Content Validity of a Scale Measuring Psychosocial Stress Factors among Infertile Women in Treatment

Maya Rathnasabapathy[1], Deviga Subramani[2]

Abstract

Infertility is highly prevalent in women. While women are focusing on self-development and financial independence before starting a family and the age of marriage is on the rise resulting in infertility. It is vital to understand and explore the emotional difficulties women experience at this stage. Thus, the researcher conducted in-depth interviews with 20 infertile women from infertility clinics and, using related reviews, generated 135 items. On the basis of repetition and representativeness of the construct, the scale was reduced to 95 items which were validated. The content validity using both qualitative and quantitative methods ensured the representativeness and validity of the scale. The researcher procured expert opinions for each item of the scale, and content validity ratio (CVR) and content validity index (CVI) was obtained. The scale was given to 11 experts representing Gynecology and Psychology; and was requested to comment on the items on the basis of "Relevance," "Clarity," and "Necessity." Thus, the experts had to check if the items represented the psychosocial sources of stress infertile women experienced during treatment and they also were pertaining to the dimensions. The CVR, CVI, and Kappa values were calculated. The items with CVR of 0.75 and above were retained, with 0.50 and below eliminated, and between 0.75 and 0.50 were modified and retained. The CVI for Relevance and Clarity indicates that 75 of the items are "appropriate" and five items required revision; 74 of the items are "Clear," and six items required revision. Thus, the number of items was 80 after content validity. The Fleiss's Kappa value of 1 for "Necessity," "Relevance," and "Clarity" indicates that the inter-rater agreement of the scale is "almost perfect agreement."

Keywords: Content validity, Infertility, Psychosocial, Stress.

International Journal of Infertility and Fetal Medicine (2022): 10.5005/jp-journals-10016-1285

Introduction

Infertility

In the current era, several women are struggling with infertility. While marriage represents hope, happiness, and prosperity, delay in pregnancy represents pain and undeserving of happiness for many couples. The one irrational belief that delay in pregnancy means undeserving of happiness by itself causes stress and depression. Studies have found that several such beliefs are prevalent among women in India. Studies both in India and in other countries have proven that infertile women and men experience anxiety and depression.

The process of discovering that a woman or her husband is infertile is emotionally very painful. Accepting the same becomes a huge challenge complicated further by beliefs, demands by self, close family members, and society. The process of identifying the cause of infertility and the corrective procedure is very complex and physically draining. In this process, women of India are subjected to stress from various sources such as self, family, workplace, social, etc. It is also scientifically proven that psychological interventions help women face their challenges and overcome depression during infertility.[1–3]

As part of psychosocial care, identifying and assessing the mental health of women before, during, and after assisted reproductive treatment is suggested by the European Society for Human Reproduction and Embryology (ESHRE) in their psychosocial care guideline in 2015.[4] There are a number of assessment tools available to diagnose the mental health of infertile women worldwide such as Beck Depression Inventory (BDI), concerns of women undergoing assisted reproductive technologies (CART), fertility quality of life (FertiQoL), fertility status awareness scale

[1,2]School of Social Sciences and Languages, Vellore Institute of Technology, Chennai Campus, Chennai, Tamil Nadu, India

Corresponding Author: Maya Rathnasabapathy, School of Social Sciences and Languages, Vellore Institute of Technology, Chennai Campus, Chennai, Tamil Nadu, India, Phone: +91 9444333030, e-mail: maya.r@vit.ac.in

(FertiSTAT), etc. The current study aims to establish the content validity of a scale constructed to measure the psychosocial stress of infertile women in treatment. The scale initially had 95 items. After content validity, the items were reduced to 80.

The dimensions of the questionnaire include the following aspects related to infertility such as emotional, cognitive (rational and irrational beliefs), behavioral, relational, social, spiritual, and medical. The dimensions were adapted from the ESHRE, a guideline for routine psychosocial care in infertility and medically assisted reproduction.[4] The researcher included spiritual and medical/treatment aspects as they also play an important role in influencing the mental health of infertile women. The researcher approached infertile women in infertility clinics in Chennai and conducted in-depth interviews to gather information about their experiences related to the above mental health of infertile women. The researcher also reviewed related studies and pre-existing scales to draw up items for the above-mentioned dimensions.

