ABC of Content Validation and Content Validity Index Calculation

Muhamad Saiful Bahri Yusoff

Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, MALAYSIA

ABSTRACT

There are five sources of validity evidence that are content, response process, internal structure, relation to other variables, and consequences. Content validity is the extent of a measurement tool represents the measured construct and it is considered as an essential evidence to support the validity of a measurement tool such as a questionnaire for research. Since content validity is vital to ensure the overall validity, therefore content validation should be performed systematically based on the evidence and best practice. This paper describes a systematic approach to quantify content validity in the form of content validity index based on the evidence and best practice.

Keywords: *Content validity, Content validation, Content validity index, Questionnaire validation, Validity*

CORRESPONDING AUTHOR Associate Professor Dr. Muhamad Saiful Bahri Yusoff, Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia | E-mail: msaiful_bahri@usm.my

INTRODUCTION

There are five sources of validity evidence that are content, response process, internal structure, relation to other variables, and consequence (1). Content validity is defined as the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose (1, 2). An assessment instrument refers to the particular method of acquiring data in psychological assessment such as questionnaires. The elements of an assessment instrument refer to all aspects of the measurement process that can affect the data obtained such as questionnaire items, response formats and instructions. The construct refers to the concept, attribute, domain, or variable that is the target of measurement. The assessment

purpose refers to the expected functions of the measurement tool, for examples, the Medical Student Stressor Questionnaire (MSSQ) was developed to identify the sources of stress in medical students (3) and the Anatomy Education Environment Measurement Inventory (AEEMI) was developed to measure the anatomy educational environment in medical schools (4). The relevance of an assessment tool refers to the appropriateness of its elements for the targeted constructs and functions of assessment, while the representativeness of an assessment tool refers to the degree to which its elements proportional to the facets of the targeted construct (2). Despite the two aspects of content validity (i.e., relevant and representativeness of an assessment tool), the relevant of an assessment tool that was advocated by Davis (5) has been frequently used to measure

the content validity (6, 7). It is important to note that establishing the content validity is vital to support the validity of an assessment tool such as questionnaires, especially for research purpose. Haynes et al. (2) emphasised that, "Inferences from assessment instruments with unsatisfactory content validity will be suspect, even when other indices of validity are satisfactory." The content validity evidence can be represented by the content validity index (CVI) (5–8), for instances, several recent studies (4, 9–11) established the content validity using CVI to support the validity of an assessment tool. Based on the evidence, this paper describes the best practice to quantify content validity of an assessment tool using CVI.

CONTENT VALIDATION PROCEDURE

The following are the six steps of content validation:

(a) Preparing content validation form

(b) Selecting a review panel of experts

(c) Conducting content validation

(d) Reviewing domain and items

(e) Providing score on each item

(f) Calculating CVI

Each step will be elaborated in the subsequent subchapters.

Step 1: Preparing Content Validation Form

The first step of content validation is to prepare the content validation form to ensure the review panel of experts will have clear expectation and understanding about the task. An example for the instruction and rating scale is provided in Figure 1. The recommended rating scale of relevance (5–8) has been used for scoring individual items (Figure 2). It is recommended to provide the definition of domain to facilitate the scoring process by the experts – please refer to Figure 2 for an example.

VALIDATION OF ANATOMY EDUCATION ENVIRONMENT INVENTORY: A Content Validity Study

Dear Experts,

This inventory contains 11 domains and 129 items related to anatomy education environment. We need your expert judgement on the degree of relevant of each item to the measured domains. Your review should be based on the definition and relevant terminologies that are provided to you. Please be as objective and constructive as possible in your review and use the following rating scale:

Degree of relevance:

1 = the item is not relevant to the measured domain

2 = the item is somewhat relevant to the measured domain

3 = the item is quite relevant to the measured domain

4 = the item is highly relevant to the measured domain

Figure 1: An example of instruction and rating scale in the content validation form to the experts.

Figure 2: An example of layout for content validation form with domain, its definition and items represent (measure) the domain.

Step 2: Selecting a Review Panel of Experts

The selection of individual to review and critique an assessment tool (e.g., questionnaire) is usually based on the individual expertise with the topic to be studied. Table 1 summarises the recommended number of experts with its implication on the acceptable cut-off score of CVI.

It can be concurred that for content validation, the minimum acceptable expert number is two, however most of recommendations propose a minimum of six experts. Considering the recommendations (5–8) and the author's experience, the number of experts for content validation should be at least 6 and does not exceed 10.

