

Developing the 'Imagery during exercise questionnaire-Gym Version' and assessing its content validity.

KATERINA DASKALAKI¹, MANOS STEFANAKIS², MICHALIS EFSTATHIOU³, ANASTASIA BENEKA⁴, PARASKEVI MALLIOU⁵

^{1,4,5}Department of Physical Education and Sport Science, School of Physical Education and Sport Science, Democritus University of Thrace, GREECE

^{2,3,5}Department of Health Sciences, School of Life and Health Sciences, University of Nicosia, NICOSIA CYPRUS

Published online: May 31, 2024

Accepted for publication : May 15, 2024

DOI:10.7752/jpes.2024.05134

Abstract

Mental imagery (MI) is a psychological technique traditionally applied outside exercise contexts. However, recent studies indicate its potential benefits during physical activity, particularly for enhancing positive emotions and self-regulation of exercise behavior. While existing questionnaires capture aspects of MI regarding the exercise, none are specifically designed to evaluate MI during training. Therefore, it is necessary to create a new questionnaire. Establishing its content validity (CV) is the initial step in this process. Despite its importance for subsequent validity assessments and the reliability of the measurement instrument, CV is often overlooked or superficially addressed, possibly owing to a lack of comprehensive guidelines. In response to these needs, the aim of this study was to develop the "Imagery during Exercise Questionnaire-Gym Version" (IdEQ-GV) to assess the functions of MI during exercise and to measure its CV. This study also provides a step-by-step presentation of the content validation process. The development of the IdEQ-GV involved instrument design and judgmental evidence from academic experts and lay experts (i.e., potential research subjects). The first version of the questionnaire, consisting of 34 items across 6 dimensions, was assessed by three content experts who evaluated its CV in terms of relevance, essentiality, clarity, and comprehensiveness and provided feedback. The second version, comprising 28 items across 5 dimensions, was re-evaluated by content experts and discussed in cognitive interviews with 13 gym members, who also provided feedback and scores regarding comprehensibility and comprehensiveness. After making the necessary modifications, the third version of the IdEQ-GV, containing 24 items, was evaluated by a new sample of 20 exercisers for comprehensibility. The methods used to quantify CV included the content validity index (CVI), the scale-level CVI (S-CVI), which comprises the average CVI (S-CVI/Ave) and universal agreement (UA) among experts (S-CVI/UA), and the content validity ratio (CVR). The IdEQ-GV was found to have an excellent CV regarding relevance, with an S-CVI/UA of 1.00. For clarity, the S-CVI/Ave was 0.95. For essentiality, the average CVR of the scale was 0.83. For comprehensiveness, according to the content experts' scores, the S-CVI/Ave was 0.93, while the first sample of exercisers gave it a perfect score. Regarding comprehensibility, the final sample of exercisers rated the S-CVI/Ave at 0.93. Through this analytical process, it has been established that the IdEQ-GV has excellent content validity and is now ready for widespread distribution to establish additional psychometric properties.

Keywords: Mental imagery, exercise imagery, relevance, comprehensiveness, comprehensibility, content validity index

Introduction

Mental Imagery (MI) is a psychological skill that first became known mainly due to its use by high-level athletes (Hall, Mark, Paivio & Hausenblas, 1998; Hall, Rodgers, & Barr, 1990) but later began to be viewed as an effective psychological technique in the fields of dance (Heiland & Rovetti, 2009), exercise (Hausenblas, Hall, Munroe, & Rodgers, 1999) and rehabilitation (Driediger, Hall, & Callow, 2006; Beneka et al., 2013). In its traditional form, it is used as a form of mental training outside of practice with the aim of achieving cognitive (e.g., improved technique) and motivational outcomes (e.g., increased motivation and positive emotions) (Cumming & Ramsey, 2009; Munroe-Chandler & Gammage, 2005; Watt, Klep, & Morris, 2018). When its use is investigated, it is usually examined what images people form (types of MI) and/or for what reasons/why they form these images (functions of MI) (Cumming & Ramsey, 2009). Depending on its type and/or function, MI can be: Cognitive Specific (CS; MI referring to technique improvement), Cognitive General (CG; related to strategies and tactics), Motivational Specific (MS; concerns goals and the behaviors that can lead to their achievement), Motivational General-Arousal (MG-A; related to the control of levels of arousal, e.g. reducing anxiety, increasing excitement) and Motivation General-Mastery (MG-M; refers to in managing

difficult or demanding situations, e.g. maintaining concentration and positivity) (Cumming & Ramsey, 2009; Daskalaki, Beneka, Mamoukari, & Malliou, 2021; Munroe-Chandler & Gammage, 2005).

In the exercise field, there are three questionnaires that examine exercise imagery (EI), that is what kind of images do the individuals form about exercise in moments unrelated to practice [i.e., the “Exercise Imagery Questionnaire-Aerobic Version” (EIQ-AV) (Hausenblas et al., 1999), the “Exercise Imagery Inventory (EII)” (Giacobbi, Hausenblas, & Penfield, 2005) and “EII-revised (EII-R)” (Giacobbi, Tuccitto, Buman, & Munroe-Chandler, 2010). Nevertheless, the EIQ-AV and EII-R do not assess all types/functions of MI, while all three of them focus on the use of MI as it is traditionally understood, i.e., during non-exercise moments. The value of investigating EI lies in the fact that its use has been found to predict leisure-time exercise behavior and self-efficacy beliefs. As it has been supported, by investigating the “relationship between different imagery types and variables typically associated with exercise participation”, researchers will be able to design interventions for improving exercise adherence (Cumming, 2008).

A very interesting finding in recent years is that MI is also used by gym members during training. In addition to reasons across all fields (e.g. improving technique, achieving goals, managing arousal levels and emotions), exercisers were found to form mental images during exercise for additional purposes, e.g., for distracting themselves from fatigue and boredom and for making the lesson more interesting (Daskalaki et al., 2021). This demonstrates that during exercise, MI serves additional motivational functions that are not seen in non-training conditions, confirming that the existing questionnaires cannot capture all functions MI may serve. Based on the above, it becomes evident that for the evaluation of the use of MI by exercisers during training, the existing questionnaires are not suitable, so a new one should be created. The necessity of creating a questionnaire of this type derives also from the fact that in the absence of such a measurement tool, relationships between the use of MI during training and exercise behavior cannot be established, leaving the field of MI during exercise majorly unexplored. As it has been reported, EI (i.e., imagery outside of practice) is considered one strategy that aids individuals to self-regulate their exercise behavior (Cumming, 2008). Considering that the maintenance of physical activity levels is crucial for reducing the risk of various health problems (Themistokleous, Hadjisavvas, Efsthathiou, & Stefanakis, 2024) and noting the high dropout rates in sport and fitness centers (Gonçalves & Diniz, 2015; Zarotis, Athanailidis, Arvanitidou, & Mourtziotis, 2017), the implementation of more interventions to retain the already existing members is imperative. Since MI during exercise may enhance positive feelings (Daskalaki et al., 2021) it is important to examine for what reasons gym members form mental images during their work-outs, in order to gain information on their needs, tendencies and preferences, which can then be exploited to incorporate effectively this technique during exercise classes, as it has been already suggested in the recent literature (Daskalaki, Papadimitriou, & Malliou, 2024).

