et al., 2008). Furthermore, weak correlations (.10 < r < .30) have been observed between action imagery generation (measured with the MIQ-3) and the imagery subscales (skill, strategy, goal, affect, and mastery) of the Sport Imagery Ability Questionnaire (SIAQ; Williams et al., 2012). A comprehensive analysis of convergent validity combining the latest versions of the most common action imagery questionnaires (VMIQ-2, MIQ-3, and SIAQ) has not been performed yet. Therefore, convergent validity and discriminant validity will be explored in more detail in Study 2.

The VMIQ-2 (Roberts et al., 2008) has been used in studies to exclude participants who are not able to generate vivid imagery (Callow et al., 2013; O'Shea & Moran, 2019). Furthermore, it can be used to evaluate whether the vividness of action imagery changes over time (Abraham et al., 2019). In applied settings, the assessment of vividness of action imagery may predict individual improvements after action imagery practice. Action imagery practice refers to the repetitive and systematic use of action imagery to improve executed performance (Driskell et al., 1994). It has been shown that more vivid imagers benefit more from action imagery practice than less vivid imagers (Isaac, 1992). Looking at it the other way round, high-level athletes reported more vivid images of actions than low-level athletes (Roberts et al., 2008). Therefore, higher ratings in vividness of action imagery were expected in active subjects than in nonactive subjects in Study 2.

The VMIQ-2 is designed to assess vividness of action imagery in active and nonactive adults. In Study 1, two younger (18–35 years) student samples were compared. In Study 2, an older (18–65 years) student and nonstudent sample was assessed. The present studies aimed to validate a computer-based version of the German version of the VMIQ-2 (Dahm et al., 2019). In the present study, computer-based testing does not imply adaptive testing (Green et al., 1984), but only presenting the items of the print version in its original order. Hence, in the present study, computer-based testing refers to the presentation of items on a notebook screen and that participants enter their responses using an external computer mouse. In contrast, in the print version, the items are presented on paper and data are entered with a pencil. It was the aim of the present studies to disentangle whether there are differences between computer-based assessments and print assessment per se, thereby avoiding other possible confounds (e.g., online computer-based assessments vs. offline print assessments or adaptive computer-based assessments vs. nonadaptive print assessments). Furthermore, Study 2 goes beyond that comparison by including subtle adaptations to improve the assessment.

The advantages of computer-based testing are threefold. First, in paper–pencil questionnaires, participants enter data in a first step. In a second step, to analyze the data, an experimenter enters the data into a computer. In computer-based testing, there is no need for a second step for data entry. This saves time and prevents possible copying errors. Second, response times can be assessed more accurately in computer-based testing than in paper–pencil tests. For instance, it is possible to investigate whether response times differ between items. Third, in computer-based testing, it is possible to give immediate feedback. For instance, if items were left out by the participant, the items can be highlighted visually. As such, missing data can be prevented. Furthermore, individual criteria may be set and analyzed in real-time during the assessment. For instance, in studies, participants may be excluded at the very start of the study. In addition, participants may receive an immediate individual feedback about their scoring characteristics at the end of the assessment.

Some guidelines (Groth-Mamat & Schumaker, 1989; Schroeders, 2009; Ziegler, 2020) suggest that computerized testing methods should be checked for validity and equality with the print version (Bugbee, 1996). Previous investigations have shown that print and computer-based versions may differ in item difficulties and cognitive workload (Noyes & Garland, 2008). Furthermore, information and communication technology literacy has been shown to contribute on differences between the print and computer-based assessments in reading comprehension (Wang et al., 2007, 2008). Considering action imagery, it has been shown that the methodological setting can influence the assessment (Anuar et al., 2016). More vivid images were reported if participants incorporated the physical position and received external materials that were appropriate for the required action than if participants were sitting in a quiet room without additional materials (Anuar et al., 2016).

The present validation of a computer-based version of the German VMIQ-2 (Dahm et al., 2019) involved two studies. In Study 1, a between-subject design was applied. The original data of the German print version (Dahm et al., 2019) were compared with the data of a computer-based version using confirmatory factor analyses. Furthermore, test–retest reliability was compared between both versions. In Study 1, metric measurement invariance was expected to indicate that the print and computer-based versions measure the same construct. Scalar (or even strict) measurement invariance was not expected because the samples may differ per se as they were not randomly picked from the same population. To overcome the latter issue, a within-subject design was applied in Study 2. Using the same sample for the print and computer-based versions, equal means and equal variances were expected indicating strict measurement invariance. Furthermore,

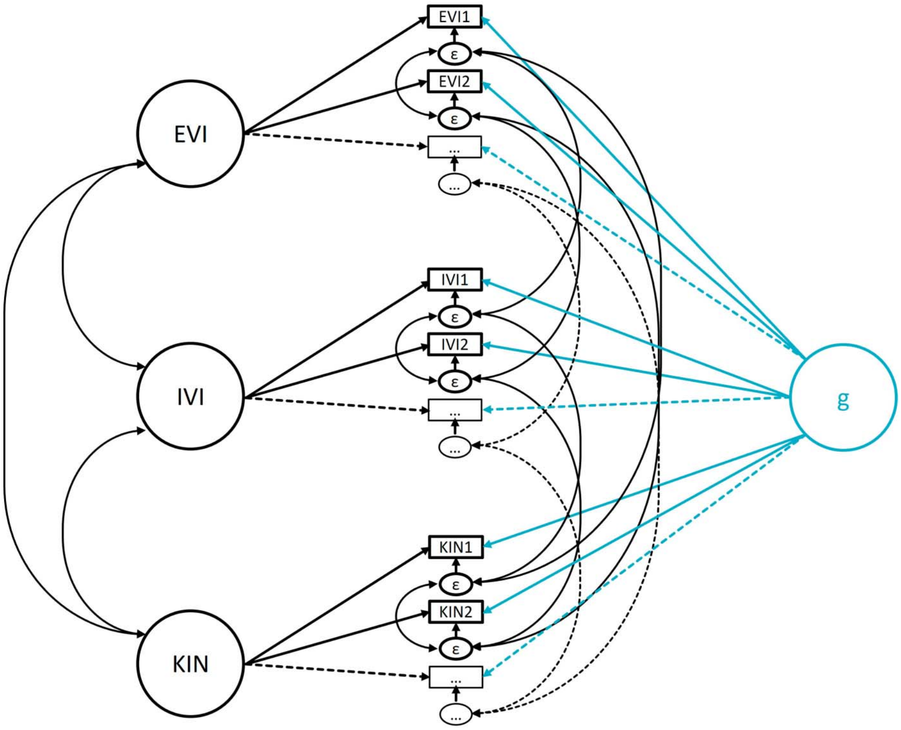

Figure 1. Depiction of the factor model. Circles correspond to latent factors (EVI = external visual imagery, g = general action imagery, IVI = internal visual imagery, KIN = kinesthetic imagery). Ellipses correspond to error variances. Each latent factor on the left side involved 12 items, as indicated with the dashed line. The higher-order g-factor was only included in the g-factor models where correlations between the specific factors were set to zero. Only in the optimized three-factor models, the error variances of action categories (e.g., EVI1 and EVI2) were grouped together.

equal correlations with other constructs were expected to indicate tau equivalence (Allen & Yen, 1979). Study 2 goes beyond previous studies by examining the nomological net of vividness of action imagery with several questionnaires as criterion variables to check for discriminant and convergent validity. In addition, construct validity is tested in Study 2 by comparisons of active and nonactive subpopulations.

