Research Article
Artigo de Pesquisa
Artículo de Investigación

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

Brazilian version of "self-efficacy for home exercise programs scale" for the elderly population

Versão brasileira da "self-efficacy for home exercise programs scale" para a população idosa

Versión brasileña de "self-efficacy for home exercise programs scale" para la población de ancianos

Desiree Rodrigues da Veiga[I]; *Kelsey Jean Picha*[II]; *Flávio Rebustini*[I]

[I]*Universidade de São Paulo, São Paulo, Brazil;* [II]*University of Kentucky, Kentychy, United States of America*

ABSTRACT

Objective: to seek evidence of validity of the Brazilian version Self-efficacy for home exercise programs scale" for the elderly population. **Method:** evidence of content, response process and internal structure were analyzed. Two expert panels were used (equivalences and content) as it is an assembly. Three scale formats were analyzed: original, 5 points and 7 points. For the response process and internal structure, 174 older people participated. The Internal Structure was tested by exploratory factor analysis. **Results:** Seven of the 12 items underwent wording adjustments, and one item was added following expert guidance. The response process indicated that the version with 5 points was pointed out by the elderly as the best to be answered. In the internal structure, the one-dimensional model with the 5-point scale obtained the best indicators. **Conclusion:** it is recommended to adopt the instrument in a unidimensional way and use the 5-point scale.

Descriptors: Psychometrics; Validation Study; Aged; Exercise; Self-Efficacy.

RESUMO

Objetivo: buscar evidências de validade da versão brasileira *Self-efficacy for home exercise programs scale"* para a população idosa. **Método**: foram analisadas as evidências de conteúdo, processo de resposta e estrutura interna. Foram utilizados dois painéis de especialistas (equivalências e conteúdo) por ser um *assembly.* Três formatos da escala foram analisados: original, 5 pontos e 7 pontos. Para o processo de resposta e estrutura interna participaram 174 idosos. A Estrutura interna foi testada pela análise fatorial exploratória. **Resultados**: Sete dos 12 itens passaram por ajuste de redação, e um item foi adicionado por orientação dos especialistas. O processo de resposta indicou que a versão com 5 pontos foi a apontada pelos idosos como a melhor para ser respondida. Na estrutura interna, o modelo unidimensional com a escala de 5 pontos obteve com os melhores indicadores. **Conclusão**: recomenda-se a adoção do instrumento de forma unidimensional e o uso da escala de 5 pontos.

Descritores: Psicometria; Estudo de Validação; Idoso; Exercício Físico; Autoeficácia.

RESUMEN

Objetivo: buscar evidencias de validez de la versión brasileña de "Self-efficacy for home exercise programs scale" (escala de autoeficacia de programas de ejercicios en el hogar) para la población anciana. Método: se analizaron evidencias de contenido, proceso de respuesta y estructura interna. Se utilizaron dos paneles de expertos (equivalencias y contenido) por tratarse de una asamblea. Se analizaron tres formatos de escala: original, 5 puntos y 7 puntos. 174 ancianos participaron en el proceso de respuesta y la estructura interna. Se probó la Estructura Interna mediante análisis factorial exploratorio. **Resultados**: Siete de los 12 ítems se sometieron a ajustes de redacción y se agregó un ítem siguiendo la opinión de expertos. El proceso de respuesta indicó que los ancianos indicaron la versión con 5 puntos como siendo la mejor para ser respondida. En la estructura interna, el modelo unidimensional con escala de 5 puntos obtuvo los mejores indicadores. **Conclusión**: se recomienda adoptar el instrumento de forma unidimensional y utilizar la escala de 5 puntos.

Descriptores: Psicometría; Estudio de Validación; Anciano; Ejercicio Físico; Autoeficacia.

INTRODUCTION

In general terms, quality of life can be considered a very broad and complex concept that encompasses various issues such as: physical and psychological health, independence level, social relations, cognitive efficacy, productivity, continuity of family and occupational roles, and adequate environmental conditions[1]. It is a consensus that the periodic practice of physical activity is a determining factor for health promotion, in addition to assisting in the treatment and prevention of various chronic diseases. Their high incidence and prevalence in several populations is already considered as a public health issue[2].

For the aged population, worldwide initiatives for health promotion emerged through the practice of physical activities that were developed by the World Health Organization (WHO), in 1996, when the institution developed the protocol called "*The Heiderberg Guidelines for Physical Promotion Among Older Persons*"[3], which contributes diverse scientific evidence of the benefits of physical activity in the biological, psychological and social aspects, in addition to the importance of public policies in encouraging physical activity. The document also deals with the motivating factors and barriers for physical activity in this population group[3].

Corresponding author: Desiree Rodrigues da Veiga. E-mail: desireeveiga@usp.br.
Editor in chief: Cristiane Helena Gallasch; Associate Editor: Magda Guimarães de Araujo Faria

Received: Mar 5th 2023 – Accepted: Jun 7th 2023

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

In Brazil the Brazilian Society of Sports Medicine, in partnership with the Brazilian Society of Geriatrics and Gerontology, prepared the document called "Physical Activity and Older Adults' Health – Official stance", whose purpose would be to expand the recommendation of physical activity practices by health professionals who deal with aged individuals[4]. After 20 years, in 2021, the Ministry of Health launched the "Guide to Physical Activity for the Brazilian Population", where specific recommendations can be found on physical activities for older adults[5].

It is noted that physical inactivity in aged people is not always due to an individual's lack of intentional practice but that, in many cases, it can be the result of sociocultural impositions which have nothing to do with functional disability. Older adults are culturally attributed dependence and immobility, which make them significantly reduce their physical activity levels.

With the recent COVID-19 pandemic, which started at the end of 2019 and arrived in Brazil at the beginning of 2020, we had a change in the performance of physical exercises in mid-March: the gyms were closed due to acknowledging the public calamity status. This movement forced many people to abandon their physical exercises or to practice them at their homes with the online guidance of Physical Education professionals.

In a recent scoping review study, articles that contained recommendations about the practice of physical activities and exercises for Brazilians during the COVID-19 pandemic were analyzed[6]. A total of 22 papers were found where recommendations for the practice of physical activity and exercise rely on the use of digital resources for their guidance and include the recommendation for their performance in home environments and. only when possible, in outdoor settings, respecting safety and social distancing measures.

The *Self-Efficacy for Home Exercise Programs Scale* (SEHEPS) is the first scale targeted at assessing self-efficacy for home-based exercise programs. The authors' idea[7] would be using it for the medical prescription of physical exercises with rehabilitation purposes. The scale was reviewed by a group of specialists comprised by sports trainers, physiotherapists and a specialist in self-efficacy, in order to provide diverse evidence regarding its face and content validity. They decided to modify the answer scale from 11 to 7 points (0-6) to reduce the discrimination levels in the scale; the scale anchors vary from 0 (Not confident) to 6 (Very confident). As the labels were combined and the verbal anchors are only used in three points of the scale: The "Not confident" label was placed between 0 and 1, the "A little confident" label is between 2, 3 and 4; and "Very confident", between 5 and 6. The fact that we do not have verbal labels at all points of the scale can be a problem in terms of the respondents' understanding about the answer that best fits their opinion in the item[8,9] and may convey discrepant information about the meaning of the intervals. Despite the little attention received on this aspect, countless articles[10-14] have pointed out inconveniences in the analysis of the internal structure due to problems related to writing, to the scale format and to response induction for the participants, generating response bias.

As the SEHEPS scale format does not display the label at all points of the scale, we decided to search for diverse validity evidence for the Brazilian aged population by testing two additional versions of the scale. This process gains relevance due to the resistance to implementing substantial changes in the instrument, even with technical support to do so. Another aspect to be noted is that the proposal with focus is an aged population different of the original target audience of the instrument, and that we have no information on how the format of the scales can affect the responses of the elderly, because we did not detect studies that have been developed to analyze the effects of the scale on the response pattern in older adults.

METHOD

This study seeks diverse validity evidence regarding the Brazilian version of SEHEPS for Older Adults and followed the *Standards* recommendations[15,16], with data collection procedures conducted between 2020 and 2022. Three evidence sources were tested: Content, which included the transcultural adaptation; Response Process; and Internal Structure, described below.