When creating a new questionnaire, and before it is given to the population, it is necessary to obtain evidence of its validity and reliability (Zamanzadeh et al., 2015). Regarding validity, it has different forms (Ouzouni & Nakakis, 2011), and the most common of them are content validity, criterion validity, and construct validity (Zamanzadeh et al., 2015). Content validity (CV) refers to the assessment of the content of a questionnaire and is the extent to which its items cover the different dimensions of the concept or, in other words, of the variable being measured. CV, essentially, involves the evaluation of a new instrument regarding the elements that are chosen to be included in it to measure a concept (Galani, 2013). Content experts (i.e., academics of professionals specialized in the relative scientific area) and/or lay experts (i.e., members from the target population) are asked to evaluate specific elements of the questionnaire by giving a score and to also provide feedback. Various elements of the instrument can be subject to evaluation, according to the purpose of the researchers, including: the “relevance” or “representativeness” (Rodrigues, Adachi, Beattie, & MacDermid, 2017), “clarity”, “simplicity”, “ambiguity” (Yaghmaie, 2003), “sensitivity” (Guillot-Valdés, Guillén-Riquelme, & Buela-Casal, 2022) of the items, as well as the “sufficiency”/ “comprehensiveness” of the dimensions (Fernández-Gómez et al., 2021; Zamanzadeh et al., 2015). The selected variables are often scored in a 4-point scale (Yaghmaie, 2003), except for essentiality which is usually scored in a 3-point scale (Taherdoost, 2016). This process is extremely valuable since it makes it possible to identify the strengths and weaknesses of an instrument (Fernández-Gómez et al., 2021), to ensure that key items have not been omitted (Artino, La Rochelle, Dezee, & Gehlbach, 2014), and to make the necessary modifications, e.g. add, eliminate, modify and/or clarify items (Fernández-Gómez et al., 2021).

As it is reported, CV is not only a prerequisite for the other forms of validity (Korakakis et al., 2021) but without it, it is impossible to establish the reliability of a measurement instrument. Although it is a necessary and critical process and should receive the highest priority when creating a new questionnaire, it is usually studied superficially and sporadically, perhaps because guidelines on its process are rarely provided in a single source (Zamanzadeh et al., 2015). As it has also been reported, in the absence of a CV study, researchers would distribute an untested instrument to the population (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003), so the elaboration of CV must precede the distribution of a measurement tool to the public (Ouzouni & Nakakis, 2011). As it is also outlined by Yaghmaie (2003), “statements such as: ‘the validity of the instrument had been tested by experts’ or ‘CV was determined through a review of literature or panel experts’ are unacceptable and invalid statements”. Content validation includes specific processes that must be followed and presented, concrete

methods for providing objective scores of its various parameters, while quantification of CV is absolutely required (Dongare et al., 2019). Haynes, Richard, & Kubany (1995) emphasized that “inferences from assessment instruments with unsatisfactory CV will be suspect, even when other indices of validity are satisfactory” (Yusoff, 2019).

Responding to these needs, the aim of the present research was to develop the “Imagery during Exercise Questionnaire-Gym Version” (IdEQ-GV) for the investigation of the functions of MI during exercise and to assess its CV, providing at the same time useful information on the CV process through a step-by-step explanation of the procedure.

Materials & Methods

The development of IdEQ-GV involved instrument design and judgmental evidence (Rodrigues et al., 2017) from both academic experts and lay experts (i.e., members of the target population), as usually is recommended (Fernández-Gómez et al., 2021).

Instrument Development (1st version of IdEQ-GV)

The purpose of IdEQ-GV is to measure how often and for what reasons exercisers are using MI during working out. The domains and items of IdEQ-GV were based: A) On an extensive literature review in Google Scholar, Pub Med, Research Gate, etc. Articles referring to MI were identified, while, also, previous questionnaires examining MI in exercisers [i.e., EIQ-AV (Hausenblas, et al., 1999), EII (Giacobbi et al., 2005), “EII-R” (Giacobbi et al., 2010) but also in dancers (e.g. Nordin & Cumming, 2006) were carefully studied, following the recommendations of the literature to study other questionnaires in and around the area of interest (Kishore, Jaswal, Kulkarni, & De, 2021). Some items from these sources were used and adapted accordingly for the purpose of IdEQ-GV (Artino et al., 2014). Also, theoretical approaches regarding the possible use of MI within the exercise fields were taken into account (Munroe-Chandler & Gammage, 2005), as well as qualitative data from previous research (e.g. Giacobbi et al., 2003). The generation of the items was also majorly based on the qualitative findings of Daskalaki et al. (2021) which recorded reasons why exercisers are using MI during exercise, and on the answers of 61 gym practitioners of an unpublished yet research (conducted by members of the present research group) who -in one of its questions- described in their own words for what reasons they create mental images while exercising. Using the qualitative findings of these studies, the suggestion to consult members from the target population when developing items for new questionnaires (Cortini, Galanti, & Fantinelli, 2019) was fulfilled.

Three members of the research group, with extensive experience in exercise in theoretical and applied contexts, created a pool of items and arranged all items systematically to form a preliminary draft of the questionnaire (Kishore et al., 2021). The 1st version of IdEQ-GV included 34 items and 6 dimensions that corresponded to the 5 dimensions found in the literature (that is CS, CG, MS, MG-A, MG-M MI) and a dimension named “other reasons” that included additional items that were not falling in the 5 previous categories. All items would require a response to the question “Considering the last 3 months, please answer how frequently you use your imagination when exercising for the reasons listed below”, on a 7-point Likert Scale, from 1 (never) to 7 (always). According to the researchers, IdEQ-GV demonstrated face validity, i.e., it appeared to measure what it was supposed to measure (Kishore et al., 2021).

Consulting content experts

Consulting “content experts” (Zamanzadeh et al., 2015), i.e., individuals who are knowledgeable of the subject matter by means of their academic or professional background, is an important step for assessing the quality of an instrument and establishing CV. Through a procedure known as “evaluation through expert judgment”, experts are asked to assess different aspects of the instrument through a numerical rating scale and to also provide their reasoned opinions and suggestions (Fernández-Gómez et al., 2021). Taking into account that the variables to be scored, the terminology used in relation to them, but also the way they are scored, differ from survey to survey, the research group, following the recommendations of Daskalaki, Koutra, Karagiannakidou, Papadimitriou, & Malliou (2023), decided which parameters were important to be evaluated for this specific questionnaire, and wrote down how they define these parameters, in order to provide specific instructions to the experts.