Study 1

In Study 1, differences between the print and computer-based versions were intentionally kept at a minimum. To test for equivalence, it is essential to restrict the use of technology to the essential parts of the measurement (Schroeders, 2009).

Each item of the VMIQ-2 involves two theoretical aspects: imagery modality and imagined action (Figure 1). Therefore, a multitrait multimethod (MTMM) approach with a correlated trait–correlated uniqueness model (Kenny, 1976) assuming 3 factors (EVI, IVI, KIN) and 12 actions has been shown to fit better than models without the MTMM approach (Dahm et al., 2019; Roberts et al., 2008). Hence, not only items of the same modality factor incorporate shared variances, but also items of the same action (imagined in different modalities). Furthermore, six action categories have been proposed with each category represented by two items (Dahm et al., 2019; Roberts et al., 2008). Therefore, shared variances were expected for items of the same action category (imagined in the same modality).

For the factors of the VMIQ-2, the MTMM model with three factors has been shown to fit better than a single-factor model (Dahm et al., 2019; Roberts et al., 2008). However, a better fit of the correlated-factor model over the one-factor model does not necessarily indicate that the factors are distinct (Gignac & Kretzschmar, 2017). To our knowledge, a g-factor model has not been tested yet although it has been assumed due to the interfactor correlations ($.39 < r < .63$; Dahm et al., 2019; Roberts et al., 2008). In a nested/bifactor model, a g-factor may indicate general action imagery ability, whereas the subscales indicate preferences for EVI, IVI, and KIN.

Methods: Study 1

Participants

On the basis of internal consistency (α = .90) and interfactor correlations (r = .40) of previous studies (Dahm et al., 2019; Roberts et al., 2008), a minimal sample size of 220 was estimated (Kretzschmar & Gignac, 2019). Two

Table 1. Sociodemographic data of the samples in Experiment 1

Sociodemographic variables	Print	Computer
Age in years, *M* (*SD*)	24 (5)	24.5 (4)
Handedness index, *M* (range)	86.9 (−100, 100)	92.3 (40, 100)
Sex, *N*		
Female	175	166
Male	79	112
Highest educational degree, *N*		
Compulsory school	3	6
Secondary school	4	7
Apprenticeship	16	24
High school	187	170
College/academy	5	6
University	39	65

Note. A handedness index of −100 indicates *complete left handedness*, 0 indicates *ambiguous handedness*, and +100 indicates *complete right handedness* (Oldfield, 1971).

hundred seventy-eight participants filled out the computer-based version. A subsample of 187 participants completed the computer-based version again after 6 weeks. The print version data of 254 participants of a previous study were reanalyzed (Dahm et al., 2019). A subsample of 78 participants completed the print version again after 6 weeks. Both computer-based and print data were collected in individual one-to-one assessments in a laboratory. Age, handedness index (Oldfield, 1971), sex, and the highest educational degree of the main samples are shown in Table 1. All participants gave informed consent. Ethical approval was given by the local ethics committee of the university. Students received course credit for their participation.

Material and Assessment

The German version of the VMIQ-2 was used (Dahm et al., 2019). The VMIQ-2 includes 36 items (Roberts et al., 2008). In the introduction, the concepts of EVI, IVI, and KIN are explained. Twelve actions are then each imagined from an external visual perspective, from an internal visual perspective, and kinesthetically. After each imagination, participants rate the vividness of the imagined action on a rating scale from 1 (*perfectly clear and vivid*) to 5 (*no imagination, I only know that I am thinking about the action*). Note that low scores indicate vivid imagery which may be contraintuitive because high abilities are often indicated by high scores in other questionnaires (e.g., the MIQ-3; Williams et al., 2012). The proposed action categories (Isaac et al., 1986) are daily actions (walking, running), actions that involve precision (kicking a stone, bending to pick up a coin), actions that involve overcoming an obstacle (running upstairs, jumping sideways), actions that involve a manipulation of objects (throwing a stone into water, kicking a ball in the air), fast actions that involve balance (running downhill, riding a bike), and actions that involve the control of objects/balance in the air (swinging a rope, jumping off a high wall).

In the computer-based version, participants were informed about the total number of pages of the questionnaire in advance. Furthermore, the response behavior allowed an individual item order and response corrections for the 12 items on each page. To prevent biases of the previous mouse position, items were shown one below the other. The computer-based VMIQ-2 is provided at https://osf.io/xtmd9/. The freeware OpenSesame (Mathôt et al., 2012) which is available at https://osdoc.cogsci.nl needs to be installed in advance of the assessment. This software has the advantage that it runs on Windows, Linux, and MacOS systems. To use the script for data preparation, the freeware R (R Development Core Team, 2019) is available at https://www.r-project.org.

Data Analysis

Satorra–Bentler corrected confirmatory factor analyses were performed with the R package *lavaan* (Rosseel, 2012). The following fit indices were calculated: the scaled Satorra–Bentler χ^2 statistics (S-B χ^2), ratio of χ^2 to degrees of freedom (χ^2/df), the robust comparative fit index (*CFI*), standardized root mean square residual (SRMR), and the robust RMSEA. All models were analyzed with separate groups (print and computer-based). First, the original three-factor model (Roberts et al., 2008), an optimized three-factor model that takes action categories into account, a nested g-factor model (Eid et al., 2018; Gignac & Kretzschmar, 2017), and an optimized nested g-factor model that takes action categories into account were compared. Second, the best fitting model was used to test for measurement invariance across the print and computer-based versions. Measurement invariance was tested in a stepwise procedure by testing configural, metric, scalar, and strict invariance between samples (Putnick & Bornstein, 2016).

