



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: [www.elsevier.com/locate/mex](http://www.elsevier.com/locate/mex)



## Method Article

# Method of preparing a document for survey instrument validation by experts



N. Elangovan\*, E. Sundaravel

*School of Business and Management, Christ University, Bengaluru, India*

### ABSTRACT

Validation of a survey instrument is an important activity in the research process. Face validity and content validity, though being qualitative methods, are essential steps in validating how far the survey instrument can measure what it is intended for. These techniques are used in both scale development processes and a questionnaire that may contain multiple scales. In the face and content validation, a survey instrument is usually validated by experts from academics and practitioners from field or industry. Researchers face challenges in conducting a proper validation because of the lack of an appropriate method for communicating the requirement and receiving the feedback.

In this Paper, the authors develop a template that could be used for the validation of survey instrument.

In instrument development process, after the item pool is generated, the template is completed and sent to the reviewer. The reviewer will be able to give the necessary feedback through the template that will be helpful to the researcher in improving the instrument.

© 2021 The Author(s). Published by Elsevier B.V.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

### ARTICLE INFO

*Method name:* Instrument validation method

*Keywords:* Face validity, Content validity, Expert validation, Information requirements, Feedback form

*Article history:* Received 1 July 2020; Accepted 22 March 2021; Available online 8 April 2021

\* Corresponding author.

*E-mail address:* [elangovan.n@christuniversity.in](mailto:elangovan.n@christuniversity.in) (N. Elangovan).

## Specifications table

Subject Area	Psychology
More specific subject area	Scale or Instrument validation by experts
Method name	Instrument validation method
Name and reference of original method	American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. [1]. Standards for educational & psychological testing. Washington, DC: Author. Boateng et al. ([3]. Best practices for developing and validating scales for health, social, and behavioral research: a primer. <i>Frontiers in public health</i> , 6, 149. Willis and Lessler [28]. Question appraisal system QAS-99. <i>National Cancer Institute</i> .
Resource availability	<i>The template is developed in MS Office word. Researchers can use any word processing tool to create the document</i>

### \*Method details

## Introduction

Survey instruments or questionnaires are the most popular data collection tool because of its many advantages. Collecting data from a huge population in a limited time and at a lower cost, convenient to respondents, anonymity, lack of interviewer bias and standardization of questions are some of the benefits. However, an important disadvantage of a questionnaire is poor data quality due to incomplete and inaccurate questions, wording problems and poor development process. The problems are critical and can be avoided or mitigated [14].

To ensure the quality of the instrument, using a previously validated questionnaire is useful. This will save time and resources in development process and testing its reliability and validity. However, there can be situations wherein a new questionnaire is needed [5]. Whenever a new scale or questionnaire needs to be developed, following a structured method will help us to develop a quality instrument. There are many approaches in scale development and all the methods include stages for testing reliability and validity among them.

Even though there are many literatures available on the reliability and validity procedures, many researchers struggle to operationalize the process. Collingridge [8] wrote in the Methodspace blog of Sage publication that he repeatedly asked professors on how to validate the questions in a survey and unfortunately did not get an answer. Most of the time, researchers send the completely designed questionnaire with the actual measurement scale without providing adequate information for the reviewers to provide proper feedback. This paper is an effort to develop a document template that can capture the feedback of the expert reviewers of the instrument.

This paper is structured as follows: Section 1 provides the introduction to the need for a validation format for research, and the fundamentals of validation and the factors involved in validation from various literature studies are discussed in Section 2. Section 3 presents the methodology used in framing the validation format. Section 4 provides the results of the study. Section 5 presents explanation of how the format can be used and feedback be processed. Finally, Section 6 concludes the paper with a note on contribution.