According to the standardized procedures proposed, a cover letter inviting content experts to participate in the content validation process was prepared (Kishore et al., 2021) which together with a copy of the questionnaire and a CV assessment form (response form) was sent via email to 6 experts. In the cover letter, the purpose and a brief description of the new instrument were provided, as well as how the dimensions and items were generated (Almanasreh, Moles, & Chen, 2019; Zamanzadeh et al., 2015). The experts were also informed that they should evaluate the questionnaire by using the assessment form to provide a score regarding: a) the relevance of each item, i.e., whether each sentence fits logically and is related to the domain where it has been included, with possible scores: 1=not relevant, 2=somewhat relevant, 3=quite relevant, and 4=very relevant (Rodrigues et al., 2017), b) the clarity of each item (i.e., if it is easily understood, if it is clearly formulated) with possible scores: 1=Not clear, 2=Needs major revisions to be clear, 3=Needs minor revisions to be clear, 4=Clear (Kishore et al., 2021; Zamanzadeh et al., 2015), c) the essentiality of each item (i.e., if the item is necessary)

with possible scores: 1=not essential, 2=useful, but not essential, 3=essential (Taherdoost, 2016), d) the comprehensiveness of each dimension (i.e., if the items included in each dimension are sufficient or if an important item may be missing) with possible scores per dimension: 1=The items are insufficient, 2=Many items must be added in order to fully assess the dimension, 3=A few items must be added in order to fully assess the dimension, 4=The items are sufficient to measure the dimension (Fernández-Gómez et al., 2021). Content experts were also requested to write their comments in case they didn't provide the highest score and to also make suggestions for improvements (Almanasreh, et al., 2019). A sample of the content validation form is provided in Table 1.

Table 1. Sample of the content validation form

	Relevance				Clarity				Essentiality		
Q1	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Q2	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Q3	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Q4	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Comprehensiveness of dimension: 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>											
Please write your comments here for anything not rated with the highest score.											

3 out of the 6 content experts responded positively and provided feedback (from the rest 3, one informed that had a busy schedule, while the others did not respond). However, 3 experts are an acceptable number according to the literature (Almanasreh, et al., 2019; Lynn, 1986). The experts who reviewed the questionnaire were academics, specialized in exercise training and exercise psychology and with experience in psychometry. Therefore, they had the cognitive sufficiency to evaluate the tool, to counter-propose methods for its improvement (Iseris, 2016), and to guarantee “a higher scientific quality of the study” (Fernández-Gómez et al., 2021). It was decided that expert judgment would be requested in as many rounds as necessary until the acceptable cut-off score for relevance and comprehensiveness would be achieved, before administrating the instrument to lay experts. From that point on, their scores and remarks regarding clarity and essentiality would be used to further inform the procedure, as it will be detailed later.

Scoring of the CV of the questionnaire following content experts' judgment

The quantification of the CV following content experts' judgment was done as follows: For quantifying the relevance and the clarity of the items and of the scale (Zamanzadeh et al., 2015), the CV index (CVI) measure was used. CVI is the most widely reported approach for measuring CV in instrument development (Rodrigues et al., 2017) and it also provides useful information for each item, which can be exploited for making decisions regarding each item, e.g. to retain, modify or delete it (Zamanzadeh et al., 2015). This index was originally designed for the evaluation of the relevance or representativeness of items (Lynn, 1986) but nowadays is also used for rating other parameters, e.g. clarity (Kishore et al., 2021). CVI can be calculated for each item separately, i.e., at an item level (I-CVI) as well as for the whole questionnaire, i.e., at a scale level (S-CVI) (Zamanzadeh et al., 2015).

I-CVI is calculated as the number of experts that gave a score of 3 or 4 to each item, divided by the total number of experts (Almanasreh, et al., 2019; Zamanzadeh et al., 2015). It expresses the proportion of agreement on the relevance of each item (or clarity or representativeness etc.), which is between zero and one. The acceptable cut-off score (i.e. the indicated score for maintaining an item) of I-CVI depends on the number of experts and for three to five experts is 1 (Lynn, 1986; Yusoff, 2019).

S-CVI is defined as “the proportion of total items judged content valid” or “the proportion of items on an instrument that achieved a rating of 3 or 4 by the content experts” (Zamanzadeh et al., 2015) and refers to the whole scale (Almanasreh, et al., 2019). There are two methods to calculate S-CVI: The first one is the Average CVI (S-CVI/Ave) which is calculated by taking the sum of the I-CVIs (I-CVI1+I-CVI2+I-CVI3+ ...+I-CVIN) divided by the total number of items (more ways of calculating it can be found on Polit & Beck, 2016). The second one is the Universal Agreement (UA) among experts (S-CVI/ UA). S-CVI/UA is calculated by adding all items with I-CVI equal to 1 (i.e. that achieved a score of 3 or 4 by all experts) divided by the total number of items (Almanasreh, et al., 2019; Rodrigues et al., 2017). The minimum acceptable standard for S-CVI/ Ave is 0.90 and for S-CVI/UA 0.80 (Polit & Beck, 2006).

For quantifying the essentiality of the items, the CV Ratio (CVR) (Lawshe, 1975) was used. CVR varies between 1 and -1 and a higher score indicates greater agreement among the experts (Galanis, 2013; Rodrigues et al., 2017). The formula for its calculation is $CVR = (Ne - N/2) / (N/2)$, where “Ne” is the number of experts indicating an item as “essential” and “N” is the total number of experts. The cut-off score of CVR depends on the number of experts, which according to the Lawshe table when the experts are five must be 0.99 (Lawshe, 1975). The CVR was generated for each item and the whole scale (average CVR) (Rodrigues et al., 2017).

For quantifying the comprehensiveness of each dimension, the number of experts that gave a score of 3 or 4 to the comprehensiveness of each dimension was divided by the total number of experts (Zamanzadeh et al., 2015), and of the whole questionnaire, the S-CVI/ Ave and the S-CVI/UA were used as previously described.

Consulting lay experts (members of the target population)

Consulting lay experts (i.e., subjects from the target population) is an important stage of the content validation process (Zamanzadeh et al., 2015) since it ensures that the population for whom the questionnaire is being developed is represented (Rubio et al., 2003). As it becomes obvious from the literature, the management of this procedure depends on its purpose. In the present research, the goal was not to develop from scratch the items of the questionnaire (case in which interviews or focus groups would be strongly suggested even before consulting content experts) (Artino et al., 2014; Cortini et al., 2019). The goal was to investigate the comprehensiveness of the questionnaire (and to possibly add missing items which were not indicated by the content experts), its comprehensibility, and to further refine it by addressing the short-comings which arose from content experts' judgment regarding clarity and essentiality of items. Adapting the procedure accordingly, two phases were followed with lay experts (i.e., gym exercisers): the first included semi-structured interviews, and the second the completion of the questionnaire digitally.