To compare the models, Satorra–Bentler difference tests were planned in case the configural model revealed a nonsignificant χ^2. However, due to the high number of degrees of freedom, previous analyses of the VMIQ-2 revealed significant χ^2 (Dahm et al., 2019; Roberts et al., 2008). Hence, nonsignificant χ^2 results were not expected. Therefore, alternative fit indices were considered in case the models revealed a significant χ^2. The alternative fit indices are considered good with (a) a χ^2/df ratio lower than 3:1, (b) a CFI higher than .95, (c) a SRMR lower than .1, and (d) a RMSEA lower than .05 (Hair et al., 2019).

McDonald's ω (Dunn et al., 2014) was calculated for internal consistency. The concordance correlation coefficient

(Lin, 1989) was calculated for test–retest reliability. The significance level was set at .05 for all analyses.

Results: Study 1

The original three-factor model involved the assumptions that the latent factors (EVI, IVI, KIN) correlate and that items of the same action correlate. In addition to the assumptions of the original three-factor model, items of the same action category were expected to correlate in the optimized three-factor model. In addition to these assumptions, an independent higher-order general-factor and orthogonal specific factors were expected in the g-factor models (Eid et al., 2018). The model fits of the original three-factor model, the optimized three-factor model, the original g-factor model, and the optimized g-factor model are shown in Table 2.

All models indicated a significant S-B χ^2 ($p < .001$) indicating poor model fit. Nevertheless, the alternative fit indices indicated acceptable fit for the g-factor model and the optimized 3-factor model. The original 3-factor model revealed CFI (CFI = .92) below the critical value of .95 and RMSEA (RMSEA = .056) above the critical value of .05 and was therefore rejected. Although the optimized g-factor model resulted in tendentially better fit indices than the optimized 3-factor model, the g-factor model was discarded after inspection of the factor loadings. The theoretical concept of the g-factor would have been a factor that loads (similarly high) on all items indicating general action imagery ability. However, the g-factor loaded higher on the items of one of the specific factors than on the items of the other two factors. Furthermore, and because this is critical, the factor loadings of this specific factor (loading on the g-factor) were incoherent (range from −.13 to +.77 in the print version; range from −.06 to +.53 in the computer-based version), resulting in low reliability for this specific factor (EVI: ω = .83; IVI: ω = .64; KIN: ω = .88). Therefore, the optimized 3-factor model was chosen for further analysis.

In addition to the assumptions of the optimized three-factor model, the metric invariance model was additionally constrained to equivalent factor loadings in the print and computer-based versions (Table 2). Because the alternative fit indices of the metric model remained acceptable, a scalar model was analyzed that accounts for equal intercepts (means) in the print and computer-based versions. Because the alternative fit indices of the scalar model remained acceptable, a strict model was analyzed that accounts for equal item residuals (sum of specific and error variance).

Descriptive characteristics of the subscales involving means, *SD*s, skew, kurtosis, McDonald's ω (Dunn et al., 2014), minimum factor loadings, and maximum factor loadings are shown in Table 3.

Interfactor correlations between EVI, IVI, and KIN were calculated with Pearson correlations (Table 4). To compare interfactor correlations of the print and computer-based versions, Fisher's *z* tests were calculated. Correlations between KIN and visual imagery (IVI and EVI) did not differ significantly between the print and computer-based versions. However, the correlation between the visual perspectives EVI and IVI was significantly higher in the computer-based version than in the print version. Therefore, another model was analyzed for the computer-based version that considered EVI and IVI as one single factor for vision. However, the model fits were poor, S-B χ^2 (539) = 1,083, $p <. 001$ $\chi^2/df = 2$, CFI = .89, SRMR = .07, RMSEA = .071, 90% CI (.065, .077).

Test–retest reliability was calculated for EVI, IVI, KIN, and the general VMIQ-2 score (Table 4). Fisher's *z* tests revealed no significant differences in test–retest reliability between the print and the computer-based versions.

Discussion: Study 1

The psychometric parameters of the print and the computer-based versions of the VMIQ-2 (Dahm et al., 2019) were compared in a between-subject design.

Table 2. Robust and scaled model fits of confirmatory factor analyses testing for invariance between the print and computer-based versions of the Vividness of Movement Imagery Questionnaire-2

Model	S-B χ^2 (*df*)	*p*	χ^2/df	CFI	SRMR	RMSEA [90% CI]
Three-factor original	1771 (1,110)	<.001	1.6	.92	.065	.056 [.051, .061]
g-factor original	1,555 (1,044)	<.001	1.5	.94	.054	.051 [.046, .056]
g-factor optimized	1,325 (1,008)	<.001	1.3	.96	.06	.041 [.034, .047]
Three-factor optimized	1,459 (1,074)	<.001	1.4	.96	.059	.043 [.038, .049]
Three-factor metric	1,495 (1,107)	<.001	1.4	.96	.065	.043 [.037, .048]
Three-factor scalar	1,546 (1,140)	<.001	1.4	.95	.066	.043 [.037, .048]
Three-factor strict	1,631 (1,230)	<.001	1.3	.95	.068	.042 [.036, .047]

Table 3. Characteristics of the factors of the print and computer-based versions of the VMIQ-2

	VMIQ-2 print (N = 254)			VMIQ-2 computer (N = 240)		
	EVI	IVI	KIN	EVI	IVI	KIN
M	2	1.8	2	1.9	1.7	1.9
SD	0.7	0.6	0.7	0.7	0.6	0.7
Skew	0.6	0.9	0.8	0.4	0.6	1.2
Kurtosis	0.2	0.7	0.8	−0.5	−0.5	2.2
McDonald's ω	.91	.9	.91	.92	.91	.93
a_{min}	.59	.6	.64	.59	.59	.65
a_{max}	.74	.73	.73	.73	.72	.75

Note. a_{min} = minimum factor loading, a_{max} = maximum factor loading. Factor loadings stem from the strict 3-factor model. EVI = external visual imagery, IVI = internal visual imagery, KIN = kinesthetic imagery, VMIQ-2 = Vividness of Movement Imagery Questionnaire.

Table 4. Comparison of the interfactor correlations and test–retest reliability in the print and computer-based versions

	Print	Computer	Z	p
Interfactor correlations				
EVI × IVI	.53	.69	2.85	.004
EVI × KIN	.43	.48	0.69	.486
IVI × KIN	.58	.57	0.17	.868
Test–retest reliability				
EVI	.62	.66	0.75	.454
IVI	.61	.64	0.54	.587
KIN	.64	.67	0.58	.562
General	.69	.77	1.9	.057

Although χ^2 was significant, the alternative fit indices indicated an acceptable fit of the expected three-dimensional structure with 12 actions from six action categories. All models of measurement invariance (configural, metric, scalar, and strict) showed a χ^2/df ratio lower than 3, CFI equal or higher than .95, SRMR lower than .10, and RMSEA lower than .05 (Hair et al., 2019). Hence, the computer-based version and the print version did not significantly differ in factor loadings and latent factor means. Similarly, the test–retest reliability did not significantly differ between both versions. This indicates that both versions measure the same dimensions for vividness of action imagery.