## Review of literature

A questionnaire is explained as “an instrument for the measurement of one or more constructs by means of aggregated item scores, called scales” [21]. A questionnaire can be identified on a continuum of unstructured to structure [14]. A structured questionnaire will “have a similar format, are usually statements, questions, or stimulus words with structured response categories, and require a judgment or description by a respondent or rater” [21]. Research in social science with a positivist paradigm began in the 19th century. The first use of a questionnaire is attributed to the Statistical Society of London as early as 1838. Berthold Sigismund proposed the first guidelines for questionnaire development in 1856, which provided a definite plan for the questionnaire method [13]. In 1941, The

British Association for the Advancement of Science provided Acceptance of Quantitative Measures for Sensory Events [26] provided a much pervasive application or questionnaire in research, similar to Guttman scale [15], Thurstone Scale [27] and Likert Scale [18].

Carpenter [6] argued that scholars do not follow the best practices in the measurement building procedure. The author claims that “the defaults in the statistical programs, inadequate training and numerous evaluation points can lead to improper practices”. Many researchers have proposed techniques for scale development. We trace the prominent methods from the literature. Table 1 presents various frameworks in scale development.

Reeves and Marbach-Ad [22] argued that the quantitative aspect of social science research is different from science in terms of quantifying the phenomena using instruments. Bollen [4] explained that a social science instrument measures latent variables that are not directly observed, although inferred from observable behaviour. Because of this characteristic of social science measures, there is a need to ensure that what is being measured actually is measuring the intended phenomenon.

The concept of reliability and validity was evolved as early as 1896 by Pearson. The validity theory from 1900 to 1950 basically dealt with the alignment of test scores with other measures. This was operationally tested by correlation. The validity theory was refined during the 1950s to include criterion, content and construct validity. Correlation of the test measure to an accurate criterion score is the criterion validity. In 1955, criterion validity was proposed as concurrent validity and predictive validity. Content validity provides “domain relevance and representativeness of the test instrument”. The concept of construct validity was introduced in 1954 and got increased emphasis, and from 1985 it took a central form as the appropriate test for validity. The new millennium saw a change in the perspectives of validity theory. Contemporary validity theory is a metamorphosis of epistemological and methodological perspectives. Argument-based approach and consequences-based validity are some new concepts that are evolving [24].

American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) jointly developed ‘Standards for educational and psychological testing’. It is described as “the degree to which evidence and theory support the interpretations of test scores for posed uses of tests” [1].

Based on the ‘Standards’, the validity tests are classified on the type of evidence. Standards 1.11 to 1.25, describe various evidence to test the validity [1]. Table 2 presents different types of validities based on evidence and their explanation.

Souza et al. [25] argued that “there is no statistical test to assess specifically the content validity; usually researchers use a qualitative approach, through the assessment of an experts committee, and then, a quantitative approach using the content validity index (CVI).”

Worthington and Whittaker [29] conducted a content analysis on new scales developed between 1995 and 2004. They specifically focused on the use of Exploratory and Confirmatory Factor Analysis (EFA & CFA) procedures in the validation of the scales. They argued that though the post-tests in the validation procedure, which are usually based on factor-analytic techniques, are more scientific and rigorous, the preliminary steps are necessary. Mistakes committed in the initial stages of scale development lead to problems in the later stages.

Messick [20] proposed six distinguishable notions of construct validity for educational and psychological measurements. Among the six, the foremost one is the content validity that looks at the relevance of the content, representativeness and technical quality. In a similar way Oosterveld et al. [21] developed taxonomy of questionnaire design directed towards psychometric aspects. The taxonomy introduces the following questionnaire design methods: (1) coherent, (2) prototypical, (3) internal, (4) external, (5) construct and (6) facet design technique. These methods are related “to six psychometric features guiding them face validity, process validity, homogeneity, criterion validity, construct validity and content validity”. The authors presented these methods under four stages: (1) concept review, (2) item generation, (3) scale development and (4) evaluation. After the definition of the construct in the first stage, the item pool is developed. The item production stage “comprises an item review by judges, e.g., experts, or potential respondents, and a pilot administration of the preliminary questionnaire, the results of which are subsequently used for refinement of the items”.

