ESTEEM Journal of Social Sciences and Humanities
Vol. 6, No. 2, 18-29, September 2022
https://ejssh.uitm.edu.my

Development of Mathematical Thinking Test: Content Validity Process

Nurul Akmal Md Nasir[1], Parmjit Singh[2], Geethanjali Narayanan[3], Aisyah Hani Mohd Habali[4*] and Nor Syazwani Rasid[5]

[1,2,3,5]Faculty of Education, Universiti Teknologi MARA, Cawangan Selangor, Kampus Puncak Alam, 42300 Puncak Alam, Selangor, Malaysia
[4]Centre of Foundation Studies, Universiti Teknologi MARA, Cawangan Selangor, Kampus Dengkil, 43800 Dengkil, Selangor, Malaysia*
**Corresponding author: [4]aisyahhani6629@uitm.edu.my*

ABSTRACT

ARTICLE HISTORY

Received:
27 June 2022
Accepted:
28 July 2022
Published:
30 September 2022

KEYWORDS
Content validity
Content validation
Content validity index
Content validity ratio
Mathematical thinking test.

Validity is the extent to which an instrument measures what it is to measure. The instrument's validity is assessed by confirming the instrument's capability to measure the intended measurement. Content validity, face validity, construct validity, and criterion validity are the four forms of validity that are often discussed. Content validity is the degree to which a measurement tool reflects the measured content, which will justify the tool's validity. The importance of content validity in the instrument has made it an essential step in instrument development. Content validity is a vital criterion in designing and ensuring instrument development's success. Since content validity is essential in ensuring overall validity, content validation should be carried out in a structured manner based on evidence and best practice. This paper outlines a methodical strategy and a detailed content validation process for developing a mathematical thinking test. Seven evaluators were selected to review two sets of the mathematical thinking test. Three content validity measurements are applied to assess the items' validity in the tests: content validity index (CVI), kappa statistic and content validity ratio (CVR). Based on the results, the item content validity index (I-CVI) ranged from 0.857 to 1, and the scale content validity index (S-CVI/Ave) was 0.957. The Kappa statistics were also excellent since items ranged from 0.849 to 1.000. Content validity ratio (CVR) scores revealed that all items are the utmost critical by the content experts. Findings of the content validity study showed that all items in both mathematical thinking tests were deemed appropriate for assessing students' mathematical thinking.

e-ISSN 2600-7274

1. INTRODUCTION

Various evaluation instruments and methods are available to test students' learning of mathematics in today's curriculum reformation (Hatfield et al., 2003). Assessment plays a role as a tool to reveal all the potential possessed by students (Sabri et al., 2019). A vital factor of a successful curriculum is based on the assessment (Drijvers et al., 2019). Standard student assessment methods were more concerned with correct or incorrect scores than with students' mathematical understanding (Rosli et al., 2013; Berenson & Carter, 1995). Multiple-choice, true or false, or recall questions, ironically used to measure the correctness of a student's knowledge base, were the traditional mathematics questions in examinations, primarily used to examine the accuracy of a student's knowledge base (Watt, 2005). According to Lake (2015), these questions are notoriously unreliable for student assessment but used outside of an exam context. Consequently, many students memorised the procedure and neglected an in-depth understanding of the mathematical concepts and thinking processes when solving mathematics problems. Worrying trends proven among current students determine over-reliance on algorithms and procedures. This phenomenon would discourage learners' intuitive knowledge in making sense of mathematics (Singh et al., 2021; Singh et al., 2016). The ideal assessment should make the students reflect on their mathematical ideas and formulate ideas by assimilating their prior knowledge during involvement with the evaluation (Van de Walle et al., 2009). Students will actively engage in all mathematical standards processes while solving practical problems. They will create and reconstruct their knowledge through problem-solving, reasoning, communication, connections and representation (Van de Walle et al., 2009; NCTM, 2000).

To address this issue, there should be an instrument that can assess students' mathematical thinking. Mathematical thinking is beyond computational skills. It is the underpinning, fundamental conceptual thinking ability and how it relates to the real world. It relates to logical, analytic thinking and quantitative reasoning (Devlin, 2012). Therefore, the instrument for assessing students' mathematical thinking must emphasise procedural knowledge and include a variety of mathematical processes when solving tasks or problems. In relation to this, this study proposes constructing a mathematical thinking test instrument which can assess the student's ability in mathematical thinking. The validity processes are conducted to ensure the proposed instruments are relevant as assessment tools to measure students' ability in mathematical thinking.