However, the correlation between EVI and IVI was higher in the computer-based version than in the print version. On the one hand, the high correlation may indicate that people do not distinguish as much between visual perspectives when they look at a screen than when they look at paper. On the other hand, the high correlation may be caused by methodological issues. Longer text passages are more difficult to read at the screen than on paper (Mazzeo & Harvey, 1988) and are therefore more likely to be skipped. Therefore, it is possible that in the computer-based version participants did not read the introduction of the VMIQ-2 as thoroughly as in the print version. If participants did not read the instructions thoroughly, they may not have detected the difference between EVI and IVI within the introductory instructions, despite this information being provided. Without knowing about this distinction, participants may have responded as if the EVI and IVI items were the same items. As it was not possible to go back to the EVI items to search for differences, this may have blurred the data. Therefore, in Study 2, a comprehension question was added directly after the introduction of the VMIQ-2. Moreover, an option was added to go back and repeat the previous page of the questionnaire.

Study 2

In Study 2, the print and computer-based versions were tested in a within-subject design. A replication of the acceptance of the optimized 3-factor model from Study 1 was expected. As in Study 1, strict measurement invariance was expected to indicate equivalence of the print and computer-based versions.

The within-subject design made it possible to check for tau equivalence (Allen & Yen, 1979), namely the correlations between vividness of action imagery and other constructs which were expected to be similar in the print and computer-based versions. The selected questionnaires were expected to depict the whole range of strong ($r > .50$), medium ($.30 < r < .50$), weak ($.10 < r < .30$), and no correlations (Cohen, 1992) with the VMIQ-2. Strong correlations between EVI, IVI, and KIN with the same factors of the Movement Imagery Questionnaire 3 (Williams et al., 2012) were expected due to construct similarity. Medium correlations ($.30 < r < .50$) with the Sport Imagery Ability Questionnaire (Simonsmeier & Hannemann, 2017) were expected because it measures action imagery but different types of subdimensions (Pithan & Dahm, 2015). Weak correlations with physical self-concept (Stiller et al., 2004), general self-efficacy (Hinz et al., 2006), mindfulness (Michalak et al., 2016), and self-rated concentration (Bankstahl & Görtelmeyer, 2013) were expected because self-ratings of abilities and self-awareness may be influenced by similar biases such as the social desirability bias and the self-protecting bias (Dahm, 2020). Correlations with general personality traits (Rammstedt & John, 2007), rumination (Trapnell & Campbell, 1999), and sport orientation (Elbe, 2004) were expected to be close to zero ($-.10 < r < .10$).

Because eye fixations and theta band voltage density during reading indicated rather lower effort at a tablet than on paper (Kretzschmar et al., 2013), it was expected that reading durations and durations to fill out the questionnaire are faster in the computer-based version than in the print version. Furthermore, it was expected that comprehension accuracy does not differ significantly between the print and computer-based versions (Kretzschmar et al., 2013).

In addition, an exploratory analysis was planned to investigate whether active subjects differ from nonactive subjects in their vividness of action imagery. Active subjects are here defined as subjects doing a lot of sports in contrast to nonactive subjects who do not perform any sport activities. Because more vivid imagers show greater execution benefits after action imagery practice than less vivid imagers (Isaac, 1992), the reverse may also hold true. Hence, subjects who perform actions frequently may differ in how they imagine actions from subjects who perform actions infrequently. Regarding the assessment medium, similar effects were expected for the print and computer-based versions.

Methods: Study 2

Participants

As in Study 1, a minimal sample size of 220 was estimated. In total, 297 participants were recruited in Study 2 who did not participate in Study 1. Data were collected in individual one-to-one assessments in private settings. The data of six participants were not taken into analysis due to missing data points. The data of another six participants were not taken into analysis because they filled out the VMIQ-2 items both under three seconds per item and without variance, which may indicate poor attention to the questionnaire (Maniaci & Rogge, 2014). The remaining 285 participants (179 female, 106 male) were on average 29.8 years old (SD = 10.8). All participants gave informed consent. Ethical approval was given by the local ethics committee of the university. Students received course credit for their participation.

Material and Assessment

As in Study 1, the German version of the VMIQ-2 (Dahm et al., 2019) was used in print and computer-based. The timescale between completing each version of the VMIQ-2 lasted between 20 and 45 minutes depending on participants' work pace. The computer-based version was adapted according to the considerations of Study 1. To reassure that the introduction of the VMIQ-2 was understood, a comprehension question was added. Participants were asked which dimensions of action imagery will be rated on the three pages of the VMIQ-2: (a) external visual imagery, (b) internal visual imagery, (c) KIN, or (d) acoustic imagery. If the answer was correct (a, b, and c), participants continued with the questionnaire. If the answer was incorrect, the introduction was shown again. Furthermore, going back and forth between the three pages was allowed by using the arrow keys on the keyboard.

Participants filled out the VMIQ-2 at the beginning and at the end of the study, once in print and once on the computer (counterbalanced across participants). In between, participants completed other questionnaires such as a German version of the Movement Imagery Questionnaire 3 (Williams et al., 2012), the Sport Imagery Ability Questionnaire (Simonsmeier & Hannemann, 2017), the Physical Self-Concept Scale (Stiller et al., 2004), the General Self-Efficacy Scale (Hinz et al., 2006), the Attention and Performance Self-Assessment Scale (Bankstahl & Görtelmeyer, 2013), a short version of the Big Five Inventory (Rammstedt & John, 2007), the Rumination Reflection Questionnaire (Trapnell & Campbell, 1999), the Sport Orientation Questionnaire (Elbe, 2004), and the Five Facet Mindfulness Questionnaire (Michalak et al., 2016). The order of the additional questionnaires was random. Participants received instructions on a computer indicating the upcoming questionnaire which was then handed out by the experimenter. The study lasted about 1 hour.

Data Analysis

As in Study 1, measurement invariance was tested in a stepwise procedure by testing configural, metric, scalar, and strict invariance between samples (Putnick & Bornstein, 2016) comparing alternative fit indices. Pearson correlations were calculated and compared using Fisher's z-tests. Reading duration and the duration to fill out the questionnaire were analyzed with between-subject t-tests. Reading comprehension was analyzed with a χ^2 test. To analyze whether participants' sport and exercise behavior influences vividness of action imagery, a mixed-model ANOVA was calculated. Participants indicated whether they exercised sports less than 1 hour per week (nonactive group: $N = 74$), between 1 and 4 hours per week (moderately active group: N = 133), or more than 4 hours per week (active group: N = 78).