Content validity evidence/Cross-cultural study (Transcultural Adaptation – TCA)

Most of the current papers on cross-cultural instruments follow the stages of the process proposed by Beaton *et al.*[17], involving six stages: translation, translation synthesis, back-translation, experts' committee, pre-test and submission to a scientific committee. However, the recommendations proposed by Gjersing *et al.*[18] were used for the SEHEPS stages.

The process to adapt instruments can involve three levels[19]. The "adoption" is based on the assumption that the fewer interventions in the original format, the more similar will the new version be, implementing the fewest possible modifications to the instrument. The "adaptation" aims for the similarities of both versions to be only achieved by modifications in the instrument, such as changing wording of the item and inclusion or removal of item; and, finally, "assembly", which is concerned with improving the instrument for its target context, time and population and is not

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

intended to be fully equivalent or even imitate the instrument developed in the original culture, complementing the issues that are important in the original with new suggestions that may render the instrument suitable for the new target population[20]. This latter includes alterations in the format of instructions or of the scale and addition of items among others, and can also be called as "Severe!" by some authors[21,22].

It is also important to note that the translation is sometimes used as a synonym of test adaptation, although they are two different processes. Translating is linked to transforming the text from one language to another In turn, adaptations seek to verify that adequacy does not only refer to language but also to transforming the characteristics for them to adapt to the target population[20].

The instrument went through the translation and back-translation stages with two independent translators and a third one for synthesis in each stage, all respecting fluency between the languages and nativity. It was decided to apply two experts' panels: the first one focused on the equivalences and, therefore, on the relationship between the original version and the one for Brazil; and the second panel, which aimed at evaluating the modifications made in the instrument and in the format of the scale. The specialists received the finalized version in the first panel with three scale formats to be evaluated: Original (O) – therefore with the labels integrating more than one numbering, as previously pointed out, which can affect the internal structure analyses due to bringing about participant response bias; a 7-item version with all labels (7I); and a 5-item version with all labels.

The specialists were selected through their Lattes curricula and had as experience some contact with assistance to older adults, development of health promotion actions and/or of Health Education focused on the guidance of physical activity and exercises, and experience with measuring instruments and with multidisciplinary specialists[23] from different regions of the country[24,25]. Those who accepted the invitation received an electronic form developed in the *QuestionPro®* software, which was sent via email.

To evaluate the specialists' answers, the CVR (Content Validity Ratio) developed by Lawshe[26] was used, where the minimum score for the items' adherence to the latent variable will be based on the critical CVR values based on the final number of specialists[27].

Response process

Studies using this evidence are uncommon[28], where Response Process evidence is only reported in 1.8% of 283 tests analyzed in a previous study[29]. They are even rarer in Brazil.

Due to the rarity of studies on the response process, even more so with older adults, and to performance of the "assembly", it was decided to carry out the response process with all participants of the internal structure evidence stage; therefore, all respondents to the instruments also answered questions about the response process.

Internal structure validity evidence

The internal structure evidence refers to the extent to which the relationships between the test items and components are in line with the construct[16].

The participants were recruited via social networks, WhatsApp® groups and in sports events. They received the link to *QuestionPro®* for them to answer the instrument in its three versions and to evaluate all three versions as pointed out in the response process description.

Exploratory Factor Analysis (EFA)

The first analysis stage aims at verifying if the data can be subjected to factor analysis; in other words, to factorability by means of the Measures of Sampling Adequacy (MSA) of the sample. For this stage, they were evaluated by means of Bartlett's sphericity, matrix determinant and Kaiser-Meyer-Olkin (KMO) for the total sample and for the items, as recommended[30].

The dimensionality testing was performed with Parallel Analysis through the *Optimal Implementation of Parallel Analysis* (PA) with *Minimum Rank Factor Analysis*, which minimizes the common variance of the residuals[31], from the polychoric matrix.

The analyses were performed with a bootstrap of 5,000. Using internal replication is a contemporary guideline that intends to test stability and replication of a model[32]. In addition to it, the G-H index was adopted to assess replicability of the model.

The factors were extracted by means of the RULS (*Robust Unweighted Least Squares*) technique, which reduces residuals of the matrices, more robust in non-normal data. In addition, *Unidimensional Congruence* (UNICO>0.95), *Explained Common Variance* (ECV>0.80) and *Mean of Item Residual Absolute Loadings* (MIREAL<0.30) were adopted as

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

unidimensionality evaluation indicators[33]. In case the instrument proved to be multidimensional, oblique Promin rotation[34] and *Unique Directional Correlation* (ETA) by Pratt's Measure[35] would be used as a complementary way to evaluate the model.

Reliability

Reliability was measured by means of three indicators: Cronbach's alpha[36], Omega[37] and *Overall Reliability of fully-Informative prior Oblique N-EAP scores* (ORION)[38].

To keep or remove items from the model, the magnitude of the factor loadings, commonalities, absence of "cross-loading, Heywood cases" and interpretability of the factors were evaluated. The databases were analyzed individually for each of the versions, named as follows: Version 1: Original scale (O); Version 2: 7-item scale (7I); and Version 3: 5-item scale (5I).

It is noted that this study was carried out in compliance with the ethical principles according to the current National Health Council resolutions, and that it was approved by the Research Ethics Committee of the institution involved.

RESULTS

The results are presented according to the methodological procedure stages.

Content Validity evidence - Translation and back-translation

Figure 1 presents the result of Translations 1 and 2 and the synthesis made by the third translator.

	Original	**Translator 1**	**Translator 2**	**Synthesis of the translations**
	How confident are you that you could perform the prescribed exercises correctly...	Você tem segurança de que poderia praticar os exercícios prescritos corretamente...	O quão confiante você está de que realizará os exercícios prescritos...	Quão confiante você está de que realizará corretamente os exercícios prescritos...
1	...as often as prescribed by your clinician?	...na mesma frequência prescrita por seu clínico?	...na frequência prescrita pelo seu médico?	...com a frequência prescrita pelo seu médico?
2	...when you are bored by the program?	...quando se sente aborrecido(a) pelo programa?	...quando estiver entediado com o programa?	...quando estiver entediado com o programa?
3	...when you feel pain when exercising?	...quando sente dores durante o exercício?	...quando sentir dor durante os exercícios?	...quando sentir dor durante os exercícios?
4	...when you have to exercise alone?	...quando tem de se exercitar sozinho(a)?	...quando tiver de se exercitar sozinho?	...quando tiver que se exercitar sozinho(a)?
5	...when you do not enjoy it?	...quando não tem mais satisfação?	...quando não estiver gostando?	...quando não estiver gostando?
6	...when you are given written exercise instruction?	...quando as instruções são dadas por escrito?	...quando as instruções forem dadas por escrito?	...quando as instruções dos exercícios forem dadas por escrito?
7	...when you are too busy with other activities?	...quando está muito ocupado(a) com outras atividades?	...quando estiver ocupado com outras atividades?	...quando estiver muito ocupado(a) com outras atividades?
8	...when you are given video exercise instruction?	...quando as instruções são dadas por vídeo?	... quando as instruções forem dadas por vídeo?	...quando as instruções dos exercícios forem dadas por vídeo?
9	...when you feel tired?	...quando se sente cansado(a)?	....quando estiver cansado?	...quando se sentir cansado(a)?
10	...when you feel stressed?	...quando se sente estressado(a)?	...quando estiver estressado?	...quando se sentir estressado(a)?
11	...when you feel depressed?	...quando se sente deprimido(a)?	...quando estiver deprimido?	...quando se sentir deprimido(a)?
12	...when you do not have supervision or clinician feedback?	...quando não tem supervisão ou acompanhamento clínico?	...quando não for supervisionado ou não tiver um retorno do médico?	...quando não tiver supervisão ou retorno do médico?

FIGURE 1: Results of the translation process. São Paulo, SP, Brazil, 2022.