Content Validity

During scale construction and validation, it is important to ensure that each item represents the mental health of infertile women in India. Face validity is a component of content validity. It refers to the degree to which respondents or users judge that the items of an assessment instrument are appropriate to the targeted construct and assessment objectives.[5–7] It is commonly thought to measure the acceptability of the assessment instrument to users and administrators.

Content validity is determined by the adequacy with which an observation instrument samples the behavioral domain of interest.[8] Content validity is defined as the extent to which an instrument adequately samples the research domain of interest when attempting to measure phenomena.[9,10] Three types of validity are considered when using a test or questionnaire to measure an individual's knowledge or attitudes. These are content validity, criterion validity, and construct validity.

Content validation provides evidence about the construct validity of an assessment instrument.[6] Construct validity is the degree to which an assessment instrument measures the targeted construct (i.e., the degree to which variance in obtained measures from an assessment instrument is consistent with predictions from the construct targeted by the instrument). Most targets of measurement in psychological assessment, regardless of their level of specificity, are constructs in that they are theoretically defined attributes or dimensions of people.[11] Criterion validity concerns the relationship of a test to specified criteria and is composed of predictive and concurrent validity.

In psychological assessment, the importance of content validation for the validation of the target construct varies depending on how precisely the construct is defined and the degree to which "experts" agree about the domain and facets of the construct. Content validation is particularly challenging for constructs with fuzzy definitional boundaries or inconsistent definitions.[12] Content validity also affects the latent factor structure of an assessment instrument. Content validity is important for any aggregated measure derived from an assessment instrument (e.g., factor or scale score, summary score, or composite score).

Qualitative and Quantitative Method in Content Validity

Content validity can be done in two possible ways. Studies in the field of psychology usually use experts' comments, feedback, and suggestions in ensuring the representability of the items to the construct of the scale. Scientists such as Lawshe[13] and Lynn[14] from the medical and nursing disciplines suggested methods to make the content validity process scientific or empirical by quantifying the aspects of content validity.

In the qualitative content validity method, content experts' and target groups' recommendations are adopted on observing grammar, using appropriate contextual words, applying correct and proper order of words in items, and appropriate scoring.[15] In the quantitative content validity method, confidence is maintained in selecting the most important and correct content in an instrument, which is quantified by the CVR.

Content Validity Ratio

The experts are requested to specify whether an item is necessary for operating a construct in a set of items or not. They are requested to score each item from 1–3 with a three-degree range of "not necessary, useful but not essential, essential," respectively.

CVR varies between 1 and −1. A higher score indicates further agreement of members of the panel on the necessity of an item in an instrument. The formula of CVR is $CVR = (Ne − N/2)/(N/2)$, in which Ne is the number of panelists indicating "essential" and N is the total number of panelists. The numeric value of CVR is determined using Lawshe Table. For example, if the number of panelists is 15 members, and if CVR is bigger than 0.49, the item is accepted.[16]

Content Validity Index

In reports of instrument development, the most widely reported approach for content validity is the CVI.[14,17,18] The CVI, a proportion agreement procedure, allows two or more raters to independently review and evaluate the relevance of a sample of items to the domain of content represented in an instrument. Panel members rate instrument items in terms of "Clarity" and "Relevance" as per the theoretical definition of the construct on a 4-point ordinal scale (1) not relevant, (2) somewhat relevant, (3) quite relevant, and (4) highly relevant).

A researcher then tallies the proportion of cases in which the raters agree and determines the stability of their agreement.[14] Researchers are instructed to collapse four ordinal response rankings into two dichotomous categories of responses ("content invalid" and "content valid"), and the CVI becomes a two-category nominal scale.[10,14,19] Davis[17] recommends a CVI of 0.80 for new measures.