Semi-structured interviews with lay experts

The purpose of this stage was to determine: 1) if lay experts understood the instructions given and the items *as it was intended by the researchers* (Rodrigues et al., 2017), and in a similar fashion (Artino et al., 2014), 2) if they felt that there were overlapping items (which would provide further information on the essentiality of items), c) if they felt that the dimensions were comprehensive, d) if they were satisfied with the score options of the items, e) which was their score regarding the comprehensibility and the comprehensiveness of the questionnaire.

Semi-structured interviews were conducted during 4 meetings with gym members in 2 different gyms of Athens. A researcher with experience in semi-structured interviews, holding a Master's degree in Communication, Media, and Culture and also a PhD in Exercise Science, was guiding the procedure. A convenience sample was used and the meetings consisted of one triad, two dyads, and one mini-group of 6 participants, combinations that permitted to take advantage of the different dynamics of diverse kinds of groups (Cortini et al., 2019).

13 participants (women=9 and men=2) participated in the interviews, from 32 to 69 years old (53.54 ± 8.9), who were exercising from 3 months to 45 years (median 3 years) and during the last three months for 4.91 ± 2.2 hours weekly. All meetings started with an explanation of the procedure. The participants provided consent to be audio-taped, and they were informed that the audio material would be used only by the researcher in order to keep notes at a later time. Then, the participants were given the 2nd version of IDEQ-GV, were asked to read the instructions (which also included the definition of MI) and (after the second meeting) to complete the questionnaires. The time needed for completing them ranged from 6 to 10 minutes (median 7 minutes). Upon completion, the first verbal probe used in all meetings was "Did you anticipate any problems in any given moment in understanding the questions? Was everything clear?" Since a basic goal of this process was to control if the items were clear and the language familiar to the participants, comprehension/interpretation verbal probes were mostly used (e.g. what does the term "to relax" mean to you?) (Artino et al., 2014). These probes were primarily employed for items that didn't achieve the highest possible score from the experts, and for items that, following the open-ended questions of the researcher, raised a discussion among the participants. Interviewees were also asked to indicate if any important issue had been omitted (Terwee et al., 2018).

The participants scored each item regarding its comprehensibility on a 4-point scale with possible scores: 1=the sentence is not comprehensible at all, 2=the sentence is a little comprehensible, 3=the sentence is quite comprehensible, 4= the sentence is completely comprehensible, and also the comprehensiveness of each dimension (with the same scoring options previously described for content experts). After the meetings, a written record of the useful and relevant comments that came up during the discussions was created (Vogt, King, & King, 2004). Taking into account the comments of the gym members, individual items were modified accordingly (Artino et al., 2014).

Online survey with lay experts

In this phase, the purpose was to assess the comprehensibility of the 3rd version of IDEQ-GV through an online survey. A digital version of the questionnaire was created in Google Forms, and a different sample of gym members (convenience sample) was asked to complete it and then provide a score regarding the comprehensibility of the instructions, responding to the question: "On a scale of 1 (not at all comprehensible) to 10 (completely comprehensible), how comprehensible were the questionnaire instructions?". They were also required to score the items of the questionnaire regarding comprehensibility with the same score options as lay experts of the previous phase (i.e., on a 4-point scale).

20 participants (women=14, men=6) from 42 to 67 years old (53.06 ± 7.99 years) who were exercising from 2 months to 35 years (10.25 ± 10.61 years), and during the last three months for 4.5 ± 2.09 hours weekly, filled the questionnaire and provided a score.

Results

Final CV results regarding relevance, clarity, essentiality, and comprehensiveness, as these were scored by content experts, and regarding comprehensiveness according to the score of lay experts who participated in the interviews, are from the 2nd version of IdEQ-GV (Table 2). Regarding comprehensibility (understandability), the scores that refer to its 2nd version (i.e., of the exercisers participating in the interviews) are mentioned for reasons of clarifying that comprehensibility was also judged during cognitive interviews, as it is suggested (Terwee et al., 2018). The final scores regarding comprehensibility are from the (final) 3rd version (i.e., of the exercisers who completed the questionnaire during the online survey). In Table 2, the items of the 2nd version of IdEQ-GV, as well as the I-CVI of all items and dimensions are disposed, and these are the final scores that are used for the establishment of the CV of IdEQ-GV. Below, apart from mentioning the scores, it is also analyzed how the results of each phase impacted on the content of IdEQ-GV.

Feedback and CV according to content experts

At first, content experts provided feedback regarding the 1st version of IdEQ-GV. After the first round of judgment, the dimension “other reasons” was removed because one expert indicated that this domain was very generic, and one item of this category was removed to “Technique”, from which another item was also removed. In total, 6 items were deleted, a new item was added, and 4 items were slightly rephrased.

The 2nd version of IdEQ-GV included 28 items and 5 dimensions which were: 1. Technique (4 items), 2. Strategies/Tactics (5 items), 3. Energy/Relaxation (5 items), 4. Concentration/Self-Confidence/Positivism in difficult situations (7 items), 5. Goals (appearance, health, improvement of physical condition) (7 items). The revised questionnaire was sent again to three experts for a second round (one of them was different from the 1st round) which provided the following scores:

Regarding relevance, all items (Q1-Q28) received the highest score (3 or 4) from all experts. Therefore, regarding relevance, the I-CVI for all items, the S-CVI/Ave and S-CVI/UA was 1.00, achieving the highest possible score and indicating that IdEQ-GV has an excellent CV regarding relevance (Polit & Beck, 2006).

For clarity, all items received a score of 4, except for items Q11, Q22, Q23 and Q28 which were scored with 2 by one expert. Consequently, 24 items had an I-CVI of 1.00 and 4 items an I-CVI of 0.66. Therefore, the S-CVI/Ave was 0.95, and the S-CVI/UA was 0.85, which is considered excellent at a scale level (Polit & Beck, 2006). For comprehensiveness, all dimensions received a score of 3 or 4 from all experts, with the exception of “Energy-Relaxation” which received a score of 2 from one expert. Consequently, the I-CVI was 1.00 for all dimensions, except for “Energy-Relaxation” which had a score of 0.66. Therefore, S-CVI/Ave was 0.93, and S-CVI/UA was 0.8, scores that are considered excellent at a scale level (Polit & Beck, 2006).