There were some discrepancies in terms of the header between T1 and T2, such as the following: "*Você tem segurança de que poderia praticar os exercícios prescrito corretamente...*" and "*O quão confiante você está de que*

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

realizará os exercícios prescritos...", which was synthesized as follows: "*Quão confiante você está de que realizará corretamente os exercícios prescritos...*".

In item 1, T1 suggested the translation as follows "*...na mesma frequência prescrita por seu clínico?*" and T2 proposed "*…na frequência prescrita pelo seu médico?*", whereas the T3 synthesis chose the "*…na frequência prescrita pelo seu médico?*" option; there were also discrepancies in item 4 between T1 and T2: "*…quando não tem mais satisfação?*" and "*…quando não estiver gostando?*"; the T3 synthesis decided to keep the same translation as in T2. In item 2, the T1 and T2 translations were as follows: "*...quando se sente aborrecido(a) pelo programa?*" and "*...quando estiver entediado com o programa?*"; and the "*…quando estiver entediado com o programa?*" option was chosen in the T3 synthesis. The other items presented no substantial discrepancies and did not alter understanding or comprehension about the topic.

First experts' panel: Equivalence

For the first experts' panel (Equivalence), we sent an invitation letter and the form via a link to *QuestionPro®* to 50 professionals from all five Brazilian regions, by means of the Lattes Platform. We received 11 forms fully filled-in, as follows: one specialist from the Midwest region, two from the North region, another two from the South region and six from the Southeast region. We had no answers from specialists in the Northeast region.

The versions were forwarded to analyze their semantic, idiomatic, experimental and conceptual equivalence. Table 1 presents the CVR values corresponding to the equivalence areas.

TABLE 1: CVR calculation corresponding to the 1st experts' panel: Equivalence. São Paulo, SP, Brazil, 2022.

	Items evaluated	**S**	**I**	**E**	**C**
Instructions	Please enclose in a circle your confidence level to finish the exercises prescribed for you to do at home.	1.00	0.82	0.45	1.00
Labels	Labels	1.00	1.00	1.00	1.00
Statement	How confident are you that you could perform the prescribed exercises correctly...	0.82	1.00	0.64	1.00
1	...as often as prescribed by your clinician?	1.00	0.82	0.82	0.82
2	...when you are bored by the program?	1.00	1.00	1.00	1.00
3	...when you feel pain when exercising?	1.00	1.00	1.00	1.00
4	...when you have to exercise alone?	1.00	1.00	1.00	1.00
5	...when you do not enjoy it?	0.64	1.00	1.00	1.00
6	...when you are given written exercise instruction?	1.00	1.00	1.00	1.00
7	...when you are too busy with other activities?	0.82	1.00	1.00	1.00
8	...when you are given video exercise instruction?	1.00	1.00	1.00	0.82
9	...when you feel tired?	0.82	1.00	1.00	1.00
10	...when you feel stressed?	0.82	1.00	1.00	1.00
11	...when you feel depressed?	0.82	1.00	1.00	1.00
12	...when you do not have supervision or clinician feedback?	1.00	1.00	1.00	1.00

Notes: CVR – Content Validity Ratio; S - Semantic ; I – Idiomatic; E – Experimental; C - Conceptual

All the items for the four equivalence areas obtained values above 0.59 (critical value for 11 specialists).

Some alterations were suggested. The term "*completing*" would be better understood as "*realizar*" in Portuguese. "*prescritos*" was changed to "*orientados*". For item 1 of the scale, one of the specialists asked about who should prescribe physical exercises: "Physical exercises are not prescribed by Physical Education trainers?". In items 5, 7, 9, 10 and 11, it was suggested to include the pronoun "you"; thus, as an example, item 5 should be changed from "*Quando não estiver gostando dos exercícios*" to "*Quando **você** não estiver gostando?*". "*Quando você estiver recebendo instruções dos exercícios por vídeo*" was suggested for item 8.

Complementary questions were also adopted for the equivalence areas (Table 2).

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

TABLE 2: Evaluation of the equivalence areas in general terms São Paulo, SP, Brazil, 2022.

	Questions	N	Y	IDK	NR
1	Do the items have the same or very similar meanings in both languages?	0	11	0	0
2	Are there grammatical cues that can render the items more easy or more difficult in the target language version?	6	3	1	1
3	In there any grammatical structure in the version of the source language items that lacks a parallel in the target language?	9	1	0	1
4	Is there any word in the items that, when translated, ceases to have one meaning and starts having more than one common meaning?	9	1	1	0
5	Are there differences between the versions of the target and source language items in relation to the use of metaphors, idiomatic expressions or colloquialisms?	11	0	0	0
6	Will the item formats and tasks required from the respondent be equally familiar in both linguistic versions?	2	9	0	0
7	When the passage is translated from the source to the target language, do the words and phrases in the translated version convey similar content and ideas to the source version?	0	11	0	0
8	Does the passage include content or require skills that may not be familiar to some respondents in any of both languages or cultural groups?	8	3	0	0
9	Are there cultural differences that would affect the chances for one answer to be chosen when the item is presented in the source or target language versions?	8	2	1	0
10	Are there cultural differences that would affect the chances for one answer to be chosen when the item is presented in the source or target language versions?	11	0	0	0
11	Does the item concept or writing have approximately the same familiarity and meaning in the source and target language versions?	0	11	0	0

Notes: N – No; Y – Yes; IDK – I Don't Know; NR – Not Relevant

In a general way, the results indicated adequate equivalence between the original and Brazilian versions. Even so, some notes were made: in the Instructions, it was pointed out that the expression "*nível de confiança*" is not a term commonly used for the aged population and that it might generate doubts.

On item 1, one specialist highlighted the wording "*as often as prescribed by your clinician*?" which may have a sense of "frequency" or represent "submission to a medical guideline", without necessarily being related to the "frequency" of that event. The suggestion was to change "*frequência*" to "*sempre*". Example: "*Sempre que prescrito pelo seu....*". Another specialist suggested reviewing the wording on the prescription of physical exercises by physicians, as physical exercise can be prescribed by other types of professionals, predominantly by Physical Education ones. Regarding "*frequência prescrita*", the suggestion was to include "*frequência orientada*" because it is a more frequently used expression when talking about physical exercises.

Second experts' panel: Content

For the second experts' panel, we sent an invitation letter and the form via a link to *QuestionPro®* to 114 professionals from all five Brazilian regions, by means of the Lattes Platform and directly to the professionals' email addresses. We received answers from 87 professionals, of which 12 filled-in the form in full. The specialists were distributed across the following regions: one from the South, three from the Northeast (Bahia, Ceará and Rio Grande do Norte); one from the North (Amazonas), and seven from the Southeast (Minas Gerais, Rio de Janeiro, and five from the state São Paulo). There was no specialist from the Midwest region in this stage.

The specialists in the first part of the form evaluated each of the items regarding the following: relevance of the items for the self-efficacy assessment; clarity in the wording of the items; and whether the items are worded correctly. Table 3 shows the CVR values corresponding to the items.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

TABLE 3: CVR calculation for the items' pertinence, clarity and writing. São Paulo, SP, Brazil, 2022.

		Pertinence	Clarity	Writing
1	...com a frequência prescrita pelo seu médico?	0.67	0.67	0.17
2	...quando estiver entediado com o programa?	0.67	0.33	0.50
3	...quando sentir dor durante os exercícios?	1.00	1.00	0.83
4	...quando tiver que se exercitar sozinho(a)?	1.00	1.00	1.00
5	...quando não estiver gostando?	1.00	1.00	0.83
6	...quando as instruções dos exercícios forem dadas por escrito?	0.83	1.00	0.83
7	...quando estiver muito ocupado(a) com outras atividades?	1.00	1.00	1.00
8	...quando as instruções dos exercícios forem dadas por vídeo?	0.83	1.00	0.67
9	...quando se sentir cansado(a)?	1.00	1.00	0.83
10	...quando se sentir estressado(a)?	1.00	1.00	1.00
11	...quando se sentir deprimido(a)?	1.00	0.83	0.83
12	...quando não tiver supervisão ou retorno do médico?	1.00	1.00	0.67

Note: CVR – Content Validity Ratio.