Number of Experts

This step entails confirmation by a specific number of experts, indicating that instrument items and the entire instrument have content validity. For this purpose, an expert panel is appointed. Determining the number of experts has always been partly arbitrary. Guion,[20] Hambleton & Rogers,[21] Lawshe,[16] Lynn,[14] and Tittle[22] recommend the use of multiple judges for content validity and quantifying judgments using formalized scaling procedures. At least five people are recommended to have sufficient control over the chance agreement. The maximum number of judges has not been determined yet; however, it is unlikely that >10 people will be used. Anyway, as the number of experts increases, the probability of chance agreement decreases. After determining an expert panel, we can collect and analyze their quantitative and qualitative viewpoints on the relevancy or representativeness, clarity, and comprehensiveness of the items to measure the construct operationally defined by these items to ensure the content validity of the instrument.[14,23,24]

The literature is diverse with respect to the number of content experts needed. Lynn[14] recommended a minimum of three. However, others have suggested a range of up to 20 experts.[10,25] As noted by Grant and Davis[18] the number of panel experts depends on the desired level of expertise and diversity of knowledge. We recommend using at least three experts for each group (professionals and lay experts) with a range of up to 10. This yields a sample size of six to 20. Using a larger number of experts may generate more information about the measure.

The article gives an overview of how the content validity process using both qualitative and quantitative approaches contributes to ensuring the representativeness of the items. The scale was developed using the data collected from in-depth interviews conducted with infertile women and a review of related literature. The researcher gave the 95 item scale to 11 experts requesting them to review the items and give their opinion. The experts represented all relevant fields related to psychology and infertility such as

gynecologists, nurses, psychologists, counselors, psychiatrists, academicians from English, social work, and psychology. A sample of all the items, either in printed form or through e-mails, was distributed to the experts.

The experts were given the following instructions: Respected madam/sir, The questionnaire is for infertile women who are undergoing treatment for infertility. The objective of the research is to study the psychosocial impact of infertility. The items are drawn using in-depth interviews and appropriate reviews representing the seven dimensions of psychosocial stress such as (1) Emotional, (2) Cognitive, (3) Behavioral, (4) Relational, (5) Social, (6) Spiritual, and (7) Medical. There are 15 items in the emotional dimension, 14 items in the cognitive dimension, 17 in the behavioral dimension, 20 in the relational dimension, eight in the social dimension, 9 in the spiritual dimension, and 12 in the medical dimension.

Dimensions

- Emotional dimension includes all the positive and negative feelings and emotions that women experience during their treatment.
- Cognitive dimension includes the rational and irrational beliefs that women have about infertility and the delay in pregnancy.
- Behavioral dimension aspect includes the activities of women that are unhealthy and unhelpful as well as helpful in dealing with infertility treatment.
- Relational dimension includes spouse and immediate family members' influence that contribute to stress and worry, or support in dealing with infertility treatment.
- Social dimension includes the neighborhood and society's expectations and pressure on the women dealing with infertility treatment.
- Spiritual dimension includes the beliefs about God and the almighty that helps or hinders infertility treatment.
- Medical dimension includes the difficulties the infertile women might experience during treatment.

I request you to validate the items on the basis of the following: (1) Relevance, (2) Clarity, and (3) Necessity. Your comments as part of the expert review process will be valuable.

- Relevance is demonstrated by an item's ability to represent the content domain as described in the theoretical definition. The ratings are from 1–4, in which 1 represents "Not relevant," 2 represents "somewhat relevant," 3 represents "Quite relevant," and 4 represents "Highly relevant."
- Clarity of an item is evaluated on the basis of how clearly an item is worded. The ratings are again from 1–4, in which 1 represents "Not clear," 2 represents "Item needs some revision," 3 represents "Clear but need minor revision," and 4 represents "Very clear."
- The third criterion is Necessity, in which the experts are requested to specify whether an item is necessary for operating a construct in a set of items or not. The rating is from 1–3, where 1 represents "Not necessary," 2 represents "Useful but not "Essential," and 3 represents "Essential."