Results: Study 2

Characteristics of the factors such as means, SDs, skew, kurtosis, McDonalds's ω (Dunn et al., 2014), minimum

Table 5. Characteristics of the factors of the print and computer-based versions of the VMIQ-2 (*N* = 240)

	VMIQ-2 print			VMIQ-2 computer		
	EVI	IVI	KIN	EVI	IVI	KIN
M	1.9	1.8	1.9	1.9	1.7	1.9
SD	0.7	0.7	0.7	0.7	0.6	0.7
Skew	0.6	1.1	1	0.7	1.2	1.1
Kurtosis	−0.2	1.5	0.7	0.1	2.3	1.9
McDonald's ω	.93	.93	.93	.93	.92	.93
a_{min}	.69	.68	.66	.69	.66	.67
a_{max}	.75	.79	.77	.75	.77	.78

Note. a_{min} = minimum factor loading, a_{max} = maximum factor loading. Factor loadings stem from the strict three-factor model. EVI = external visual imagery, IVI = internal visual imagery, KIN = kinesthetic imagery, VMIQ-2 = Vividness of Movement Imagery Questionnaire.

factor loadings, and maximum factor loadings of the print and computer-based versions of the VMIQ-2 are shown in Table 5.

As in Study 1, the alternative fit indices remained acceptable in the metric, scalar, and strict models (Table 6). Hence, the print and the computer-based versions of the VMIQ-2 did not significantly differ in factor loadings (metric), means (scalar), and variances (strict).

Correlations of the same factor between the print and computer-based versions were $r_{EVI \times EVI} = .83$, $r_{IVI \times IVI} = .83$, and $r_{KIN \times KIN} = .74$. Correlations of different factors between the print and computer-based versions were $r_{EVI\_PP \times IVI\_PC} = .63$. and $r_{IVI\_PP \times EVI\_PC} = .60$ ($Z = 0.57$, $p = .566$), $r_{EVI\_PP \times KIN\_PC} = .56$ and $r_{KIN\_PP \times EVI\_PC} = .5$ ($Z = 1$, $p = .322$), and $r_{IVI\_PP \times KIN\_PC} = .54$. and $r_{KIN\_PP \times IVI\_PC} = .54$ ($Z = 0, p > .99$). Interfactor correlations between EVI, IVI, and KIN did not differ significantly between the print and computer-based versions (Table 7). Correlations between other constructs and EVI, IVI, and KIN of the print and computer-based versions are shown in Table 7.

Additionally, the reading durations of the introduction of the VMIQ-2 and the durations to fill out the VMIQ-2 were analyzed. Only participants who filled out the questionnaire for the first time were taken into account for these analyses. Moreover, participants were split into two groups: those who filled out the computer-based version first (*N* = 128) and those who filled out the print version first (*N* = 122). Between-subject *t*-tests were calculated. The reading duration was significantly shorter in the computer-based version (*M* = 73 s, *SD* = 47 s) than in the print version (*M* = 92 s, *SD* = 42 s), t (248) = 3.4, $p = .001$, 95% CI [−8, 30], $d = 0.43$. Analogously, the duration to fill out the questionnaire was significantly shorter in the computer-based version (*M* = 189 s, *SD* = 97 s) than in the print version (*M* = 229 s, *SD* = 124 s), t (248) = 2.9, $p = .004$, 95% CI [13, 68], $d = 0.36$.

Furthermore, repetitions of the comprehension question were analyzed. A χ^2 test indicated no significant difference in the number of subjects who repeated the comprehension question in the computer-based version (32 out of 128) and the print version (25 out of 122), $\chi^2(1) = 0.72$, $p = .396$.

Means and standard errors of vividness of action imagery ratings in the nonactive, moderately active, and active group are shown in Figure 2. A mixed-model ANOVA with the between-subject factor activity (nonactive, moderately active, active) and the within-subject factors questionnaire version (computer-based, print) and modality (EVI, IVI, KIN) was performed on vividness of action imagery. The significant main effect modality, $F_{2,564} = 17.8$, $p < .001$, $\eta_p^2 = .06$, was modified by the significant interaction between activity, modality, and version, $F_{2,564} = 2.8$, $p = .026$, $\eta_p^2 = .02$. In both versions, the ratings were significantly lower in IVI than in EVI ($p_{max} = .003$), except for the ratings in the computer-based version of the nonactive group ($p = .089$). Additionally, in the moderately active group, the ratings were significantly lower in IVI than in KIN ($p_{max} = .003$). Furthermore, in the active group, the ratings were significantly lower in KIN than in EVI in the print version ($p = .002$), but not in the computer-based version ($p = .714$). In the print version, ratings in KIN were significantly lower in the active group than in the moderately active group ($p = .009$) and the nonactive group ($M = 2$, $p = .011$), whereas the latter did not significantly differ from each other ($p = .981$). In the computer-based version, these differences in KIN did not reveal significance ($p_{min} = .131$). The ratings did not significantly differ between the computer-based and print versions ($p_{min} = .075$), except for EVI ratings in the active group which were significantly lower in the print version than in the computer-based version ($p = .041$). The remaining effects were not significant (activity: $F_{2,282} = 2.9$,

Table 6. Robust and scaled model fits of confirmatory factor analyses testing for invariance between the print and computer-based versions of the Vividness of Movement Imagery Questionnaire-2

Model	S-B Χ² (*df*)	*p*	Χ²/*df*	*CFI*	*SRMR*	RMSEA [90% CI]
3-factor optimized	1,543 (1,074)	<.001	1.4	.95	.063	.049 [.043, .054]
3-factor metric	1,568 (1,107)	<.001	1.4	.95	.064	.048 [.042, .053]
3-factor scalar	1,596 (1,140)	<.001	1.4	.96	.064	.046 [.041, .052]
3-factor strict	1,640 (1,230)	<.001	1.3	.96	.064	.043 [.037, .048]