Item 1 for Writing (0.17) and item 2 for Clarity (0.33) and Writing (0.50) were below the critical value (0.56).

The specialists made suggestions in relation to pertinence of the items, namely: Similarity between the terms "*entediados*" and "*cansado*" and the expression "*quando não estiver gostando*". Another term pointed out was "*programa*" to designate guided physical activities; a suggestion would be to use the expression "*exercícios prescritos*" instead of "*programa*". The specialists indicated to substitute the term "*médico*", as it excludes other health professionals who are legally authorized to prescribe in Brazil. It was suggested to review the items with a negative connotation, as they entered into conflict with the "*quão confiante...*" statement of the instrument. It was also proposed to review the item that deals with the practice of physical activities when depressed, as self-efficacy independent depression is out of the person's control and management ability, as is also the case with pain.

In assessing clarity of the items in the statement, it was requested to review the expression "*Quão confiante*"; and we again had a question about the term "*programas*", as this is not common in the Brazilian context of physical activity. As well as the following expression: "*Com a frequência prescrita pelo seu médico*" since, depending on the aged person's schooling level, it is possible that the word "*frequência*" is not properly understood; therefore, "*o número de vezes que o seu médico recomendou*" was suggested as an option.

Another suggestion pointed out would be to replace the word "*prescrita*" by "*receitada*", as it can be a more usual expression, thus easing comprehension of the individuals evaluated. In item 2 it was suggested to complete the writing by specifying the type of program, as well as adding "*gostando de quê?*" in item 5. In the items with negative words, alterations such as "*...quando não estiver gostando?*" by "*...quando estiver insatisfeito?*" were also requested. In addition, for item 12 it was suggested to divide it; creating an item for supervision by the health professional and another one for the health professional's feedback.

There were suggestions for changes in the word "*prescrição*", as it is repeated in the same sentence; it may be that patients with lower schooling levels do not understand the meaning, so that the specialist suggested reviewing this term. It was also suggested to substitute "*médico*" with "*algum professional*". In item 5, it was suggested to complete the sentence, as it is not clear what the participant is enjoying: it can be the activity, the day, the body, instructions or even the trainer. In items 3, 7 and 8 it was suggested to avoid repeating the word "*exercício*" and to change it to "*atividade*", for example. To improve writing of item 8, it was suggested to replace "*...quando as instruções dos exercícios forem dadas por vídeo?*" by "*oferecidas por vídeo*". In item 9, "*Quando se sentir cansado (a)?*", specify if tiredness is due to the older adults' overall status, to everyday life, or to the exercises themselves. Reviews were also recommended in items 2 and 11; it was suggested to substitute the terms "*entediado*" and "*deprimido*" for others that are more accessible to the target population.

The CVR results for the complementary general questions of the instrument are presented in Table 4.

Table 4: Evaluation of the instrument in general terms São Paulo, SP, Brazil, 2022.

	Questions	No	Yes	CVR
1	Does the instrument measure the Self-Efficacy latent variable?	0	12	1.00
2	Does the instrument only measures what is intended to assess?	0	12	1.00
3	Is the language adopted suitable for the target population, in this case the Brazilian aged population?	5	7	0.17
4	Is the number of items enough for the measuring purpose?	1	11	0.83
5	Is the scale coherent with the proposal?	1	11	0.83
6	Is the scale size coherent with the latent variable?	2	10	0.67

Note: CVR – Content Validity Ratio.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

They again indicate that the criterion that assessed language adequacy for the target population resulted in a CVR (0.17) below the critical value. Clearly, this is the result of the various suggestions to adjust the writing. The specialists' suggestions were met to solve item 3 from Table 4, when they were asked to evaluate the language adopted.

As a result of testing three possible scale configurations, a set of questions were submitted to the specialists (Table 5).

TABLE 5: The specialists' considerations in relation to the scale options. São Paulo, SP, Brazil, 2022.

	Questions	Original	7-item	5-item
1	The labels better describe how I can feel.	0	3	9
2	Better label visualization.	0	3	9
3	Better understanding regarding the options included in the labels.	0	2	10
4	The layout of labels and numbers improves visualization of the answers.	0	2	10
5	Considering the scale structure in a general way after having answered all three options, which one do you believe is the most suitable for the answers to the items, thinking about the target population of older adults?	0	2	10

For the five questions formulated to the panel, the 5-item scale was the best evaluated. In opposition, the original scale received no indication for any of the questions made.

Changes implemented in the instrument

Due to the extensive suggestions made by the specialists, the instrument underwent various adaptations that can be seen in Figure 2: Synthesis of the translations, alterations suggested, and final writing.

	SYNTHESIS OF THE TRANSLATIONS	ALTERATIONS SUGGESTED	FINAL WRITING
	Quão confiante você está de que realizará corretamente os exercícios prescritos...	Qual o seu nível confiança para realizar corretamente os exercícios indicados...	Qual o seu nível confiança para realizar corretamente os exercícios indicados...
1	...com a frequência prescrita pelo seu médico?	...na quantidade que foi informada pelo profissional?	...na quantidade que foi informada pelo profissional?
2	...quando estiver entediado com o programa?	...quando estiver desmotivado com exercícios?	...quando me sinto motivado com os exercícios?
3	...quando sentir dor durante os exercícios?	...quando sentir dor durante os exercícios?	...quando sinto dor durante os exercícios?
4	...quando tiver que se exercitar sozinho(a)?	...quando tiver que realizar os exercícios sozinho(a)?	...quando tenho que realizar os exercícios sozinho(a)?
5	...quando não estiver gostando?	...quando não estiver gostando?	...quando estou gostando dos exercícios?
6	...quando as instruções dos exercícios forem dadas por escrito?	...quando as instruções dos exercícios forem passadas por escrito?	...quando as instruções são passadas por escrito?
7	...quando estiver muito ocupado(a) com outras atividades?	...quando estiver muito ocupado(a) com outras atividades?	...quando estou muito ocupado(a) com outras atividades?
8	...quando as instruções dos exercícios forem dadas por vídeo?	...quando as instruções dos exercícios forem passadas por vídeo?	...quando as instruções são passadas por vídeo?
9	...quando se sentir cansado(a)?	...quando se sentir cansado(a) por qualquer motivo?	...quando me sinto cansado(a) por qualquer motivo?
10	...quando se sentir estressado(a)?	...quando se sentir estressado(a) por qualquer motivo?	...quando me sinto estressado(a) por qualquer motivo?
11	...quando se sentir deprimido(a)?	...quando se sentir deprimido(a) por qualquer motivo?	...quando me sinto deprimido(a) por qualquer motivo?
12	...quando não tiver supervisão ou retorno do médico?	...quando não tiver supervisão ou retorno do profissional?	...quando tenho supervisão?
13			...quando tenho retorno do profissional?

FIGURE 2: Changes made in the scale. São Paulo, SP, Brazil, 2022.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

The specialists suggested that items 2 and 5 should be grouped into a single one; however, it was decided to keep them separated after the review. As the scale is ascending, item 2, which is a reverse item, was modified to "*quando me sinto motivado*" and item 5, to "*quanto gosto do exercício*". The agreement between the anchor and the complement (items) was reviewed to maintain coherence of the writing. Items 2, 3, 4, 7, 9, 10 and 11 were altered. In items 6 and 8, writing was modified to the present tense. Item 12 was divided into two items, due to having the preposition "or", which indicates two factors being evaluated by item[39,40], thus making the instrument have 13 items. The sentences in items 5, 12 and 13 were turned into positive.

Response Process and Internal Structure evidence

Characterization of the sample

The collection procedure was performed from August 31st to November 30th, 2022. In total, the questionnaire had 2,211 views, with 412 individuals starting the answers, 240 giving up after starting answering, and 174 finishing the questionnaire, with a 41.75% completion rate. The mean time for filling it out was 11 minutes, including the sociodemographic questionnaire and all three versions of the instrument.