For essentiality, all items received a score of 3, except for items Q5, Q6, Q7, Q8, Q9, Q24 and Q27 which were scored with 2 by one expert. Therefore, 21 items had a CVR of 1.00, seven had a score of 0.33, and the average CVR of the scale was 0.83. Although nonessential items can be eliminated, in this case, they were not (Rodrigues et al., 2017), since all items were found to have an excellent I-CVI with regards to relevance, but also in case they were removed the whole dimension “Strategy-Tactics” would be eliminated, leaving the domain of CG MI unrepresented. Also, although the “Energy-Relaxation” dimension was scored with 2 from one expert, the explanation given for this score was that item Q11 was not complete, without any other feedback for including additional items. For these reasons, the 2nd version of IdEQ-GV was considered by the research group as ready to be judged by lay experts. The I-CVIs for all variables, according to the judgment of content experts, are presented in Table 2.

Feedback and CV according to lay experts during semi-structured interviews

According to the score given by gym members, for comprehensiveness, all dimensions had a CVI of 1.00. Regarding comprehensibility all items had an I-CVI of 1, except for Q6, Q11, and Q18 which received a score of 2 in one instance each, thus having an I-CVI of 0.90, which for this number of raters (i.e., 13) is considered excellent (Rodrigues et al., 2017). Also, all exercisers were absolutely satisfied with the instructions of the questionnaire and with the scoring options of the items (i.e., 1-7).

Although gym members evaluated IdEQ-GV in the most favorable way, using the open-ended questions specific matters arose that permitted a further refinement of the instrument. Specifically, Q2, although comprehensible, it was also criticized by some respondents to be more appropriate as an item for a professional instructor, not for simple exercisers, while others argued that the item was appropriate because independently of the outcome, the goal of an exerciser would be to approach perfection. The solution proposed by lay experts was to rephrase Q2 into “improve/perfect my technique” sustaining that in this way would apply to all of them.

Regarding Q4, when participants were asked to rephrase the question, there was a tendency to try to rephrase the term “flow” and not the whole question, which indicated that the item was not comprehensible the way the researchers intended. Some participants proposed to add more examples in the parenthesis, so Q4 was rephrased as follows: “To achieve a particular characteristic in my movement (i.e., make the movement more dynamic, more graceful, have a better flow, etc.).

For Q11, some exercisers indicated that they thought that “to relax” referred to psychological/mental relaxation, others to physical, and others to holistic relaxation. The item was intended to mean “psychological/mental relaxation” so it was established that this item was ambiguous and not complete (as it was

also indicated by one of the experts). Given that the participants argued that other items (*i.e.*, Q12, Q13) were enough to embrace the idea of psychological relaxation, Q11 was eliminated, even though according to content experts' score, it was essential.

Q18 gave also rise to discussion because some exercisers were not sure if the word "exercise" was referring to one single exercise or to a whole exercise session (in Greek language and in this specific phrasal context both interpretations can be given when using the word "exercise", which was also our intention). Exercisers agreed that it would be better to use both terms in other similar questions, *i.e.* during "exercise/exercise session", and their advice was followed.

When asking participants if there were items that were overlapping, a consideration was made about Q25 and Q28, arguing that Q28 included by way of its examples in the parenthesis the meaning of Q25. Therefore, Q25, although comprehensible, it was eliminated. Further, through interpretation verbal probes asking the meaning of Q24 and Q27, (*e.g.* what does the term "muscular body" mean to you?), it was established that although the participants perceived that these two questions did not share the same meaning, in many cases interpreted differently these items than was intended and not in a similar way in between them, meeting the considerations of one academic expert who, however, perceived the two items as being the same. The research group modified Q24 as follows: "my muscles will be more well-defined/ more visible", and eliminated Q27.

Regarding the instructions and the definition of MI in the questionnaire, they were judged as clear by the majority of the respondents, however, through individual questions of some participants, it was judged useful to add two new phrases to the definition. In the 4th meeting, it was observed that the same problems were discussed and the same solutions were proposed, so a meeting with a new group would not add new important information to the achieved results (Cortini *et al.*, 2019; Rodrigues *et al.*, 2017). The researchers also eliminated Q7 because it was decided that it is not applicable to all types of training held in a gym (*e.g.* programs without music or using not rhythmical music). After these modifications, the 3rd version of IdEQ-GV had 5 dimensions and 24 items.

CV regarding comprehensibility according to lay experts through an online survey

The 3rd version of IdEQ-GV demonstrated a high degree of comprehensibility. The instructions and the definition provided received a score of "10" from 12 participants, "9" from 5 participants, and "8" from 3 participants, indicating that it was highly comprehensible attaining a score of 9 or 10 from 85% of the participants, (Fernández-Gómez *et al.*, 2021), and an average score of 9.45.

15 items attained a score of 3 or 4 from 19 participants and an I-CVI of 0.95. 9 items attained a score of 3 or 4 from 18 participants and an I-CVI of 0.90 (Table 2). These scores establish excellent I-CVI for this number of raters (Lynn, 1986) regarding comprehensibility. S-CVI/Ave was 0.93 which is considered excellent (Polit & Beck, 2006).

Table 2. Scores regarding Relevance, Clarity, Essentiality, Comprehensiveness and Comprehensibility of IdEQ-GV

Items of the 2 nd version of IdEQ-GV	Scores from content experts (2 nd version of IdEQ-GV)			Scores from lay experts (3 rd version of IdEQ-GV)
	Relevance	Clarity	Essentiality	Comprehensibility
	I-CVI	I-CVI	(CVR)	(I-CVI)
TECHNIQUE				
Q1. To correct mistakes in the technical execution of the exercise.	1	1	1	0.95
Q2. To perfect my technique.	1	1	1	0.95
Q3. To activate the right muscles.	1	1	1	0.9
Q4. To achieve a particular characteristic in my movement (<i>e.g.</i> better flow).	1	1	1	0.9 (in its modified version)
Comprehensiveness of dimension: According to content experts: 1.00 According to lay experts: 1.00				
STRATEGY/TACTICS				
Q5. To make it easier to remember the sequence of steps/exercises.	1	1	0.33	0.95
Q6. To be able to maintain the desired pace of exercise.	1	1	0.33	0.9
Q7. To synchronize with the beat of the music.	1	1	0.33	ELIMINATED
Q8. To better handle the space around me.	1	1	0.33	Q7=0.9
Q9. To better manage my strength to complete the workout (<i>e.g.</i> vary the intensity or speed of execution).	1	1	0.33	Q8=0.95
Comprehensiveness of dimension: According to content experts: 1.00 According to lay experts: 1.00				