**Table 1**  
Frameworks of Scale development.

Author & Framework	Steps	Remarks
Churchill [7] Paradigm for Developing Better Measures of Marketing Constructs	8-Step process. (1) specify domain of construct, (2) generate a sample of items, (3) collect data, (4) purify measure, (5) collect data, (6) assess reliability, (7) assess validity and (7) develop standards.	He recommended a multi-item measure to diminish the difficulties of a single-item measure. Experts are consulted during the item development stage. A focus group of 8 to 10 participants are triggered for an open discussion on the concept. When a researcher wants to include items, experienced researchers can attest identical statements. Every statement will be reviewed for the preciseness of words, double-barreled statements, positive and negative statements, socially acceptable responses and even to remove the item.
Hinkin [16] Three stages scale development	Following are the stages of the scale construction: (1) Item generation, (2) Scale development under which Design of developmental study, scale construction and reliability assessment are the steps, (3) Scale evaluation.	The study recommended the use of subject matter experts in developing the conceptual definition.
Hinkin et al. [17] Seven-step scale development procedure.	(1) Item Generation, (2) Content Adequacy Assessment, (3) Questionnaire Administration, (4) Factor Analysis, (5) Internal Consistency Assessment, (6) Construct Validation and (7) Replication.	The authors propose 'content adequacy assessment' as a necessary step in scale development. They are of the concern that this step is being overlooked and researchers land in trouble after collecting large datasets. The authors argue that there are several content assessment methods and recommend using experts in a content domain for the assessment.
Rossiter [23] C-OAR-SE scale development.	The steps of the framework are as follows: (1) Construct definition, (2), Object classification, (3) Attribute classification, (4) Rater identification, (5) Scale formation, and (6) Enumeration and reporting.	This framework has been exclusively proposed for scale development in marketing research where the construct is defined in terms of object, attribute and rater entity (OAR). The scale depends on only content validity than any other types of validity and places more emphasis on reasonable arguments and the agreement of experts. The author distinguishes content validity from face validity and argues that "content validity is conducted before the scale is developed, that the items will properly represent the construct", whereas "face validity is a post hoc claim that the items in the scale measure the construct". The author presented a prototype of an expert judge's rating form.
DeVellis [9] Eight-step scale construct method	(1) Determine clearly what it is you want to measure, (2) Generate the Item pool, (3) Determine the format for measurement, (4) Have the initial item pool reviewed by experts, (5) Consider the inclusion of Validation items, (6) Administer Items to a development sample, (7) Evaluate the items and (8) Optimize scale length.	The author proposes an exclusive step in which the generated items are validated by experts. The expert panel is required to evaluate how each item is relevant to measure the concept based on the working definition of the construct. The experts are also expected to assess the clarity and conciseness of the items. The experts can also indicate any missing phenomenon that the researcher failed to include. However, the final decision on considering the expert's comments is with the researcher.

(continued on next page)

**Table 1** (continued)

Author & Framework	Steps	Remarks
Carpenter [6] 10 step scale development and Reproting	(1) Research the intended meaning and breadth of the theoretical concept, (2) Determine sampling procedure, (3) Examine data quality, (4) Verify the factorability of the data, (5) Conduct Common Factor Analysis, (6) Select factor extraction method, (7) Determine the number of factors, (8) Rotate factors, (9) Evaluate items based on a priori criteria and (10) Present results.	The author claims that "Interviews, focus groups, and expert feedback are critical in the item generation and dimension identification process" and recommends that "the pool of items needs to be concise, clear, distinct, and reflect the chosen conceptual definition".

**Table 2**

Types of validity.