The study participants were older adults aged between 60 and 95 years old, with a mean of 65 (Standard Deviation=6.72) and 68.53% female subjects. There was predominance of white race and/or skin color (72.56%), married individuals (41.57%), Catholic (51.81%), Higher Education (35%), incomes of more than 5 minimum wages (36.13%) and, in, 53.89% with retirement as income source. In relation to participants' origin, we had representatives from all five regions of the country, with a majority (85.71%) from the Southeast, 73.55% living in the state of São Paulo, 3.23% from the North, 3.24% from the Northeast, 2.59% from the Midwest, and 5.17% from the South.

When asked about pre-existing diseases, hypertension was recurrently stated by 20% of the participants, followed by anxiety with 9.72% and by hypothyroidism with 9.03%. Most of the sample reported taking some medication (75.30%) and, when asked about the type of access to health treatments, 27.44% use individual insurance and 23.78% resort to the Unified Health System (*Sistema Único de Saúde*, SUS). Regarding consumption of alcoholic beverages, 60.48% indicated not making use of them, and 97.62% reported not smoking.

We found that 88.10% of the participants performed physical activities and that, although having practiced them before, 11.09% were not doing so at that moment. For most of the participants, the purpose was health-related (38.42%), followed by aesthetics and medical indication (10.35%), and of moderate intensity for 66.87% of the subjects. The place indicated for the practice was at home for 24.64%; the most frequent types of activity were walking (34.59%), gymnastics (18.90%), bodybuilding (12.79%) and other types (13.66%), with 42.57% of the participants performing them alone, 35.64% in a group and 21.78% monitored by a professional. They perform the activities mostly in the morning (61.20%), afternoon (27.32%) and evening (11.48%), and 45.58% of the participants stated that they receive guidance from a trainer for the practices.

Response process

After the sample participants answered all three versions of the instrument, namely: Version 1, Original scale (O); Version 2, 7-item scale (7I); and Version 3, 5-item scale (5I), they were asked to answer some questions about these 3 versions. We decided to carry out this stage along with the internal structure collection stage due to the difficulties imposed by the pandemic in relation to contact with the participants and to the research design allowing the stages to be performed concomitantly.

In addition to that, the response process questions dealt with the comparison between ease of use of all three versions. Table 6 shows the results of this stage.

Table 6: Response process, overall evaluation of the instrument. São Paulo, SP, Brazil, 2022.

	O		7I		5I	
	n	%	n	%	n	%
Easier to answer	50	32.47%	32	20.78%	**72**	**46.75%**
Easier to understand	45	29.22%	31	20.13%	**78**	**50.65%**
Faster to answer	35	22.44%	31	19.87%	**90**	**57.69%**
More confidence in the answers	44	28.21%	27	17.31%	**85**	**54.49%**

Notes: O – Original; P7 – 7-item; P5 - 5-item.

For all four questions applied, the 5-item version obtained the best results. It is interesting to point out that the 7-item version had worse indices than the original version.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

Internal structure validity evidence

Factorability

The global adequacy indices of the sample for the "O" version were as follows: KMO=0.75, Bartlett's Sphericity=1,094.0 (DoF=78; p<0.001) and Matrix Determinant=0.000337. The following was obtained for the "7I" version: KMO=0.75, Bartlett's Sphericity=1,183.9 (DoF=78; p<0.00001) and Matrix Determinant=0.00000158. In turn, the "5I" version obtained KMO=0.79, Bartlett's Sphericity=1,213.5 (DoF=78; p<0.00001) and Matrix Determinant=0.0000159. In the "O" version, the result can indicate an item redundancy problem (p>0.000001)[41].

i. Dimensionality

Despite the original proposal for unidimensional model, the parallel analysis with the 13 items indicated the possibility of a bidimensional model by the mean of the eigenvalues and of a unidimensional one by the 95th percentile. Given this, we decided to perform *post hoc* dimensionality analyses[39,42].

Therefore, another two dimensionality techniques were applied: the *Bayesian Information Criterion* (BIC)[43] and HULL[44]. Again, we had both options; while BIC indicated a bidimensional model, HULL indicated a unidimensional model for all three versions of the scale.

For the "closeness of dimensionality" data, a multidimensional model is reinforced for all three databases. For the "O" version: UNICO=0.92; ECV=0.74 and MIREAL=0.35, which can indicate the existence of multidimensionality. In the "7I" version: UNICO=0.88; ECV=0.68 and MIREAL=0.48. Finally, version 3, with 5 items: UNICO=0.89; ECV=0.71 and MIREAL=0.42. As the techniques diverge and enable more than one configuration in the dimensions, we decided to explore the solutions with one and two dimensions.

ii. Analysis of the unidimensional version

The results obtained in the analysis are presented in Table 7.

TABLE 7: Factor loadings and the h² unidimensional model. São Paulo, SP, Brazil, 2022.

Item		O		7I		5I	
		λ	h²	λ	h²	λ	h²
1	na quantidade que foi informada pelo profissional?	**0.65**	0.43	**0.68**	0.47	**0.67**	0.45
2	quando me sinto motivado com os exercícios?	**0.63**	0.40	**0.69**	0.48	**0.72**	0.52
3	quando sinto dor durante os exercícios?	0.42	0.18	**0.66**	0.44	**0.57**	0.33
4	quando tenho que realizar os exercícios sozinho(a)?	**0.67**	0.45	**0.81**	0.67	**0.81**	0.66
5	quando gosto dos exercícios?	**0.74**	0.55	**0.68**	0.46	**0.62**	0.39
6	quando as instruções são passadas por escrito?	**0.61**	0.37	**0.75**	0.57	**0.80**	0.64
7	quando estou muito ocupado(a) com outras atividades?	**0.58**	0.33	**0.55**	0.31	**0.56**	0.31
8	quando as instruções são passadas por vídeo?	**0.63**	0.39	**0.65**	0.42	**0.67**	0.46
9	quando me sinto cansado(a) por qualquer motivo?	**0.70**	0.49	**0.64**	0.41	**0.71**	0.50
10	quando me sinto estressado(a) por qualquer motivo?	**0.66**	0.43	**0.58**	0.34	**0.68**	0.47
11	quando me sinto deprimido(a) por qualquer motivo?	**0.58**	0.33	**0.63**	0.40	**0.59**	0.35
12	quando tenho supervisão?	**0.56**	0.31	**0.62**	0.39	**0.63**	0.39
13	quando tenho retorno do profissional?	**0.57**	0.32	**0.52**	0.27	**0.44**	0.19

Notes: O – Original; P7 – 7-item; P5 - 5-item; λ – Factor loadings; h² – Commonalities.

The initial data from the "O" version presented factor loadings varying from 0.43 to 0,74, when the recommendation is a minimum of 0.30, for a sample comprised by more than 300 individuals. As our sample has less than 300 participants, the cutoff value corresponding to the factor loadings (λ) rose to 0,45, as recommended in the literature[45]. The commonalities (h²) varied between 0.18 and 0.55. The model has 47.20% of its variance explained.

For the "7I" version, the factor loadings varied from 0.52 to 0.82 and commonality ranged between 0.27 and 0.67. The model has 49.14% of its variance explained.

For the "5I" version, the factor loadings varied from 0.56 to 0.82 and commonality ranged between 0.20 and 0.66. The model has 54.02% of its variance explained.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

The "5I" version is the only one with more than 50% of its variance explained and has better factor loadings and commonalities than the "O" and "7I" models.

The reliability indices varied across the analysis databases from 0.91 to 0.89 for Cronbach's alpha and for Omega in all three versions. The indices corresponding to quality of the factor solution also showed adequate thresholds. The indices across all three formats are mostly similar.

However, the explained variance for the "5I" model is determinant, as it indicates that it better measures the latent variable, as shown in Table 8.

TABLE 8: Synthesis of the unidimensional model for the versions. São Paulo, SP, Brazil, 2022.