ENERGY/RELAXATION				
Q10. To increase my energy/get energized.	1	1	1	Q9=0.95
Q11. To relax.	1	0.66	1	ELIMINATED
Q12. To feel calm/peaceful	1	1	1	Q10=0.95
Q13. To reduce stress/anxiety.	1	1	1	Q11=0.9
Q14. To forget my problems.	1	1	1	Q12=0.9
Comprehensiveness of dimension				
According to content experts: 0.66				
According to lay experts: 1.00				
FOCUS/CONFIDENCE/POSITIVISM IN DIFFICULT SITUATIONS				
Q15. To focus more on exercise.	1	1	1	Q13=0.95
Q16. To be able to continue exercising for longer/ not to stop.	1	1	1	Q14=0.95
Q17. To be distracted from tiredness and/or muscle pain caused by exercise.	1	1	1	Q15=0.95
Q18. To alleviate the monotony of exercise.	1	1	1	Q16=0.9
Q19. To make exercise more enjoyable.	1	1	1	Q17=0.9
Q20. To make time go faster.	1	1	1	Q18=0.9
Q21. To increase my confidence that I will be able to complete the exercise/workout.	1	1	1	Q19=0.95
Comprehensiveness of dimension				
According to content experts: 1.00				
According to lay experts: 1.00				
GOALS (APPEARANCE, HEALTH, IMPROVEMENT OF PHYSICAL CONDITION)				
During the exercise I imagine that through the exercise:				
Q22. I am doing good to my health.	1	0.66	1	Q20=0.95
Q23. I will get a more beautiful body.	1	0.66	1	Q21=0.95
Q24. I will get a more muscular body.	1	1	0.33	Q22=0.95 (in its modified version)
Q25. I will get a stronger/durable body.	1	1	1	ELIMINATED
Q26. I will lose weight/maintain my weight at desired levels.	1	1	1	Q23=0.95
Q27. I will get a more "well defined" body.	1	1	0.33	ELIMINATED
Q28. My physical condition will improve (eg. endurance, strength, flexibility, etc.).	1	0.66	1	Q24=0.95
Comprehensiveness of dimension				
According to content experts: 1.00				
According to lay experts: 1.00				

Discussion

The present research developed a new questionnaire named IdEQ-GV and provided evidence of its CV, detailing all the processes followed. After two rounds of consultation with content experts, the scale attained an excellent CV regarding relevance. All dimensions were judged as completely comprehensive by all content experts, attaining a score of 1.00, except for the dimension "Energy-Relaxation" which received a score of 0.66, therefore S-CVI/Ave was 0.93, and the S-CVI/UA was 0.8, which are acceptable (Polit & Beck, 2006). Additionally, all dimensions of the scale were considered absolutely comprehensive by lay experts (who participated in interviews) who scored comprehensiveness with the highest possible score, corresponding to a CVI of 1.00 at a dimension level and at a scale level. An excellent CV regarding comprehensibility was also achieved, as this was established by 20 gym members.

The whole process of content validation proved the necessity of following this procedure. The 1st version of IdEQ-GV had 6 dimensions and 33 items, the 2nd had 5 dimensions and 28 items, and the 3rd (final) version had 5 dimensions and 24 items. The content validation process from content experts revealed some weaknesses that were corrected, so the questionnaire distributed to lay experts was already substantially improved. The comments of 13 gym members through interviews held in 4 different meetings made it possible to identify problematic items, to remove some of them and to modify others accordingly, a procedure that ensured the high comprehensibility of the tool when it was evaluated by a different sample of 20 gym subscribers. The standard set for concluding the procedure of obtaining feedback from content experts was that the scale achieved an excellent CV regarding relevance and comprehensiveness. Indeed, in the 2nd version of IdEQ-GV, all items were judged as relevant. Regarding comprehensiveness, 4 of the 5 dimensions achieved the highest score. The dimension "Energy/Relaxation" received a lower score, however, no additional proposals

were made regarding new items, while the S-CVI/Ave was 0.93, and the S-CVI/UA was 0.8, which both are acceptable (Polit & Beck, 2006). Although also clarity and essentiality were evaluated, it was decided to not necessarily eliminate items that did not achieve the acceptable cut-off score regarding these two parameters, if researchers had concrete reasons for maintaining them.

In more detail, regarding essentiality, the main reason for not eliminating Q5, Q6, Q7, Q8 and Q9 was that these items were the only items that represented the CG function of MI. In their absence, a whole category of MI would be missing from the questionnaire, replicating what has already happened with the questionnaires EIQ-AV (Hausenblas et al., 1999) and EII (Giacobbi et al., 2005) (which assess the use of MI during non-exercise moments). Both EIQ-AV and EII have been criticized for the fact of not assessing all functions that EI may serve (Giacobbi, et al., 2010; Munroe-Chandler & Gammage, 2005). Also, there were no other proposals regarding alternative items for this category, which, anyway, from all content experts was scored as comprehensive. As for the remaining items that didn't achieve the minimal requirement of perfect agreement at an item level (i.e., Q24 and Q27), it was decided to pay special attention to them during the upcoming interviews. Indeed, Q24 and Q27 were proven to be ambiguous to lay experts, so the first one was rephrased and the second one was eliminated. As it is also discussed by Rodrigues et al (2017), the calculation of CV based on essentiality is less common, while, in their research, they retained items that had a lower CVR than that indicated as acceptable. Also, according to COSMIN methodology (Terwee et al., 2018) which, nevertheless, targets mainly the assessment of CV of patient-reported outcome measures (PROMs), relevance, comprehensiveness and comprehensibility are the three most important aspects of CV, while they underline that comprehensibility should be evaluated by lay experts, not by professionals.

Regarding clarity, as it becomes obvious by Table 2, item Q11 which had received an I-CVI of 0.66 from the content experts was removed, even though it had received an I-CVI of 1.00 from the lay experts who participated in the interviews: interviews made it possible to establish that exercisers did not interpret the item in the same way, although, each one of them had his/her own interpretation of the item and considered it to be comprehensible. As for Q22, Q23 and Q28, according to the research group were clear enough to describe what they intended, which was also established through the cognitive interviews, so it was decided that these items would not be eliminated. As it was previously done by other researchers (Fernández-Gómez et al., 2021; Guillot-Valdes et al., 2022), we controlled the comprehensibility of IdEQ-GV employing a survey. Fernández-Gómez et al. (2021) immediately after the content experts' judgment distributed the validated questionnaire in a sample of 102 women, and the degree of understandability of each item was measured by means of what percentage of them evaluated positively the items in a dichotomous yes/no response (a score equal to or greater than 85% indicated high understandability, from 80% to 85% medium understandability, and less than 80% low understandability. Guillot-Valdes et al. (2022) used a scale of 0 to 10 regarding understanding of the items from their sample and concluded to an average comprehension of 9.82 out of 10. As they remarked, all items achieved a score of 9 or higher, indicating that they were easily understandable, so it was not necessary to modify or delete any of them. Dantas Silva, Brito, & Pereira (2022) applied a 4-point scale regarding comprehensibility and demanded a total agreement for retaining an item. In the present research, it could be sustained that a rigorous procedure with regards to lay experts was followed since we employed both cognitive interviews and also an online survey. According to COSMIN methodology (Terwee et al., 2018), the comprehensiveness of the PROM or subscale should be assessed using the final set of PROM items, and if items were added or eliminated during psychometric or pilot testing, it is necessary to conduct a new pilot test. As they report, using textual data alone—a survey—is not sufficient to evaluate comprehensibility. In the case of the present research, we conducted the final evaluation regarding comprehensibility, after consulting lay experts via interviews. The comprehensiveness of the scale according to lay experts had been decided in that stage, and the items that were eliminated were mainly overlapping with others, so there was no need to reassess the comprehensiveness in a different stage. The modifications that were made were acceptable to members of the target population during the cognitive interviews, and the new items were totally comprehensible to them. The new (quantitative) survey was mainly undertaken in order to cross-check the questionnaire with a different sample of exercisers, and mainly because two phrases were added to the definition and two items were collapsed into a rephrased one, which was though discussed with the interviewees. The fact that 20 gym members gave such a high score regarding comprehensibility, revealed that the questionnaire is ready for further distribution.