Types of evidence	Explanation
Content-oriented evidence	"Validity evidence can be obtained from an analysis of the relationship between the content of a test and the construct it is intended to measure".
Evidence regarding cognitive processes	"Evidence concerning the fit between the construct and the derailed nature of the performance or response actually engaged in by test takers".
Evidence regarding internal structure	"Analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based".
Evidence concerning relationships with conceptually related constructs	"Evidence based on relationships with other variables provides evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations". This includes convergent and discriminant validity.
Evidence regarding relationships with criteria	"Evidence of the relation of test scores to a relevant criterion". This includes concurrent and predictive validity.
Evidence based on consequences of tests	"The validation process involves gathering evidence to evaluate the soundness of these proposed interpretations for their intended use".

(Source: [1])

## What needs to be checked?

This paper mainly focuses on the expert validation done under the face validity and content validity stages. Martinez [19] provides a clear distinction between content validity and face validity. "Face validity requires an examination of a measure and the items of which it is composed as sufficient and suitable 'on its face' for capturing a concept. A measure with face validity will be visibly relevant to the concept it is intended to measure, and less so to other concepts". Though face validity is the quick and excellent first step for assessing the appropriateness of measure to capture the concept, it is not sufficient. It needs to be interpreted along with other forms of measurement validity.

"Content validity focuses on the degree to which a measure captures the full dimension of a particular concept. A measure exhibiting high content validity is one that encompasses the full meaning of the concept it is intended to assess" [19]. An extensive review of literature and consultation with experts ensures the validity of the content.

From the review of various literature studies, we arrive at the details of validation that need to be done by experts. Domain or subject matter experts both from academic and industry, a person with expertise in the construct being developed, people familiar with the target population on whom the instrument will be used, users of the instrument, data analysts and those who take decisions based on the scores of the test are recommended as experts. Experts are consulted during the concept development stage and item generation stage. Experts provide feedback on the content, sensitivity and standard settings [10].

During the concept development stage, experts provide inputs on the definition of the constructs, relating it to the domain and also check with the related concepts. At the item generation stage,

experts validate the representativeness and significance of each item to the construct, accuracy of each item in measuring the concept, inclusion or deletion of elements, logical sequence of the items, and scoring models. Experts also validate how the instrument can measure the concept among different groups of respondents. An item is checked for its bias to specific groups such as gender, minority groups and linguistically different groups. Experts also provide standard scores or cutoff scores for decision making [10].

The second set of reviewers who are experts in questionnaire development basically check the structural aspects of the instrument in terms of common errors such as double-barreled, confusing and leading questions. This also includes language experts, even if the questionnaire is developed in a popular language like English. Other language experts are required in case the instrument involves translation.

There were many attempts to standardize the validation of the questionnaire. Forsyth et al. [11] developed a Forms Appraisal model, which was an exhaustive list of problems that occur in a questionnaire item. This was found to be tiresome for experts. Fowler and Roman [12] developed an 'Interviewer Rating Form', which allowed experts to comment on three qualities: (1) trouble reading the question, (2) respondent not understanding the meaning or ideas in the question and (3) respondent having difficulty in providing an answer. The experts had to code as 'A' for 'No evidence of a problem', 'B' for 'Possible problem' and 'C' for 'Definite Problem'. Willis and Lessler [28] developed a shorter version of the coding scheme for evaluation of questionnaire items called "Question appraisal system (QAS)". This system evaluates each item on 26 problem areas under seven heads. The expert needs to just code 'Yes' or 'No' for each item. Akkerboom and Dehue [2] developed a systematic review of a questionnaire for an interview and self-completion questionnaire with 26 problems items categorized under eight problem areas.

Hinkin [16] recommended a "best practices" of "clearly cite the theoretical literature on which the new measures are based and describe the manner in which the items were developed and the sample used for item development". The author claims that "in many articles, this information was lacking, and it was not clear whether there was little justification for the items chosen or if the methodology employed was simply not adequately presented".