Synthesis	Index	Technique	O	7I	5I
Exploratory	Correlation Matrix Adequacy	Matrix Determinant	0.0003371	0.0000015	0.0000159
		Bartlett (DoF=78)	1,094.0	1,183.9	1,213.5
		KMO (Kaiser-Meyer-Olkin)	0.75	0.75	0.79
	Explained Variance (PA)		47.20%	49.14%	54.22%
	Polychoric Correlation (r_p=)		0.031 - 0.820	0.012 - 0.899	0.035 - 0.838
Reliability	Cronbach's Alpha		0.89	0.91	0.90
	McDonald's Omega		0.89	0.91	0.90
Replicability	G-H index (Latent/Observed)		0.89/0.87	0.91/0.94	0.92/0.89
Unidimensional Evaluation	Unidimensional Congruence (UNICO)		0.92	0.88	0.89
	Explained Common Variance (ECV)		0.74	0.68	0.71
	Mean of Item Residual Absolute Loading (MIREAL)		0.35	0.45	0.42
Quality and Efficacy	Factor Determinacy Index (FDI)[a]		0.947	0.987	0.959
	Sensitivity Ratio (SR)[a]		2.956	3.314	3.393
	Expected Percentage of True Differences (EPTD)[a]		92.40%	93.30%	93.50%

It is added that all three models showed good replicability (G-H index), both latent and observed, which indicates that the results obtained are consistent and tend to be found for other subpopulation groups.

iii. Analysis of the bidimensional version

Table 9 presents the primary values of the bidimensional model.

TABLE 9: Factor loadings, commonalities and ETA corresponding to the bidimensional model. São Paulo, SP, Brazil, 2022.

	Original					7P					5P				
	λ		h²	ETA		Λ		h²	ETA		λ		h²	ETA	
	1D	2D		1D	2D	1D	2D		1D	2D	1D	2D		1D	2D
Item 1	**0.86**	-0.09	0.67	**0.82**	0.07	**0.84**	-0.08	0.66	**0.81**	0.00	**0.87**	-0.14	0.67	0.82	0.00
Item 2	**1.03**	-0.25	0.85	**0.92**	0.00	**0.94**	-0.17	0.79	**0.89**	0.00	**0.81**	-0.02	0.64	**0.80**	0.00
Item 3	0.38	0.10	0.20	0.40	0.20	0.14	**0.76**	0.68	0.24	**0.79**	0.00	**0.75**	0.57	**0.03**	0.75
Item 4	**0.79**	-0.01	0.61	**0.76**	0.18	**0.65**	0.30	0.65	**0.70**	0.40	**0.62**	0.30	0.65	0.69	**0.42**
Item 5	**0.71**	0.15	0.63	**0.72**	0.34	**0.74**	0.02	0.55	**0.74**	0.08	**0.72**	0.04	0.49	**0.70**	0.00
Item 6	0.39	0.30	0.37	0.45	0.41	**0.80**	0.05	0.67	**0.81**	0.13	**0.81**	0.08	0.72	**0.83**	0.19
Item 7	0.15	**0.51**	0.36	0.25	**0.55**	0.00	**0.80**	0.64	0.00	**0.80**	-0.01	**0.75**	0.56	**0.00**	0.75
Item 8	0.25	**0.46**	0.40	0.33	**0.53**	**0.73**	-0.01	0.53	**0.73**	0.00	**0.62**	0.14	0.48	0.65	**0.24**
Item 9	0.21	**0.58**	0.52	0.33	**0.64**	-0.04	**0.89**	0.81	0.12	**0.89**	0.02	**0.93**	0.88	0.10	**0.90**
Item 10	0.07	**0.69**	0.53	0.19	**0.70**	-0.09	**0.98**	0.90	0.00	**0.95**	0.08	**0.80**	0.71	0.19	**0.82**
Item 11	-0.13	**0.82**	0.57	0.00	**0.76**	-0.07	**0.83**	0.73	0.15	**0.84**	-0.09	**0.90**	0.74	0.00	**0.86**
Item 12	-0.11	**0.77**	0.52	0.00	**0.72**	**0.79**	-0.11	0.57	**0.76**	0.00	**0.87**	-0.19	0.65	**0.80**	0.00
Item 13	-0.12	**0.79**	0.54	0.00	**0.74**	**0.58**	0.00	0.34	**0.58**	0.00	**0.66**	-0.20	0.36	0.60	0.00

Notes: O – Original; P7 – 7-item; P5 - 5-item.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

The initial data from the "O" version presented factor loadings from 0.46 to 1.02 (Heywood case), the commonalities varied between 0.19 and 0.85 and ETA between 0.53 and 0.92. Item 2 presented a factor loading infringement with a Heywood Case; in other words, the factor loading infringed the theoretical limits of factor loading from -1 to 1. The model has 50.67% of its variance explained (Table 9).

In the bidimensional model with the original scale, items 1 to 6 were aligned in one domain and items 7 to 3 in another, with items 3 and 6 presenting values below the desirable in both dimensions. This fact is reinforced by the ETA values (Pratt's Measure), which assists in confirmation of the factor loading values.

For the "7I" version, the factor loadings varied from 0.58 to 0.98, the commonalities from 0.33 to 0.90 and ETA from 0.58 to 0.89. The model has 51.88% of its variance explained. This model also had two dimensions: one comprised by items 1, 2, 4, 5, 6, 8 and 13, and the other consisting of items 3, 7, 9, 10 and 11.

In the last database, the "5I" version factor loadings varied from 0.61 to 0.92, the commonalities from 0.33 to 0.90 and ETA from 0.49 to 0.90. The model has 53.03% of its variance explained. This model also had two dimensions: one comprised by items 1,2 , 4, 5, 6, 8, 12 and 13, and the other consisting of items 3, 7, 9, 10 and 11. The "5I" and "7I" models had the same item configuration in the domains with adequate indices, which was not the case with the "O" version.

In all three versions analyzed, the content of the domains does not allow any interpretation. Therefore, despite the quantitative alignment, there is no qualitative alignment in terms of content. These two elements are indissociable in the internal structure analysis. Consequently, due to the infeasibility of both elements, the reliability, replicability, quality and efficacy indices of the scores lose interpretive power.

Thus, the unidimensional solution founds quantitative and qualitative support. The "5I" model has better indices than the "O" and "7I" ones and is the only one with more than 50% of explained variance.

The reliability indices varied between the analysis banks from 0.91 to 0.91 for Cronbach's Alpha and for Omega in the three versions. The syntheses of the models can be seen in Table 10.

TABLE 10: Factor loadings and the h² bidimensional model. São Paulo, SP, Brazil, 2022.

Synthesis	Index	Technique	O	7I	5I
Exploratory	Correlation Matrix	Matrix Determinant	0.0003371	0.0000015	0.0000159
	Adequacy	Bartlett (DoF=78)	1,094.0	1,183.9	1,213.5
		KMO (Kaiser-Meyer-Olkin)	0.75	0.75	80
	Explained Variance (PA)		50.67%	51.88%	53.03%
	Polychoric Correlation (r_p=)		0.031 - 0.820	0.012 - 0.899	0.035 - 0.838
Reliability	Cronbach's Alpha		0.89	0.91	0.91
	McDonald's Omega		0.89	0.91	0.91
	ORIONª		0.928; 0.880	0.933; 0.954	0.926; 0.939
Unidimensional Evaluation	Unidimensional Congruence (UNICO)		0.92	0.878	0.887
	Explained Common Variance (ECV)		0.74	0.679	0.708
	Mean of Item Residual Absolute Loading (MIREAL)		0.35	0.447	0.417
Quality and Efficacy	Factor Determinacy Index (FDI)ª		0.963; 0.938	0.966; 0.977	0.962; 0.969
	Sensitivity Ratio (SR)ª		3.580; 2.708	3.742; 4.558	3.525; 3.940
	Expected Percentage of True Differences (EPTD)ª		93.9%; 91.65	94.2%; 95.6%	93.8%; 94.6%

Notes: O – Original; P7 – 7-item; P5 - 5-item; a – Values by domain.

Thus, the set of techniques and indices employed points to a set of diverse validity evidence of adequate internal structure, consistent, robust and interpretable both quantitatively and qualitatively, with a recommendation for an instrument with a 5-item numerical and labeled scale.