The scale will begin with the following instructions "The statements given below relate to your emotional experience and what you think or believe of infertility". There are no right or wrong answers. Please read them carefully and enter the score that feels right for you. Please do not take too much time to think and respond. Choose whatever comes to your mind as soon as you read the statements.

The Experts' comments, CVR, and CVI were used to review and modify the items of the scale. The probability of chance occurrence and inter-rater agreement for Relevance, Clarity, and Necessity was calculated using SPSS.

The present paper demonstrates the process for content validity and its importance during the validation of an instrument. It is said that validation is a lengthy process, in the first step of which the content validity is studied and followed by reliability evaluation (through internal consistency and test-retest), construct validity (through factor analysis), and criterion-related validity.[18] Common limitation of content validity studies is that the experts' feedback is subjective, and the study is subjected to bias that may exist among the experts. The lesser the experts, the fewer could be the opinion, and similarly, a too large number of experts may also cause inconvenience during content validity. It is at the discretion of the researcher to decide the appropriate number of experts. When the content domain is not well identified, the possibility of the items not representing the construct is possible. In the current study, experts are asked to suggest modifications or other items for the instrument, which may help minimize this limitation.[26]

RESULTS

During analysis, CVR was used to determine whether the item was to be retained in the tool or eliminated. Lawshe[16] minimum values for CVR were used to determine the ratio for each item. Most of the items had CVR equal to and above 0.75, which is good and excellent. Nineteen items had CVR below 0.75, out of which 15 items were eliminated. The eliminated items are Item 6, Item 8, Item 11, Item 14, Item 16, Item 22, Item 27, Item 28, Item 29, Item 35, Item 38, Item 40, Item 57, Item 72, and Item 81. For example, Item 27, "I regret that I should have taken care of me as a teenager," Item 8, "I do no get bowed down when people ask me about," and Item 81, "I will be sent to hell if I don't get pregnant." Reasons for low ratings by the experts were that the items were similar, repetitive, too specific, and not representative of the construct.

Item 10, Item 12, Item 20, and Item 95 were below the minimum CVR value. Based on the Lawshe table, they had to be eliminated, but the researcher decided to retain the items because of their importance in defining the scale. The items were modified using CVI for Relevance, Clarity, and Expert comments. After the researcher's decision about retaining or eliminating the items, CVI for Relevance and Clarity was calculated for each item. The items were modified on the basis of 'Relevance' and 'Clarity' and Expert comments on every item.

One suggestion given by most of the experts for the questionnaire was to change items from statements with "blanks" to "complete" statements such that all of them follow a similar pattern such as choosing options. It was also suggested to change the rating scale from "occurrence" to "agreeableness." Hence, the rating was changed from "Always, Often, Sometimes, Rarely, and Never" to "Strongly Agree, Agree, Neither Agree nor Disagree, Disagree and Strongly Disagree" and the sentences were modified accordingly.

The CVI for Relevance[14] was 0.86, and above for most of the items which is interpreted as "appropriate." Items 44, 45, 47, 48, and 49 had CVI below 0.71 which is interpreted as "Need for Revision."

Similarly, the CVI for Clarity[14] was 0.86, and above for most of the items, which are interpreted as "Clear." Items 13, 21,25, 46, 50, and 53 had CVI below 0.71, which is interpreted as "Need for Revision."

The inter-rater reliability was calculated using Fleiss Kappa, which is an extension of Cohen's Kappa used to measure agreement

between three or more raters. The calculated Kappa value was 1 for "Relevance," "Clarity," and "Necessity," which is interpreted as "almost perfect (agreement)" The finding indicates that the inter-rater agreement between the expert's judgments is high and reliable.[27]

CONCLUSION

Validation of a tool is a lengthy process. It is time-consuming and ambiguous when we resort to only qualitative feedback and comments. Psychology is a subjective field, and one single concept can be described from various perspectives. Collaborating quantitative approach along with the qualitative approach helps refine the items. The article suggests that the content validity process can be objective and helps in improving the quality of the items on a scale. The researcher in the article explains in detail the process of content validity in her study. The items of the scale were 135 initially and were reduced to 80 items after content validity. The CVR was used to determine how many items were to be eliminated and retained, in which 15 items out of 95 were eliminated, and 4 items were retained in spite of a lesser ratio because of the importance of the items. The CVI for Relevance and Clarity and expert opinion were used to modify the items. In this way, the researcher used the process of content validity to have meaningful items that represent the constructs of the study.