Further to the qualitative analysis of the items, recent developments include quantitative assessments of the items. "The content adequacy of a set of newly developed items is assessed by asking respondents to rate the extent to which items corresponded with construct definitions" [16]. Souza et al. [25] suggest using the Content Validity Index (CVI) for the quantitative approach. Experts evaluate every item on a four-point scale, in which "1 = non-equivalent item; 2 = the item needs to be extensively revised so equivalence can be assessed; 3 = equivalent item, needs minor adjustments; and 4 = totally equivalent item". The number of items with a score of 3 or 4 and dividing it with the total number of answers is used to calculate an index of CVI. The CVI value is the percentage of judges who agree with an item, and the index value of at least 0.80 and higher than 0.90 is accepted.

### **Information to be provided to the experts**

The problems with conducting a face validity and content validity may be attributed to both scale developer and the reviewer. Scale developers do not convey their requirements to the experts properly, and experts are also not sure about what is expected by the researcher. Therefore, a format is developed, which will capture the requirements information for scale validation from both the researcher and the experts.

#### *Covering letter*

A covering letter is an important part when sending a questionnaire for review. It can help in persuading a reviewer to support the research. It should be short and simple. A covering letter first invites the experts for the review and provides esteem to the expert. Even if the questionnaire for review is handed over personally, having a covering letter will serve instructions for the review process and the expectations from the reviewer.

Boateng et al. [3] recommended that the researcher specifies the purpose of the construct or the questionnaire being developed, justifying the development of new instruments by confirming that there are no existing instruments are crucial. If there are any similar instruments, how different is the proposed one from the existing instruments.

The covering letter can mention the maximum time required for the review and any compensation that the expert will be awarded. This will motivate the reviewer to contribute their expertise and efforts. Instructions on how to complete the review process, what aspects to be checked, the coding systems and how to give the feedback are also provided in the covering letter. The covering letter ends with a thank you note in advance and personally signed by the instrument developer. Information on further contact details can also be provided at the end of the covering letter.

### *Introduction to research*

Boateng et al. [3] proposed that it is an essential step to articulate the domain(s) before any validation process. They recommend that “the domain being examined should be decided upon and defined before any item activity. A well-defined domain will provide a working knowledge of the phenomenon under study, specify the boundaries of the domain, and ease the process of item generation and content validation”.

In the introduction section, the research problem being addressed, existing theories, the proposed theory or model that will be investigated, list of variables/concepts that are to be measured can be elaborated. Guion [30] defended that for those who do not just accept the content validity by the evaluations of operational definition alone, five conditions will be a tentative answer: “(1) the content domain should be grounded in behavior with a commonly accepted meaning, (2) the content domain must be defined in a manner that is not open to more than one interpretation, (3) the content domain must be related to the purposes of measurement, (4) qualified judges must agree that the domain has been sufficiently sampled and (5) the response content must be dependably observed and evaluated.” Therefore, the information provided in the ‘Introduction’ section will be helpful to the expert to do a content validity at the first step.

### *Construct-wise item validation*

After the need for the measure or the survey instrument is communicated, the domain is validated. The next step is to validate the items. Validation may be done for developing a scale for a single concept or as a questionnaire with multiple concepts of measure. For a multiple construct instrument, the validation is done construct-wise.

In an instrument with multiple constructs, the Introduction provides information at the theory level. The domain validation is done to assess the relevance of the theory to the problem. In the next section, the domain validation is done at variable level. Similar to the Introduction, details about the construct is provided. The definition of the construct, source of the definition, description of the concept, and the operational definition are shared to the experts. Experts will validate the construct by relating it to the relevant domain. If the conceptualization and definition are not properly done, it will result in poor evaluation of the items.

New items are developed by deductive method or inductive method. In deductive methods, items are generated from already existing scales and indicators through literature review. In inductive technique, the items are generated through direct observation, individual interviews, focus group discussion and exploratory research. It is necessary to convey how the item is generated to the expert reviewer. Even when the item or a scale is adopted unaltered; it becomes necessary to validate them to assess their relevance to a particular culture or a region. Even in such situations, it is necessary to inform the reviewer about the source of the items.