Table 7. Comparison of correlations of the print and computer-based versions

	VMIQ-2	Print	Computer	Z	p
Interfactor correlations					
EVI × IVI		.66	.68	0.4	.666
EVI × KIN		.60	.53	1.2	.221
IVI × KIN		.63	.61	0.4	.700
Discriminant and convergent construct validity					
MIQ-3 EVI	EVI	−.64	−.54	1.8	.067
MIQ-3 IVI	IVI	−.61	−.59	0.3	.711
MIQ-3 KIN	KIN	−.55	−.46	1.4	.151
SE	EVI	−.24	−.26	0.3	.800
	IVI	−.24	−.26	0.3	.800
	KIN	−.20	−.15	0.6	.540
SIAQ	EVI	−.29	−.33	0.5	.599
	IVI	−.37	−.37	0	>.999
	KIN	−.30	−.21	1.1	.253
PSC	EVI	−.27	−.31	0.5	.604
	IVI	−.25	−.27	0.3	.799
	KIN	−.25	−.24	0.1	.899
SOQ Competitiveness	EVI	−.10	−.15	0.6	.546
	IVI	−.13	−.13	0	>.999
	KIN	−.16	−.11	0.6	.545
SOQ Win Orientation	EVI	−.04	−.08	0.5	.634
	IVI	−.11	−.08	0.4	.719
	KIN	−.13	−.09	0.5	.631
SOQ Goal Orientation	EVI	−.16	−.20	0.5	.623
	IVI	−.19	−.22	0.4	.710
	KIN	−.14	−.11	0.4	.717
APSA	EVI	.27	.25	0.3	.799
	IVI	.24	.19	0.6	.534
	KIN	.10	.12	0.2	.810
RRQ	EVI	.10	.15	0.6	.546
	IVI	.18	.19	0.1	.902
	KIN	.05	.04	0.1	.905
FFMQ	EVI	−.30	−.28	0.3	.795
	IVI	−.25	−.28	0.4	.702
	KIN	−.17	−.16	0.1	.903
BFI Neuroticism	EVI	.13	.16	0.4	.716
	IVI	.15	.15	0	>.999
	KIN	.09	.06	0.4	.720
BFI Extraversion	EVI	−.17	−.16	0.1	.903
	IVI	−.12	−.17	0.6	.544
	KIN	−.15	−.08	0.8	.399
BFI Openness	EVI	−.08	−.03	0.6	.551
	IVI	−.08	−.10	0.2	.811
	KIN	−.09	−.13	0.5	.631
BFI Agreeableness	EVI	−.08	−.08	0	>.999
	IVI	−.04	−.10	0.7	.474
	KIN	−.07	−.14	0.8	.400
BFI Conscientiousness	EVI	−.15	−.12	0.4	.717
	IVI	−.10	−.11	0.1	.904
	KIN	−.08	−.06	0.2	.811

Note. The Vividness of Movement Imagery Questionnaire (VMIQ-2) factors were external visual imagery (EVI), internal visual imagery (IVI), and kinesthetic imagery (KIN). Pearson correlations of these factors were calculated with movement imagery (MIQ-3; Williams et al., 2012), self-efficacy (SE; Hinz et al., 2006), sport imagery ability (SIAQ; Simonsmeier & Hannemann, 2017), physical self-concept (PSC; Stiller et al., 2004), sport orientation (SOQ; Elbe, 2004), concentration (APSA; Bankstahl & Görtelmeyer, 2013), rumination (RRQ; Trapnell & Campbell, 1999), mindfulness (FFMQ; Michalak et al., 2016), and personality traits (BFI; Rammstedt & John, 2007).
For p = .05 and N = 285, the critical r = .12.

p = .056, η_p^2 = .02; version: $F < 1$; activity × version: $F < 1$; version × modality: $F < 1$; activity × modality: $F_{2,564}$ = 1.4, p = .231, η_p^2 = .01).

Discussion: Study 2

In a within-subject design, the print and computer-based versions of the VMIQ-2 (Dahm et al., 2019) revealed similar item characteristics. Model fits of the confirmatory factor analysis, interfactor correlations, and correlations with other questionnaires did not significantly differ between the print and computer-based versions. Furthermore, correlations of single factors between the print and computer-based versions were strong (r > . 5; Cohen, 1992). On average, participants created more vivid images in IVI than EVI. This was observed in both the print and computer-based versions. Hence, independent from the measurement method, the same construct – vividness of action imagery – was assessed.

Interestingly, active participants created more vivid images in KIN than in EVI, which was not observed in moderate and nonactive participants. Although this was only significant in the print version, visual inspection of the data indicates similar tendencies in the computer-based version (see Figure 2). Possibly, regular action execution involving kinesthetic information facilitated the representation of kinesthetic information during imagery in active subjects. This could not be compensated for in nonactive and moderately active subjects. Visual information of an action, in contrast, may also be acquired by action observation in nonactive subjects (Flanagan & Johansson, 2003).

In line with the expectations, participants needed less time in the computer-based version than in the print version to read the introduction of the VMIQ-2 (Kretzschmar et al.,

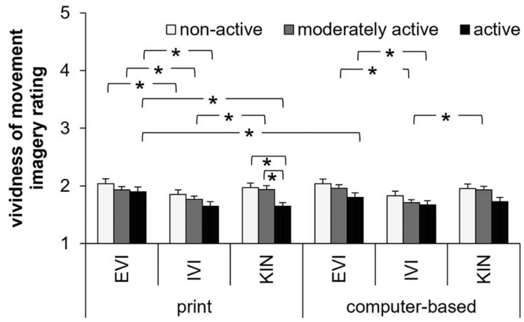

Figure 2. Means and standard errors of vividness of action imagery in the print and the computer-based versions of the VMIQ-2 separately for nonactive, moderately active, and active participant groups. VMIQ-2 = Vividness of Movement Imagery Questionnaire.

2013). However, the repetitions of the comprehension question indicated that comprehension of the introduction did not significantly differ between the print and computer-based versions. As for the reading durations, participants needed less time to fill out the computer-based version than to fill out the print version. The data indicate that the medium does influence the working speed. One might argue that filling out the questionnaire is easier by mouse clicks than by writing crosses with a pencil. Furthermore, passing over to new pages was automatized in the computer version but may have cost time in the print version. Nevertheless, reading time was also shorter in the computer-based version than in the print version. Possibly, participants feel time pressure while working with a computer but feel more relaxed when working with paper and pencil. It has also been argued that displays facilitate reading due to higher text discriminability (by color contrasts) than in the print medium (Kretzschmar et al., 2013). Alternatively, although the experimenters were instructed to prepare in advance and to hand out the pages as fast as possible, handing out the forthcoming pages to the participants may have taken time. A previous study showed that without interactions with the experimenter, reading time did not differ between print reading and screen reading (Oborne & Holton, 1988).

General Discussion

It was investigated whether the medium (print vs. computer-based) influences the assessment of vividness of action imagery. Strong correlations on the same factors were observed between the print and computer-based versions. Confirmatory factor analyses revealed measurement invariance between the print and computer-based versions in a between-subject design (Study 1) and a within-subject design (Study 2). Furthermore, test–retest reliability, interfactor correlations, and correlations with other constructs did not significantly differ between the print and computer-based versions. This goes in line with measurement invariance in paper–pencil and computer-based personality assessments of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Chuah et al., 2006).