Step 3: Conducting Content Validation

The content validation can be conducted through the face-to-face or non-face-to-face approach. For the face-to-face approach, an expert panel meeting is organised, and the researcher facilitates the content validation process through Step 4 to Step 5 (will be described later). For the non-face-to-face approach, usually an online content validation form is sent to the experts and clear instructions are provided (Figure 1) to facilitate the content validation process (Step 4 to Step 5). The most important factors need to be considered are cost, time and response rate. The cost and time might be the challenging factor to conduct the face-to-face approach because of difficulty to get all experts be together, but the response

Table 1: The number of experts and its implication on the acceptable cut-off score of CVI

Number of experts	Acceptable CVI values	Source of recommendation
Two experts	At least 0.80	Davis (1992)
Three to five experts	Should be 1	Polit & Beck (2006), Polit et al., (2007)
At least six experts	At least 0.83	Polit & Beck (2006), Polit et al., (2007)
Six to eight experts	At least 0.83	Lynn (1986)
At least nine experts	At least 0.78	Lynn (1986)

rate will be at the highest. The response rate and time might be the challenging factor for the non-face-to-face approach because of difficulty to get response on time and at risk of not getting response at all from the expert, however the cost saving is the biggest advantage. Nevertheless, based on the author's experience, the non-face-to-face approach is very efficient if a systematic follow-up is in place to improve the response rate and time.

Step 4: Reviewing Domain and Items

In the content validation form, the definition of domain and the items represent the domain are clearly provided to the experts as shown in Figure 2. The experts are requested to critically review the domain and its items before providing score on each item. The experts are encouraged to provide verbal comment or written comment to improve the relevance of items to the targeted domain. All comments are taken into consideration to refine the domain and its items.

Step 5: Providing Score on Each Item

Upon completion of reviewing domain and items, the experts are requested to provide score on each item independently based on the relevant scale (Figures 1 and 2). The experts are required to submit their responses to the researcher once they have completely provided the score on all items.

Step 6: Calculating CVI

There are two forms of CVI, in which CVI for item (I-CVI) and CVI for scale (S-CVI). Two methods for calculating S-CVI, in which the average of the I-CVI scores for all items on the scale (S-CVI/Ave) and the proportion of items on the scale that achieve a relevance scale of 3 or 4 by all experts (S-CVI/UA) (6). The definition and formula of the CVI indices are summarised in Table 2.

Prior to the calculation of CVI, the relevance rating must be recoded as 1 (relevance scale of 3 or 4) or 0 (relevance scale of 1 or 2) as shown in Table 3. To illustrate the calculation of different CVI indices, the relevance ratings on item scale by ten experts are provided in Table 3.

To illustrate the calculation for the CVI indices (refer to Table 2), the following are examples of calculation based on the data provided in Table 3.

Table 2: The definition and formula of I-CVI, S-CVI/Ave and S-CVI/UA

The CVI indices	Definition	Formula
I-CVI (item-level content validity index)	The proportion of content experts giving item a relevance rating of 3 or 4	**I-CVI** = (agreed item)/(number of expert)
S-CVI/Ave (scale-level content validity index based on the average method)	The average of the I-CVI scores for all items on the scale or the average of proportion relevance judged by all experts. The proportion relevant is the average of relevance rating by individual expert.	**S-CVI/Ave** = (sum of I-CVI scores)/(number of item) **S-CVI/Ave** = (sum of proportion relevance rating)/(number of expert)
S-CVI/UA (scale-level content validity index based on the universal agreement method)	The proportion of items on the scale that achieve a relevance scale of 3 or 4 by all experts. Universal agreement (UA) score is given as 1 when the item achieved 100% experts in agreement, otherwise the UA score is given as 0.	**S-CVI/UA** = (sum of UA scores)/(number of item)

Note: The definition and formula were based on the recommendations by Lynn (8), Davis (5), Polit & Beck (6) and Polit et al. (7)