The results of the present research confirm the necessity of content validation studies, a procedure that is majorly ignored in favor of more “prestigious” methodological designs, such as those that target construct validity. The various findings of the different stages of this research show the value of this process and highlight the need to give higher priority to the publication of detailed content validity studies which would stand as autonomous articles in their own right. Through this analytical process, a new scale was developed that will permit the measurement of MI of exercisers during working out. Although the scale was developed for exercisers who work out in organized exercise and athletic facilities, it could be probably be applicable to exercisers who are exercising without supervision in non-competitive contexts (e.g. recreational runners, cyclists, swimmers, etc.). Through IdEQ-GV, new findings are expected to be generated regarding the relationships between the use of MI during training and exercise behavior.

Conclusions

Although content validation processes are majorly overlooked and neglected, the various findings of the different stages of this research confirm the value of this procedure. IdEQ-GV was found to have an excellent CV regarding relevance, comprehensiveness and comprehensibility, which are three fundamental aspects of CV. In this way, it is now established that IdEQ-GV is ready as an instrument to be widely distributed for the establishment of additional psychometric properties. This questionnaire is the first to measure MI during exercise and it is expected to benefit majorly the field of exercise psychology as well as the applied field of exercise instruction. However, for this result to happen, various stages were followed, employing both content and lay experts, while different methods were used including written feedback, cognitive interviews and an online survey. In all phases, specific scores were obtained that permitted objectivity in the evaluation of the different parameters of the questionnaire's CV, and the procedure was concluded when acceptable cut-off scores were achieved. The step-by-step explanation of the CV procedure here presented may be used by researchers who deal with psychometrics, facilitating the decision to first establish the CV of a new questionnaire before proceeding to the assessment of its other psychometric properties and thus ensuring a higher quality of new measurement instruments.

Disclosure statement

No potential conflicts of interest were reported by the authors.

References

- Almanasreh, E., Moles, R., & Chen T. F. (2019). Evaluation of methods used for estimating content validity, *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Artino, A. R., Jr., La Rochelle, J.S., Dezee, K. J. & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87, *Medical Teacher*, 36(6), 463-474. DOI: 10.3109/0142159X.2014.889814
- Beneka, A., Malliou, P., Gioftsidou, A., Kofotolis, N., Rokka, S., Mavromoustakos, S., & Godolias, G. (2013). Effects of instructional and motivational self-talk on balance performance in knee injured. *The European Journal of Physiotherapy*, 15(2), 56-63. <https://doi.org/10.3109/21679169.2013.776109>
- Cortini, M., Galanti, T. & Fantinelli, S. (2019). Focus Group Discussion: how many Participants in a Group? *Encyclopaideia – Journal of Phenomenology and Education*, 23(54), 29-43. <https://doi.org/10.6092/issn.1825-8670/9603>
- Cumming, J., & Ramsey, R. (2009). Imagery interventions in sport. In S. D Mellalieu & S. Hanton (Eds.), *Advances in applied sport psychology: A review* (pp. 5–36). London: Routledge.
- Cumming, J. (2008). Investigating the Relationship between Exercise Imagery, Leisure-Time Exercise Behavior, and Self-Efficacy. *Journal of Applied Sport Psychology*, 20(2), 184-198. DOI: [10.1080/10413200701810570](https://doi.org/10.1080/10413200701810570)
- Dantas Silva, W.A. & Brito, T. R. S. & Pereira, C. R. (2022). COVID-19 anxiety scale (CAS): Development and psychometric properties. *Current Psychology*. 39. DOI: 10.1007/s12144-020-01195-0.
- Daskalaki, K., Papadimitriou, K., & Malliou, P-V. (2024). Use of mental imagery techniques to transform your fitness classes into a fantastic experience. *ACSM's Health and Fitness Journal*, 28(2), 37-41. DOI: 10.1249/FIT.0000000000000947
- Daskalaki, K., Koutra, Ch., Karagiannakidou, I., Papadimitriou, K., & Malliou P. (2023). Which Parameters Are Assessed When Establishing the Content Validity of New Questionnaires. 6th Conference in Functional Rehabilitation of Injuries in Athletes and Trainees, Komotini, Greece, 10-11 November 2023.
- Daskalaki, K., Beneka, A., Mamoukari P., & Malliou, P. (2021). Construction of the Questionnaire "Exercise Imagery Pilot Questionnaire-Gym Version" for the Investigation of the Imagery of Gym Members during Exercise- a Pilot Research. *European Psychomotricity Journal*. 14, 15-25. https://psychomotor.gr/wp-content/uploads/2022/01/3_DASKALAKI_15_25.pdf
- Dongare, P.A., Bhaskar, S.B., Harsoor, S.S., Kalaivani, M., Garg, R., Sudheesh, K., & Goneppanavar U. (2019). Development and validation of a questionnaire for a survey on perioperative fasting practices in India. *Indian Journal of Anaesthesia*, 63(5), 394-399. DOI: 10.4103/ija.IJA_118_19.
- Driediger, M., Hall, C., & Callow, N. (2006). Imagery use by injured athletes: A qualitative analysis. *Journal of Sports Sciences*, 24(3), 261–271. DOI: [10.1080/02640410500128221](https://doi.org/10.1080/02640410500128221)
- Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M.A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content Validation through Expert Judgement of an Instrument on the Nutritional Knowledge, Beliefs, and Habits of Pregnant Women. *Nutrients*, 12(4):1136. <https://doi.org/10.3390/nu12041136>
- Galanis, P. (2013). Validity and reliability of questionnaires in epidemiological studies, *Archives of Hellenic Medicine*, 30(1), 97-110.
- Giacobbi, P. R. J., Tuccitto, D. E., Buman, M. P., & Munroe-Chandler, K. (2010). A measurement and conceptual investigation of exercise imagery establishing construct validity. *Research Quarterly for Exercise and Sport*, 81(4), 485–493. DOI: [10.1080/02701367.2010.10599710](https://doi.org/10.1080/02701367.2010.10599710)