In Study 1, the correlation between the visual imagery factors was higher in the computer-based version than in the print version. It was assumed that in the computer-based version, participants did not read the introduction to the questionnaire as thoroughly as participants who read the print version (Mazzeo & Harvey, 1988). Possibly, they were not aware of the three factors (EVI, IVI, and KIN) when filling out the first page with EVI which may have influenced particularly the EVI ratings which were at the first page of the questionnaire. Indeed, in Study 2, shorter reading durations were observed in the computer-based version than in the print version. However, this did not significantly hamper reading comprehension. Most importantly, in Study 2, the correlation between the visual imagery factors did not significantly differ between the print and computer-based versions. This may be due to the comprehension question after the introduction which enforced 20% of the participants to re-read the introduction, making sure that all participants had understood the three factors. The present data indicate that particularly in computer-based assessments comprehension questions are a useful tool to assure that participants read and understand the instructions.

In Study 2, the ratings indicated more vivid visual images from an internal perspective than from an external perspective. It is possible that more people favor the internal perspective over the external perspective. It appears plausible that subjects are more familiar with IVI than with EVI as the internal perspective is the natural perspective during action execution.

Negative correlations between the VMIQ-2 and several other questionnaires were observed. The negative correlations stem from the inverse rating scale of the VMIQ-2 with a rating score of 1 indicating very vivid movement images and a rating score of 5 indicating almost no imagery. Basically, vividness of action imagery (Dahm et al., 2019) correlated strongly ($.46 < r < .64$) with ease of action imagery generation (Williams et al., 2012). One may have expected even stronger correlations because the imagery dimensions vividness and ease are assumed to be very similar (Cumming & Eaves, 2018; Dahm, 2020), and both questionnaires involve the same factors. However, the MIQ-3 involves different movements than the VMIQ-2 and action imagery is assumed to be movement specific (Dahm et al., 2019). Furthermore, the foregoing execution of each movement might influence the following imagination of the movement in the MIQ-3.

Moderate correlations were observed between vividness of action imagery (Dahm et al., 2019) and sport imagery ability (Simonsmeier & Hannemann, 2017), self-efficacy (Hinz et al., 2006), concentration (Bankstahl & Görtelmeyer, 2013), and mindfulness (Michalak et al., 2016). These correlations may either indicate that the constructs overlap to some degree or that the self-ratings were similarly influenced by social-desirability biases (Dahm, 2020; Gabbard & Lee, 2014). Weak correlations were observed between vividness of action imagery (Dahm et al., 2019) and competitiveness (Elbe, 2004), win orientation (Elbe, 2004), goal orientation (Elbe, 2004), rumination (Trapnell & Campbell, 1999), and the Big Five personality traits Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Rammstedt & John, 2007). These constructs appear not to overlap with vividness of action imagery and are possibly less susceptible to social-desirability biases.

Unfortunately, information and communication technology literacy (Wang et al., 2007, 2008) was not assessed in the present studies. It remains unresolved whether differences between the print and computer-based assessments would appear in a subpopulation with less information and communication technology literacy. Particularly older subjects may have lower information and communication technology literacy than younger subjects, which may further influence the print and computer-based assessments.

In conclusion, the computer-based version of the VMIQ-2 revealed acceptable model fits which replicated the three-dimensional structure of the print version. Furthermore, convergent and discriminant validity was observed and correlations with other questionnaires did not significantly differ between the computer-based version and the print version. The computer-based version is therefore a valid measurement to assess vividness of action imagery involving EVI, IVI, and KIN.

References

Abraham, A., Gose, R., Schindler, R., Nelson, B. H., & Hackney, M. E. (2019). Dynamic neuro-cognitive imagery (DNITM) improves developpé performance, kinematics, and mental imagery ability in university-level dance students. *Frontiers in Psychology*, *10*, 1–17. https://doi.org/10.3389/fpsyg.2019.00382

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Waveland Press.

Anuar, N., Cumming, J., & Williams, S. E. (2016). Effects of applying the PETTLEP model on vividness and ease of imaging movement. *Journal of Applied Sport Psychology*, *28*(2), 185–198. https://doi.org/10.1080/10413200.2015.1099122

Bankstahl, U., & Görtelmeyer, R. (2013). Measuring subjective complaints of attention and performance failures - Development and psychometric validation in tinnitus of the self-assessment scale APSA. *Health and Quality of Life Outcomes*, *11*(1), 86. https://doi.org/10.1186/1477-7525-11-86

Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, *28*(3), 282–299. https://doi.org/10.1080/08886504.1996.10782166

Callow, N., Roberts, R., Hardy, L., Jiang, D., & Edwards, M. G. (2013). Performance improvements from imagery: Evidence that internal visual imagery is superior to external visual imagery for slalom performance. *Frontiers in Human Neuroscience*, *7*(697), 1–10. https://doi.org/10.3389/fnhum.2013.00697

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, *40*(4), 359–376. https://doi.org/10.1016/j.jrp.2005.01.006

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cumming, J., & Eaves, D. L. (2018). The nature, measurement, and development of imagery ability. *Imagination, Cognition and Personality*, *37*(4), 375–393. https://doi.org/10.1177/0276236617752439

Dahm, S. F. (2020). On the assessment of motor imagery ability: A research commentary. *Imagination, Cognition and Personality*, *39*(4), 397–408. https://doi.org/10.1177/0276236619836091

Dahm, S. F. (2022). *computer-based version of the Vividness of Movement Imagery Questionnaire* [Data set]. https://osf.io/xtmd9/

Dahm, S. F., Bart, V. K. E., Pithan, J. M., & Rieger, M. (2019). Deutsche Übersetzung und Validierung des VMIQ-2 zur Erfassung der Lebhaftigkeit von Handlungsvorstellungen [German translation and validation of the VMIQ-2 for the assessment of vividness of movement imagery]. *Zeitschrift für Sportpsychologie*, *26*(4), 151–158. https://doi.org/10.1026/1612-5010/a000273

Dana, A., & Gozalzadeh, E. (2017). Internal and external imagery effects on tennis skills among novices. *Perceptual and Motor Skills*, *124*(5), 1022–1043. https://doi.org/10.1177/0031512517719611

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, *79*(4), 481–492. https://doi.org/10.1037/0021-9010.79.4.481

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, *6*(42), 1–23. https://doi.org/10.3390/jintelligence6030042

Elbe, A.-M. (2004). Testgütekriterien der deutschen Version des Sport Orientation Questionnaires [Test control criteria for the German version of the Sport Orientation Questionnaire]. *Spectrum der Sportwissenschaften*, *16*(1), 96–107.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Gabbard, C. P., & Lee, J. (2014). A comparison of movement imagery ability self-report and imagery use in a motor task: A preliminary investigation. *Journal of Imagery Research in Sport and Physical Activity*, *9*(1), 61–66. https://doi.org/10.1515/jirspa-2014-0006

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence*, *62*, 138–147. https://doi.org/10.1016/j.intell.2017.04.001

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347–360. https://doi.org/10.1111/j.1745-3984.1984.tb01039.x

Groth-Mamat, G., & Schumaker, J. (1989). Computer-based psychological testing: Issues and guidelines. *American Journal of Orthopsychiatry*, *59*(2), 257–263. https://doi.org/10.1111/j.1939-0025.1989.tb01657.x

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (Eds.), (2019). *Multivariate data analysis*. Cengage.