DISCUSSION

This study had as its main objective to adapt and seek diverse validity evidence of the Brazilian version of SEHEPS aimed at the aged population, with the original version targeted at adults and in order to evaluate the practice of physical activities directed to rehabilitation, with not only cultural adaptations required, but also contextual.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

On some scales, performing an adaptation or modification is almost like developing a new instrument, as it is much more difficult or almost impossible to obtain the desired equivalence, regardless of what the adequacy areas are: linguistic, cultural or psychometric; in the case of our study, choosing "assembly" was the best option to provide a viable, useful and valid measure of the construct for the target culture and language[46].

After the specialists carried out the evaluation, we had suggestions for changes to improve equivalence in seven of all 12 items of the instrument, thus reinforcing the differences between the cultures and the target audience of the instrument regarding the practice developed and the terms used, in addition to language.

All specialists agreed that the items have similar meanings in both languages, but also all agreed that there were differences between the versions of the items in the target and source languages in relation to the use of metaphors, idioms or colloquialisms; therefore, the necessary adjustments were made respecting equivalence of their meaning.

Despite adequacy of the items, it was not necessary to exclude any. However, for better understanding, it was necessary to include an item, respecting the need to include aspects found in the target culture and lacking in the instrument to be adapted; therefore, item 12 was divided into two items, thus making the instrument have 13 items. The work by two panels is added; the first one focused on the equivalence areas and the second, effectively on the content. Both panels had a broad representation of the Brazilian regions, in order to ensure understanding of the content in the most diverse contexts[23-25].

There are guidelines[39,40] for the development and format of response scales, mainly the need for all items of a scale to be labeled and numbered, which did not happen in the original scale.

The proposal and testing of three configurations of the scale allows looking not only at the statistical adjustment aspects, but also understanding how the participants answer and generate associations between the instrument items.

Another aspect to sustain the performance of an "assembly" lies in the fact that the target audience was completely different from the original instrument's intended population. As the purpose of the instrument, the basis for our study, was to evaluate the physical activities and exercises performed remotely by the aged population, something relatively recent in our culture and which became popular and viable with the beginning of the pandemic and lasts until these days. We have not found in the literature, to date, any instrument that evaluates the behavior of those practicing them in this new context.

Given this first element of the content phase, it should be clear that the application of two panels focused on different evaluation elements contributed extensively to improving the instrument. Countless contributions emerged from the second panel that were not pointed out in the equivalence process. Despite the possibility of evoking experts' panels at various times in the search for evidence[15,16], they are rarely adopted. Many of the amendments proposed by the second panel were not even mentioned by the equivalence panel. Also, because several questions are unique to the objectives of each stage, which can also be understood as a contribution to future studies of our research. These notes lead us to reflect on the extent, quality and accuracy of the adjustments made in studies that only use the equivalence panel as a source for evaluating the instrument contents.

It should also be noted that it is not usually the role of equivalence panels to point out deeper changes in an instrument; again, as it is not their primary objective and because the protocols do not induce or recommend that researchers direct questions to specifically technical aspects related to writing, format, guidance, integration and interpretation, in addition to the sense of equivalence. So much so that both adaptation protocols mentioned in this study[17,18] do not even address, for example, the technical issues of the scale's writing and format, although analysis of these issues is pointed out in various books[20,39]. This fact leads us to point out that, in the practice, the equivalence panel is not sufficient for an adequate adaptation process.

Conduction of the broad response process, with all subjects participating together while filling out the three versions, allowed listing diverse elements and evidence that the 5-item model of the scale was superior to the other two models, both from the quantitative (internal structure) and qualitative (content and response process) points of view.

We must remember that, despite being established for decades in the *Standard*[15,16], previously integrated in the content validity phase, the response process is little used and much remains to be developed.

After overcoming the issues related to meanings, it is possible to verify in the results of the factorial analysis that they differ between all three scale formats, pointing to a 5-item scale model with all its labels numbered and with all the appropriate and consistent indices.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

The central fact lies in the extensive testing required to reach this point. It is not distant to the fact that, at first sight, the data offered by the unidimensional model do not rule out the possibility of using all three scale formats - not allowing to ignore the original problem, already addressed in the scale format. By itself, the problem of the scale precludes treating the internal structure data of the original model without distrust.

One aspect that may go unnoticed by a reader who is not familiar with the study of evidence is the existence of *post hoc* dimensionality testing, which includes additional techniques based on the results found and the integration of various model indicators.

Faced with the possibility of more than one dimensional configuration of the instruments, it was necessary to extend the testing, as well as to assess whether the results of the factor matrix would be interpretable.

Although the elements found in the analysis reinforce and direct the best solution for a unidimensional instrument, it is necessary to bear in mind that there is the possibility that future studies can expand the instrument and improve it in order to find a model that can accurately accommodate it for it to require two dimensions, which are but preliminarily indications.

CONCLUSION

Given the extensive testing and comparison between the three models, the recommendation is to use the 5-item version of the SEHEPS scale for the aged population, which presented adequate, satisfactory, consistent and robust levels of validity evidence.

As far as it was possible to investigate, it is noted that this is the first instrument in Brazil that, during its search for diverse cross-cultural validity evidence, has gone through an “assembly”.

As previously mentioned, holding two panels of independent specialists to assess different aspects of the content and adoption of the diverse response process evidence with the full sample of respondents and application of extensive testing in the internal structure, as well as the implementation of techniques for *post hoc* for dimensionality, can contribute to the work of other researchers, mainly due to the fact that this extensive procedure and techniques are not usually employed.

REFERENCES

1. Veras RP, Oliveira M. Aging in Brazil: the building of a healthcare model. Ciênc saúde colet. 2018 [cited 2020 Oct 13]; 23:1929-36. DOI: https://doi.org/10.1590/1413-81232018236.04722018.
2. Bottcher LB. Atividade física como ação para promoção da saúde. Rev. G. & S. 2019 [cited 2020 Oct 25]; 14:98-111. DOI: https://doi.org/10.26512/gs.v0i0.23324.
3. World Health Organization. The Heidelberg guidelines for promoting physical activity among older persons. WHO, Copenhagen, [cited 2020 Oct 13]; 1-14. Available from: https://apps.who.int/iris/handle/10665/108545.
4. Nóbrega AC, Freitas EV, Oliveira MA, Leitão MB, Lazzoli JK, Nahas RM, et al. Posicionamento oficial da Sociedade Brasileira de Medicina do Esporte e da Sociedade Brasileira de Geriatria e Gerontologia: atividade física e saúde no idoso. Rev Bras Med Esporte. 1999 [cited 2020 Oct 13]; 5:207-11. DOI: https://doi.org/10.1590/S1517-86921999000600002.
5. Ministério da Saúde. Guia de Atividade Física para a População Brasileira [Internet]. Brasília: 2021. Available from: https://aps.saude.gov.br/biblioteca/visualizar/MjA1MA==.
6. Mattos SM, Pereira DS, Moreira TM, Cestari VR, Gonzalez RH. Recomendações de atividade física e exercício físico durante a pandemia Covid-19: revisão de escopo sobre publicações no Brasil. Rev Bras Ativ Fís Saúde. 2020 [cited 2021 Sep 25]; 25:e0176. DOI: https://doi.org/10.12820/rbafs.25e0176.
7. Picha KJ, Lester M, Heebner NR, Abt JP, Usher EL, Capilouto G, et al. The self-efficacy for home exercise programs scale: development and psychometric properties. J Orthop Sports Phys Ther. 2019 [cited 2020 Oct 20]; 49(9):647-55. DOI: https://www.jospt.org/doi/10.2519/jospt.2019.8779.
8. Menold N. Rating-scale labeling in online surveys: An experimental comparison of verbal and numeric rating scales with respect to measurement quality and respondents’ cognitive processes. Sociol Methods Res. 2020 [cited 2021 Apr 19]; 49(1):9-107. DOI: https://doi.org/10.1177/0049124117729694.
9. Menold N, Kemper CJ. The impact of frequency rating scale formats on the measurement of latent variables in web surveys-an experimental investigation using a measure of affectivity as an example. Psihologija. 2015 [cited 2021 Mar 10]; 48(4):431-49. DOI: https://doi.org/10.2298/PSI1504431M.
10. Robie C, Meade AW, Risavy SD, Rasheed S. Effects of response option order on Likert-type psychometric properties and reactions. Educ Psychol Meas. 2022 [cited 2022 Dec 15]; 82(6):1107-29. DOI: https://doi.org/10.1177/00131644211069406.
11. Zhang X, Noor R, Savalei V. Examining the effect of reverse worded items on the factor structure of the need for cognition scale. PloS one. 2016 [cited 2021 out 10]; 11(6):e0157795. DOI: https://doi.org/10.1371/journal.pone.0157795.
12. Sonderen EV, Sanderman R, Coyne JC. Ineffectiveness of reverse wording of questionnaire items: Let’s learn from cows in the rain. PloS one. 2013 [cited 2021 Oct 15]; 8(7):e68967. DOI: https://doi.org/10.1371/journal.pone.0068967.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