REFERENCES

1. Wischmann T, Scherg H, Strowitzki T, et al. Psychosocial characteristics of women and men attending infertility counselling. Hum Reprod 2009;24(2):378–385. DOI: 10.1093/humrep/den401
2. Boivin J. Guidelines for counselling in infertility: outline version. Hum Reprod 2001;16(6):1301–1304. DOI: 10.1093/humrep/16.6.1301
3. Haica C. The effect of psychological intervention on infertile couples quality of life during ART medical treatment.2018;8.
4. Gameiro S, Boivin J, Dancet E, et al. ESHRE guideline: routine psychosocial care in infertility and medically assisted reproduction—a guide for fertility staff. Hum Reprod 2015;30(11):2476–2485. DOI: 10.1093/humrep/dev177
5. Allen MJ, Yen WM. Introduction to measurement theory. Belmont, CA: Wadsworth. Inc.[Context Link]. 1979.
6. Anastasi A, Urbina S. Psychological testing. New Delhi: Prentice-Hall of India; 2007.
7. Nevo B. Face validity revisited. J Educ Meas 1985;22(4):287–293. DOI: 10.1111/j.1745-3984.1985.tb01065.x
8. Cronbach LJ. Test validation. In R. Thorndike (Ed.), Educational Measurement (2nd edition) 1971;443.
9. Carmines EG, Zeller RA. Reliability and validity assessment. SAGE Open Med 1979. DOI: 10.4135/9781412985642
10. Waltz CF, Strickland O, Lenz ER. Measurement in nursing research. FA Davis 1991;19–41.
11. Suen HK, Ary D. Analyzing quantitative behavioral observation data. Psychology Press 2014;376. DOI: 10.4324/9781315801827
12. Murphy KR, Davidshofer C. Psychological testing: principles and applications. Pearson.
13. Lawshe CH. Inferences from personnel tests and their validity. Am J Appl Psychol 1985;70(1):237. DOI: 10.1037/0021-9010.70.1.237
14. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382–385.
15. Safikhani S, Sundaram M, Bao Y, et al. Qualitative assessment of the content validity of the dermatology life quality index in patients with moderate to severe psoriasis. J Dermatol Treat 2013;24(1):50–59. DOI: 10.3109/09546634.2011.631980
16. Lawshe C. A quantitative approach to content validity. Pers Psychol 1975;28(4):563–575.
17. Davis LL. Instrument review: getting the most from a panel of experts. Appl Nurs Res 1992;5(4):194–197. DOI: 10.1016/s0897-1897(05)80008-4
18. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health 1997;20(3):269–274.
19. Waltz CF, Bausell BR. Nursing research: design statistics and computer analysis. FA Davis 1981.
20. Guion RM. Content validity—The source of my discontent. Appl Psychol Meas 1977;1(1):1–10. DOI: 10.1177/014662167700100103
21. Hambleton RK, Rogers HJ. Advances in criterion–referenced measurement. Advances in Educational and Psychological Testing 1991:3-43. Springer, Dordrecht.
22. Tittle CK. Use of judgmental methods in item bias studies. Handbook of methods for detecting test bias 1982;1:31–63.
23. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. West J Nurs Res 2003;25(5):508–518. DOI: 10.1177/0193945903252998
24. Yaghmaie F. Content validity and its estimation. J Med Educ 2003;3(1):25–27.
25. Gable RK, Wolf MB. Instrument development in the affective domain: measuring attitudes and values in corporate and school settings.
26. Nunnally JC. Psychometric theory. McGraw-Hill 1994;3.
27. Brennan P, Hays B. Focus on psychometrics the kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. Res Nurs Health 1992;15(2):153–158. DOI: 10.1002/nur.4770150210