Hinz, A., Schumacher, J., Albani, C., Schmid, G., & Brähler, E. (2006). Bevölkerungsrepräsentative Normierung der Skala zur Allgemeinen Selbstwirksamkeitserwartung [Standardization of the general self-efficacy scale in the German population]. *Diagnostica*, *52*(1), 26–32. https://doi.org/10.1026/0012-1924.52.1.26

Isaac, A. R. (1992). Mental practice—Does it work in the field? *The Sport Psychologist*, *6*(2), 192–198. https://doi.org/10.1123/tsp.6.2.192

Isaac, A. R., Marks, D. F., & Russell, D. G. (1986). An instrument for assessing imagery of movement: The vividness of Movement Imagery Questionnaire (VMIQ). *Journal of Mental Imagery*, *10*(4), 23–30.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202. https://doi.org/10.1017/S0140525X00034026

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, *12*(3), 247–252.

Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, *80*, 17–22. https://doi.org/10.1016/j.jrp.2019.03.007

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: Concurrent EEG-eyetracking evidence from the reading of books and digital media. *PLoS One*, *8*(2), e56178. https://doi.org/10.1371/journal.pone.0056178

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268. https://doi.org/10.2307/2532051

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mazzeo, J., & Harvey, A. L. (1988). The equivalence of scores from automated and conventional educational and psychological tests. *ETS Research Report Series*, *1988*(1), 1–27. https://doi.org/10.1002/j.2330-8516.1988.tb00277.x

Michalak, J., Zarbock, G., Drews, M., Otto, D., Mertens, D., Ströhle, G., Schwinger, M., Dahme, B., & Heidenreich, T. (2016). Erfassung von Achtsamkeit mit der deutschen Version des Five Facet Mindfulness Questionnaires (FFMQ-D) [Assessment of mindfulness with the German version of the Five Facet Mindfulness Questionnaires (FFMQ-D)]. *Zeitschrift für Gesundheitspsychologie*, *24*(1), 1–12. https://doi.org/10.1026/0943-8149/a000149

Moran, A. P., Campbell, M., Holmes, P., & MacIntyre, T. (2012). Mental imagery, action observation and skill learning. In N. J. Hodges & A. M. Williams (Eds.), *Skill acquisition in sport: Research, theory and practice* (2nd ed., pp. 94–111). Routledge (Taylor and Francis).

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

Oborne, D. J., & Holton, D. (1988). Reading from screen versus paper: There is no difference. *International Journal of Man-Machine Studies*, *28*(1), 1–9. https://doi.org/10.1016/s0020-7373(88)80049-x

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

O'Shea, H., & Moran, A. (2019). Are fast complex movements unimaginable? Pupillometric studies of motor imagery in expert piano playing. *Journal of Motor Behavior*, *51*(4), 371–384. https://doi.org/10.1080/00222895.2018.1485010

Pithan, J. M., & Dahm, S. F. (2015). Fragebögen und Testmethoden der Bewegungsvorstellung [Questionnaires and testing methods of motor imagery]. *Zeitschrift für Sportpsychologie*, *22*(3), 112–124. https://doi.org/10.1026/1612-5010/a000146

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Development Core Team. (2019). *R: A language and environment for statistical computing*. http://www.r-project.org/

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

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*(2), 200–221. https://doi.org/10.1123/jsep.30.2.200

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02

Ruiz, M. C., Watt, A. P., Samper, N. P., & Morris, T. (2018). 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

Schroeders, U. (2009). Testing for equivalence of test data across media. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment* (pp. 164–170). Office for Official Publications of the European Communities.

Simonsmeier, B. A., & Hannemann, A. (2017). Die deutsche Übersetzung des SIQ und SIAQ zur Erfassung von Vorstellungsgebrauch und Vorstellungskompetenz im Sport [German translation of the SIQ and SIAQ to assess imagery use and imagery ability in sports]. *Zeitschrift für Sportpsychologie*, *24*(3), 100–110. https://doi.org/10.1026/1612-5010/a000204

Stiller, J., Würth, S., & Alfermann, D. (2004). Die Messung des physischen Selbstkonzepts (PSK) [The measurement of physical self-concept (PSK) - the development of the PSK-Scales for children, adolescents, and young adults]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *25*(4), 239–257. https://doi.org/10.1024/0170-1789.25.4.239

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, *76*(2), 284–304. https://doi.org/10.1037/0022-3514.76.2.284

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, *67*(2), 219–238. https://doi.org/10.1177/0013164406288166

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments. *Educational and Psychological Measurement*, *68*(1), 5–24. https://doi.org/10.1177/0013164407305592

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*(3–4), 149–180.

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 and Exercise Psychology*, *34*(5), 621–646. https://doi.org/10.1123/jsep.34.5.621

Ziegler, M. (2020). Psychological test adaptation and development – How papers are structured and why. *Psychological Test Adaptation and Development*, *0*(0), 1–9. https://doi.org/10.1027/2698-1866/a000002

History
Received July 17, 2020
Revision received December 16, 2021
Accepted December 20, 2021
Published online March 4, 2022

Section: Sports & Health Psychology

Acknowledgments
Many thanks to the students involved in data collection (Fidan Arziman, Verena Christl, Nadja Hasslinger, Christoph Hoppichler, Marianna Maticevic, Katharina Neumayer, Sara Schirmer, Meike Steuer, Chantal Wari, and Belinda Yilmaz) and thanks to the participants for spending their valuable time. Furthermore, I want to thank the reviewers for their helpful comments to improve the manuscript.

Open Data
The author confirms that the data supporting the findings of this study are available within the article and its supplementary materials. Determination of the sample size, all data exclusions, all manipulations, and all measures in the study are reported in the manuscript. Furthermore, the assessment tool for the computer-based version of the VMIQ-2 is provided in the following link https://osf.io/xtmd9/ (Dahm, 2022).

ORCID
Stephan F. Dahm
https://orcid.org/0000-0002-6001-6640

Stephan F. Dahm
Institute of Psychology
Department for Psychology and Medical Sciences
UMIT – Private University for Health Sciences
Medical Informatics and Technology
Eduard Wallnöfer-Zentrum 1
6060 Hall in Tyrol
Austria
stephan.dahm@umit.at