13. Wetzel E, Böhnke JR, Rose N. A simulation study on methods of correcting for the effects of extreme response style. Educ Psychol Meas. 2016 [cited 2021 Oct 20]; 76(2):304-24. DOI: https://doi.org/10.1177/0013164415591848.
14. Weijters B, Millet K, Cabooter E. Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding. Int J Res Mark. 2021 [cited 2021 May 20]; 38(1):85-103. DOI: https://doi.org/10.1016/j.ijresmar.2020.04.002.
15. Association AER, Association AP, Association AE, Education NCoMi: Standards for Educational and Psychological Testing American Educational Research Association; 1999.
16. Association AER, Association AP, Association AE, Education NCoMi: Standards for Educational and Psychological Testing American Educational Research Association; 2014.
17. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000 [cited 2021 May 10]; 25(24):3186-91. Available from: https://journals.lww.com/spinejournal/Citation/2000/12150/Guidelines_for_the_Process_of_Cross_Cultural.14.aspx.
18. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. BMC Med Res Methodol. 2010 [cited 2021 Apr 19]; 10:13. DOI: https://doi.org/10.1186/1471-2288-10-13.
19. Van de Vijver FJ, Leung K. Methods and data analysis for cross-cultural research. Cambridge University Press; 2021.
20. Iliescu D. Adapting tests in linguistic and cultural situations. Cambridge University Press; 2017.
21. Stewart AL, Thrasher AD, Goldberg J, Shea JA. A framework for understanding modifications to measures for diverse populations. J Aging Health. 2012 [cited 2021 Jul 15]; 24(6):992-1017. DOI: https://doi.org/10.1177%2F0898264312440321.
22. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009 [cited 2021 Jun 12]; 12(4):419-29. DOI: https://doi.org/10.1111/j.1524-4733.2008.00470.x.
23. Epstein J, Osborne RH, Elsworth GR, Beaton DE, Guillemin F. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. J Clin Epidemiol. 2015 [cited 2021 Jun 15]; 68(4):360-9. DOI: https://doi.org/10.1016/j.jclinepi.2013.07.013.
24. Costa Palacio D, Rebustini F, Oliveira DB, Neto JP, Barbieri W, Sanchez TP, et al. Dental vulnerability scale in primary health care: evidence of content and structure internal validity. BMC Oral Health. 2021 [cited 2022 Sep 18]; 21:421. DOI: https://doi.org/10.1186/s12903-021-01742-6.
25. Nobile GG, Barrera SD, Rebustini F. Avaliação da alfabetização: elaboração e validação de conteúdo do IBALEC. Rev. Psicopedagogia. 2021 [cited 2023 Jan 11]; 38(117):333-45. DOI: http://dx.doi.org/10.51207/2179-4057.20210028.
26. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975 [cited 2021 May 22]; 28(4):563-75. DOI: https://doi.org/10.1111/j.1744-6570.1975.tb01393.x.
27. Wilson FR, Pan W, Schumsky DA. Recalculation of the critical values for Lawshe's content validity ratio. Meas Eval Couns Dev. 2012 [cited 2021 Apr 08]; 45(3):197-210. DOI: https://doi.org/10.1177/0748175612440286.
28. Anastasi A, Urbina S. Testagem psicológica. Artmed; 2000.
29. Cizek GJ, Rosenberg SL, Koons HH. Sources of validity evidence for educational and psychological tests. Educ Psychol Meas. 2008 [cited 2021 Aug 18]; 68(3):397-412. DOI: https://doi.org/10.1177/0013164407310130.
30. Lorenzo-Seva U, Ferrando PJ. MSA: the forgotten index for identifying inappropriate items before computing exploratory item factor analysis. Methodology. 2021 [cited 2022 Dec 11]; 17(4):296-306. DOI: https://doi.org/10.5964/meth.7185.
31. Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychol Methods. 2011 [cited 2022 Dec 11]; 16(2):209. DOI: https://doi.org/10.1037/a0023353.
32. Goretzko D, Bühner M. Robustness of factor solutions in exploratory factor analysis. Behaviormetrika. 2022 [cited 2022 Dec 11]; 49(1):131-48. DOI: https://doi.org/10.1007/s41237-021-00152-w.
33. Ferrando PJ, Lorenzo-Seva U. Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. Educ Psychol Meas. 2018 [cited 2022 Dec 11]; 78(5):762-80. DOI: https://doi.org/10.1177/0013164417719308.
34. Lorenzo-Seva U. Promin: A method for oblique factor rotation. Multivariate Behav Res. 1999 [cited 2022 Dec 11]; 34(3):347-65. DOI: https://doi.org/10.1207/S15327906MBR3403_3.
35. Wu AD, Zumbo BD, Marshall SK. A method to aid in the interpretation of EFA results: An application of Pratt's measures. Int J Behav Dev. 2014 [cited 2022 Dec 11]; 38(1):98-110. DOI: https://doi.org/10.1177/0165025413506143.
36. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 [cited 2022 Dec 11]; 16(3):297-334. DOI: https://doi.org/10.1007/BF02310555.
37. McDonald RP. Test theory: A unified treatment. Psychology Press; 2013.
38. Ferrando PJ, Lorenzo-Seva U. A note on improving EAP trait estimation in oblique factor-analytic and item response theory models. Psicológica. 2016 [cited 2022 Dec 11]; 37(2):235-47. Available from: https://psycnet.apa.org/record/2016-34732-007.
39. Bandalos DL. Measurement theory and applications for the social sciences. Guilford Publications; 2018.
40. Johnson RL, Morgan GB. Survey scales: A guide to development, analysis, and reporting. Guilford Publications; 2016.
41. Field A. Discovering statistics using IBM SPSS statistics 5th ed. 2018.
42. Auerswald M, Moshagen M. How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions. Psychol Methods. 2019 [cited 2023 Jan 17]; 24(4):468. DOI: https://doi.org/10.1037/met0000200.

DOI: http://dx.doi.org/10.12957/reuerj.2023.73954

43. Schwarz G. Estimating the dimension of a model. The annals of statistics. 1978 [cited 2023 Feb 11]; 1:461-4. Available from: https://www.jstor.org/stable/2958889.
44. Lorenzo-Seva U, Timmerman ME, Kiers HA. The Hull method for selecting the number of common factors. Multivariate Behav Res. 2011 [cited 2023 Feb 11]; 46(2):340-64. DOI: https://doi.org/10.1080/00273171.2011.564527.
45. Hair J, Black W, Anderson R, Babin B. Multivariate data analysis (8th ed.). Cengage Learning EMEA. 2018.
46. He J, van de Vijver F. Bias and equivalence in cross-cultural research. J Psychol. 2012 [cited 2023 Feb 22]; 2(2):2307-19. DOI: https://doi.org/10.1108/09504120910935093.

Author's contributions:
Conceptualization, D.R.V., K.J.P. and F.R.; methodology, D.R.V. and F.R.; software, D.R.V. and F.R.; validation, D.R.V. and F.R.; investigation, D.R.V.; formal analysis, D.R.V. and F.R.; resources, D.R.V. and F.R.; data curation, D.R.V. and F.R.; manuscript writing, D.R.V. and F.R.; writing—review and editing, D.R.V., K.J.P. and F.R.; visualization, A.R.S.I. and D.C.A.; supervision, F.R.; project administration, F.R. All authors have read and agreed to the published version of the manuscript.