Experts review each item and the construct as a whole. For each item, item code, the item statement, measurement scale, the source of item and description of the item are provided. In informing the source of the item, there are three options. When the item is adopted as it is from the previous scales, the source can be provided. If the item is adapted by modifying the earlier item, the source and the original item can be informed along with description of modification done. If the

item is developed by induction, the item source can be mentioned. First, experts evaluate each item to assess if they represent the domain of the construct and provide their evaluation and 4-point or 3-point scale. When multiple experts are used for the validation process, this score can also be used for quantitative evaluation. The quality parameters of the item are further evaluated. Researchers may choose the questionnaire appraisal scheme from many different systems available. An open remarks column is provided for experts to give any feedback that is not covered by the format. A comments section is provided at the end of the construct validation section where the experts can give the feedback such underrepresentation of the construct by the items.

#### *Validation of demography items*

The same way, the information regarding each of the demography items that will be required in the questionnaire is also included in the format. Finally, space for the expert to comment on the entire instrument is also provided. The template of the evaluation form is provided in the Appendix.

#### *Inferring the feedback*

Since the feedback is a qualitative approach, mathematical or statistical approach is not required for inferring the review. Researcher can retain, remove or modify the statements of the questionnaire as indicated by the experts as essential, not essential and modify. As we have recommended using the quality parameters of QAS for describing the problems and issues, researcher will get a precise idea on what need to be corrected. Remarks by the experts will carry additional information in form of comments or suggestion that will be easy to follow when revising the items. General comments at the end of each scale or construct will provide suggestions on adding further items to the construct.

### **Conclusion**

Despite the various frameworks available for the available to the researchers for developing the survey instrument, the quality of the same is not at the desirable level. Content validation of the measuring instrument is an essential requirement of every research. A rigorous process expert validation can avoid the problems at the latter stage. However, researchers are disadvantaged at operationalising the instrument review process. Researchers are challenged with communicating the background information and collecting the feedback. This paper is an attempt to design a standard format for the expert validation of the survey instrument. Through a literature review, the expectations from the expert review for validation are identified. The domain of the construct, relevance, accuracy, inclusion or deletion of items, sensitivity, bias, structural aspects such as language issues, double-barreled, negative, confusing and leading questions need to be validated by the experts. A format is designed with a covering page having an invitation to the experts, their role, introduction to the research and the instrument. Information regarding the scale and the list of the scale item are provided in the subsequent pages. The demography questions are also included for validation. The expert review format will provide standard communication and feedback between the researcher and the expert reviewer that can help in developing a rigorous and quality survey instruments.

### **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Acknowledgments**

[OPTIONAL. This is where you can acknowledge colleagues who have helped you that are not listed as co-authors, and funding. MethodsX is a community effort, by researchers for researchers. We highly appreciate the work not only of authors submitting, but also of the reviewers who provide valuable input to each submission. We therefore publish a standard "thank you" note in each of the articles to acknowledge the efforts made by the respective reviewers.]



## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.mex.2021.101326](https://doi.org/10.1016/j.mex.2021.101326).