Table 3: The relevance ratings on the item scale by ten experts

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10		Experts in Agreement	I-CVI	UA
Q1	1	1	1	1	1	1	1	1	1	1		10	1	1
Q2	1	0	1	1	1	1	1	1	1	1		9	0.9	0
Q3	0	0	0	0	0	0	0	0	0	0		0	0	0
Q4	1	1	1	1	1	1	1	1	1	1		10	1	1
Q5	1	1	1	1	1	1	1	1	1	1		10	1	1
Q6	1	1	1	1	1	1	1	1	1	1		10	1	1
Q7	1	1	1	1	1	1	1	1	1	1		10	1	1
Q8	1	1	1	1	1	1	1	1	1	1		10	1	1
Q9	1	1	1	1	1	1	1	1	1	1		10	1	1
Q10	1	1	1	1	1	1	1	1	1	1		10	1	1
Q11	1	1	1	1	1	1	1	1	1	1		10	1	1
Q12	1	1	1	1	1	1	1	1	1	1		10	1	1
												S-CVI/Ave	0.91	
Proportion relevance	0.92	0.83	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92		S-CVI/UA		0.83
					Average proportion of items judged as relevance across the ten experts					0.91				

i. **Experts in agreement**: just sum up the relevant rating provided by all experts for each item, for example, the experts in agreement for Q2 (1 + 0 + 1 +1 + 1 + 1 + 1+ 1 + 1 + 1) = 9

ii. **Universal agreement (UA)**: score '1' is assigned to the item that achieved 100% experts in agreement, for examples, Q1 obtained 1 because all the experts provided relevance rating of 1, while Q2 obtained 0 because not all the experts provided relevance rating of 1.

iii. **I-CVI**: the expert in agreement divided by the number of experts, for example I-CVI of Q2 is 9 divided by 10 experts that is equal to 0.9.

iv. **S-CVI/Ave** (based on I-CVI): the average of I-CVI scores across all items, for example the S-CVI/Ave [(10 + 9 + 0 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10)/12] is equal to 0.91.

v. **S-CVI/Ave** (based on proportion relevance): the average of proportion relevance scores across all experts, for example the S-CVI/Ave [(0.92 + 0.83 + 0.92 + 0.92 + 0.92 + 0.92 + 0.92 + 0.92 + 0.92 + 0.92)/10] is equal to 0.91.

vi. **S-CVI/UA**: the average of UA scores across all items, for example the S-CVI/UA [(1 + 0 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1)/12] is equal to 0.83.

Based on the above calculation, we can conclude that I-CVI, S-CVI/Ave and S-CVI/UA meet satisfactory level, and thus the scale of questionnaire has achieved satisfactory level of content validity. For more examples on how to report the content validity index, please refer to papers written by Hadie et al. (4), Ozair et al. (9), Lau et al. (10) and Marzuki et al. (11).

CONCLUSION

Content validity is vital to ensure the overall validity of an assessment, therefore a systematic approach for content validation should be done based on the evidence and best practice. This paper has provided a systematic and evidence-based approach to conduct a proper content validation.

REFERENCES

1. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. The American Journal of Medicine. 2006;119(2):166.e7–16. https://doi.org/10.1016/j.amjmed.2005.10.036

2. Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. Psychological Assessment. 1995;7(3):238. https://doi.org/10.1037//1040-3590.7.3.238

3. Yusoff MSB. A systematic review on validity evidence of medical student stressor questionnaire. Education in Medicine Journal. 2017;9(1):1–16. https://doi.org/10.21315/eimj2017.9.1.1

4. Hadie SNH, Hassan A, Ismail ZIM, Asari MA, Khan AA, Kasim F, et al. Anatomy education environment measurement inventory: a valid tool to measure the anatomy learning environment. Anatomical Sciences Education. 2017;10(5):423–32. https://doi.org/10.1002/ase.1683

5. Davis LL. Instrument review: getting the most from a panel of experts. Applied Nursing Research. 1992;5(4):194–7. https://doi.org/10.1016/s0897-1897(05)80008-4

6. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health. 2006;29(5):489–97. https://doi.org/10.1002/nur.20147

7. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 2007;30(4):459–67. https://doi.org/10.1002/nur.20199

8. Lynn MR. Determination and quantification of content validity. Nursing Research. 1986;35(6):381–5.

9. Ozair MM, Baharuddin KA, Mohamed SA, Esa W, Yusoff MSB. Development and validation of the knowledge and clinical reasoning of acute asthma management in emergency department (K-CRAMED). Education in Medicine Journal. 2017;9(2):1–17. https://doi.org/10.21315/eimj2017.9.2.1

10. Lau AS, Yusoff MS, Lee Y-Y, Choi S-B, Xiao J-Z, Liong M-T. Development and validation of a Chinese translated questionnaire: a single simultaneous tool for assessing gastrointestinal and upper respiratory tract related illnesses in pre-school children. Journal of Taibah University Medical Sciences. 2018;13(2):135–41. https://doi.org/10.1016/j.jtumed.2017.11.003

11. Marzuki MFM, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps. JMIR Human Factors. 2018;5(2):e10308. https://doi.org/10.2196/preprints.10308