- Giacobbi, P. R., Hausenblas, H. A., & Penfield, R. D. (2005). Further Refinements in the Measurement of Exercise Imagery: The Exercise Imagery Inventory. *Measurement in Physical Education and Exercise Science*, 9(4), 251–266. <http://doi.org/10.1207/s15327841mpee0904>
- Giacobbi, P. R., Hausenblas, H. A., Fallon, E. A., & Hall, C. A. (2003). Even more about exercise imagery: A grounded theory of exercise imagery. *Journal of Applied Sport Psychology*, 15(2), 160–175. DOI: [10.1080/10413200305391](https://doi.org/10.1080/10413200305391)
- Gonçalves, C. & Diniz, A. (2015). Analysis of member retention in fitness through satisfaction, attributes perception, expectations and well-being. *Revista Portuguesa de Marketing*, 38(34), 65-76. <https://cif2d.fade.up.pt/files/rpm34celinaana.pdf>
- Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2022). Content Validity through Expert Judgment for the Depression Clinical Evaluation Test. *International Journal of Clinical and Health Psychology*, 22(2):100292. DOI: [10.1016/j.ijchp.2022.100292](https://doi.org/10.1016/j.ijchp.2022.100292).
- Hall, C.R., Mack, D., Paivio, A., & Hausenblas, H. (1998). Imagery use by athletes: Development of the Sport Imagery Questionnaire. *International Journal of Sport Psychology*, 29, 73–89.
- Hall, C. R., Rodgers, W. M., & Barr, K. A. (1990). The Use of Imagery by Athletes in Selected Sports. *The Sport Psychologist*, 4(1), 1–10. DOI: <https://doi.org/10.1123/tsp.4.1.1>
- Hausenblas, H. A., Hall, C. R., Munroe, K. J., & Rodgers, W. M. (1999). Exercise imagery: Its nature and measurement. *Journal of Applied Sport Psychology*, 11(2), 171–180. DOI: [10.1080/10413209908404198](https://doi.org/10.1080/10413209908404198)
- Haynes, S.N., Richard, D., & Kubany, E.S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7(3):238. <https://doi.org/10.1037/1040-3590.7.3.238>
- Heiland, T., & Rovetti, R. (2013). Examining effects of Franklin Method metaphorical and anatomical mental images on college dancers' jumping height. *Research in Dance Education*, 14(2), 141–161. DOI: [10.1080/14647893.2012.712105](https://doi.org/10.1080/14647893.2012.712105)
- Iseris, G. (2016). Statistical methods of checking the validity and reliability of questionnaires. The case of CiGreece). *International Journal of Language, Translation and Intercultural Communication*, 5, 175–189. <https://doi.org/10.12681/ijltic.10665>
- Kishore, K., Jaswal, V., Kulkarni, V., & De, D. (2021). Practical Guidelines to Develop and Evaluate a Questionnaire. *Indian Dermatology Online Journal*, 12(2), 266-275. DOI: [10.4103/idoj.IDOJ_674_20](https://doi.org/10.4103/idoj.IDOJ_674_20).
- Korakakis, V., Kotsifaki, A., Stefanakis, M., Sotiralis, Y., Whiteley, R., & Thorborg, K. (2021). Evaluating lower limb tendinopathy with Victorian Institute of Sport Assessment (VISA) questionnaires: a systematic review shows very-low-quality evidence for their content and structural validity-part I. *Knee Surgery, Sports Traumatology, Arthroscopy*, 29(9), 2749–2764. DOI: [10.1007/s00167-021-06598-5](https://doi.org/10.1007/s00167-021-06598-5)
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, 28(4), 563-575. DOI: [10.1111/j.1744-6570.1975.tb01393.x](https://doi.org/10.1111/j.1744-6570.1975.tb01393.x)
- Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6):382-5.
- Munroe-Chandler, K. J., & Gammage, K. L. (2005). Now See This: A New Vision of Exercise Imagery. *Exercise and Sport Sciences Reviews*, 33(4), 201–205. DOI: [10.1097/00003677-200510000-00009](https://doi.org/10.1097/00003677-200510000-00009).
- Nordin, S. M., & Cumming, J. (2006). Measuring the Content of Dancers' Images Development of the Dance Imagery Questionnaire (DIQ). *Journal of Dance Medicine and Science*, 10(3&4), 85–98.
- Ouzouni, Ch. & Nakakis, K. (2011). Validity and Reliability of Measurement Instruments in Quantitative Studies, *Nosileftiki*, 50(2), 231-239.
- Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Research in Nursing & Health*, 29, 489-497. <https://doi.org/10.1002/nur.20147>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94–104. <http://www.jstor.org/stable/42659521>
- Rodrigues, I.B., Adachi, J.D., Beattie, K.A., & MacDermid, J.C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders*, 18(1): 540. DOI: [10.1186/s12891-017-1914-5](https://doi.org/10.1186/s12891-017-1914-5)
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. <http://dx.doi.org/10.2139/ssrn.3205040>
- Terwee, C.B., Prinsen, C.A.C., Chiarotto, A., Westerman, M.J., Patrick, D.L., Alonso, J., Bouter, L.M., de Vet, H.C.W., & Mokkink, L.B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, 27(5), 1159-1170. DOI: [10.1007/s11136-018-1829-0](https://doi.org/10.1007/s11136-018-1829-0).
- Themistocleous, I. C., Hadjisavvas, S., Efstathiou, M. A., & Stefanakis, M. (2024). The Application of High Intensity Interval Training in Atrial Fibrillation: A Literature Review. *International Journal of Cerebrovascular Disease and Stroke*, 7, 175. DOI: <https://doi.org/10.29011/2688-8734.100175>

- Vogt, D. & King, D. & King, L. (2004). Focus Groups in Psychological Assessment: Enhancing Content Validity by Consulting Members of the Target Population. *Psychological assessment*, 16(3), 231-43. DOI: 10.1037/1040-3590.16.3.231.
- Watt, A., Klep, D., & Morris, T. (2018). Psychometric analysis of the sport imagery ability measure. *Journal of Physical Education and Sport*, 18(1), 138-148. DOI:10.7752/jpes.2018.01018
- Yaghmaie, F. (2003). Content Validity and its Estimation. *Journal of Medical Education*, 3(1): e105015. <https://doi.org/10.22037/jme.v3i1.870>.
- Yusoff, M.S.B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*. 11(2):49–54. <https://doi.org/10.21315/eimj2019.11.2.6>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar A.R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, 4(2):165-78. DOI: [10.15171/jcs.2015.017](https://doi.org/10.15171/jcs.2015.017)
- Zarotis, G., Athanailidis, I., Arvanitidou, V., & Mourtzios C. (2017). Age-specific reasons for dropping out of the fitness-sport. *Journal of Physical Education and Sport*, 17(2): 916. DOI: 10.7752/jpes.2017.02140