## References

- [1] American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education Standards for Educational & Psychological Testing, American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education, Washington, DC, 2014 Author.
- [2] H. Akkerboom, F. Dehue, The Dutch model of data collection development for official surveys, *Int. J. Public Opin. Res.* 9 (2) (1997) 126–145.
- [3] G.O. Boateng, T.B. Neilands, E.A. Frongillo, H.R. Melgar-Quiñonez, S.L. Young, Best practices for developing and validating scales for health, social, and behavioral research: a primer, *Front. Public Health* 6 (2018) 149.
- [4] K.A. Bollen, Latent variables in psychology and the social sciences, *Annu. Rev. Psychol.* 53 (1) (2002) 605–634.
- [5] P.M. Boynton, T. Greenhalgh, Selecting, designing, and developing your questionnaire, *BMJ* 328 (7451) (2004) 1312–1315.
- [6] S. Carpenter, Ten steps in scale development and reporting: a guide for researchers, *Commun. Methods Meas.* 12 (1) (2018) 25–44.
- [7] G.A. Churchill Jr., A paradigm for developing better measures of marketing constructs, *J. Mark. Res.* 16 (1) (1979) 64–73.
- [8] Collingridge, D. (2014, September 22). Validating a Questionnaire. Retrieved from <https://www.methodsplace.com/validating-a-questionnaire/>.
- [9] R.F. DeVellis, *Scale development: Theory and applications*, Vol. 26, Sage publications, 2016.
- [10] D.M. Dimitrov, *Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields*, John Wiley & Sons, 2014.
- [11] Forsyth, B., Lessler, J.T., & Hubbard, M. (1992). Cognitive evaluation of the questionnaire. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No.(ADM), 92-1929.
- [12] F.J. Fowler, A.M. Roman, *A Study of Approaches to Survey Question Evaluation*, Center for Survey Research, University of Massachusetts, 1992.
- [13] R.H. Gault, A history of the questionnaire method of research in psychology, *Res. Psychol.* 14 (3) (1907) 366–383.
- [14] B. Gillham, *Developing a Questionnaire*, A&C Black, 2008.
- [15] L. Guttman, A basis for scaling qualitative data, *Am. Sociol. Rev.* 9 (1944) 139–150.
- [16] T.R. Hinkin, A review of scale development practices in the study of organizations, *J. Manag.* 21 (5) (1995) 967–988.
- [17] T.R. Hinkin, J.B. Tracey, C.A. Enz, Scale construction: developing reliable and valid measurement instruments, *J. Hosp. Tour. Res.* 21 (1) (1997) 100–120.
- [18] R. Likert, A Technique for the Measurement of Attitudes, *Arch. Psychol.* 140 (1932) 1–55.
- [19] L.S. Martinez, Validity, Face and Content, in: M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods*, SAGE Publications, 2017, pp. 1823–1824.
- [20] S. Messick, Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meanings, *Am. Psychol.* 50 (9) (1995) 741.
- [21] P. Oosterveld, H.C. Vorst, N. Smits, Methods for questionnaire design: a taxonomy linking procedures to test goals, *Qual. Life Res.* 28 (9) (2019) 2501–2512.
- [22] T.D. Reeves, G. Marbach-Ad, Contemporary test validity in theory and practice: a primer for discipline-based education researchers, *CBE Life Sci. Educ.* 15 (1) (2016) rm1.
- [23] J.R. Rossiter, The C-OAR-SE procedure for scale development in marketing, *Int. J. Res. Mark.* 19 (4) (2002) 305–335.
- [24] S. Shaw, V. Crisp, Tracing the evolution of validity in educational measurement: past issues and contemporary challenges, *Res. Matters* 11 (2011) 14–19.
- [25] A.C.D. Souza, N.M.C. Alexandre, E.D.B. Guirardello, Psychometric properties in instruments evaluation of reliability and validity, *Epidemiol. E Serv. de Saúde* 26 (2017) 649–659.
- [26] S.S. Stevens, On the theory of scales of measurement, *Science* 103 (2684) (1946) 677–680.
- [27] L.L. Thurstone, Attitudes can be measured, *Am. J. Sociol.* 33 (1928) 529–554.
- [28] G.B. Willis, J.T. Lessler, *Question Appraisal System QAS-99*, National Cancer Institute, 1999.
- [29] R.L. Worthington, T.A. Whittaker, Scale development research: a content analysis and recommendations for best practices, *Couns. Psychol.* 34 (6) (2006) 806–838.
- [30] R.M. Guion, Content validity—the source of my discontent, *Appl. Psychol. Meas.* 1 (1) (1977) 1–10.