Validity is the vital process that refers to the extent to which the instrument measures what it is designed to measure. Validity is a necessary procedure that relates to how well an instrument measures what it is supposed to measure. Validity is the correlation value between measurement and the actual value of a variable. If a size accurately reflects the variable's actual value, the correlation value will be high, and the research will have high validity (Chua, 2012). There are several different types of validity, such as face validity, content validity, criterion validity, construct validity, internal validity and external validity. Content validity is the degree to which a measurement tool accurately captures the construct being assessed, and it is seen as crucial proof that a measurement tool is legitimate (Yusoff, 2019). In contrast to other types of validity, content validity refers to test-based validity rather than score-based validity. It outlines the content requirements for the test and has nothing to do with the scores that were attained using that construct (Almanasreh et al., 2019). Content validity is defined as the degree to which elements of an assessment instrument are applicable to and reflective of the targeted construct for a given assessment purpose (Cook & Beckman, 2006; Haynes et al., 1995). This procedure entails a group of subject matter experts weighing in on the significance of various items inside an instrument (Ayre & Scally, 2014).

The development of the instrument itself begins the process of content validity. This process involves two phases which are instrument development and judgment. Throughout the instrumentation process, the two phases of procedures ensure that content validity is determined and quantified. This article attempts to assess the content validity of the proposed items consisting of the instrumentations. The context of mathematical thinking in this study refers to students' ability (cognitive) to solve non-routine problems. This article divides the two phases of the validation process (development and judgment) into five main steps. Two sets of mathematical thinking tests have been developed and proposed for the validation process. The construct items are generated through various sources and literature. The content validity of the generated items is next conducted with the aid of seven topic experts. The content validity index (CVI), Kappa statistics, and the content validity ratio (CVR; Lawshe test) quantify content validity. Therefore, the objective of the paper is to develop and evaluate the content validity of mathematical thinking tests using the Content Validity Index (CVI) and Kappa coefficient and Content Validity Ratio (CVR) for the two sets of mathematical thinking tests. This paper will describe in detail the best practice to quantify the content validity of an assessment tool using CVI, Kappa coefficient and CVR.

2. CONTENT VALIDATION PROCEDURE

The following are the five steps of content validation applied in this study:

a) Creating content validation form
b) Choosing expert review panels
c) Conducting content validation
d) Reviewing and Scoring Items
e) Calculating content validity

2.1 Step 1: Creating Content Validation Form

The first step of content validation is preparing a content validation form. The role of this form is to ensure that the review panel of experts will have clear expectations and understanding of the area to measure. Due to that, detailed descriptions regarding the items involved and what to measure are explained in the initial stage of the form (Figure 1). The experts also are provided with the definition of the topics and areas covered to facilitate them in the scoring process (see Figure 2 for an example).

VALIDATION OF MATHEMATICAL THINKING TESTS: A Content Validity

Dear Experts,

Background Information:

The purpose of this study is to investigate the effects of the problem-solving approach on the development of students' mathematical thinking. The context of mathematical thinking in this study refers to students' ability (cognitive) in mathematical thinking before (pre) and after (post) the intervention of the problem-solving approach. The instrument of mathematical thinking pre-post tests will be used to assess students' ability in mathematical thinking before and after the intervention process. Mathematical Thinking Test will be used to assess the effect of problem-solving approach (intervention) on participants' mathematical thinking ability. The first set of test (pre-test) will be conducted at the beginning of the study and second set (post-test), after ten weeks of problem-solving approach intervention. The duration of each test (pre-test and post-test) is 60 minutes (1 hour).

Participants:

Thirty first-year undergraduate students from B. Sc. Mathematic Education program will participate in this study. The test is expected to be conducted at the end of March and early June 2021.

Figure 1 An example of an overall description of the items in the content validation form to the experts

Section B: Evaluation of the Mathematical Thinking Pre-Test

This is an evaluation form to determine how valid the pre-test of students' mathematical thinking is in solving non-routine problems. Please read through the test items (Appendix I) carefully before rating (✓) the items below.

Rating Scale:

1 : No relevance at all
2 : The item needs some revision
3 : Relevant but needs minor revision
4 : Very relevant

a) Item Evaluation

Item	Topic	Solution (Heuristic Suggestion)	Rating				Comments
			1	**2**	**3**	**4**	
1	Numbers & Arithmetic	Pathway Heuristics (Working backwards)					
2	Numbers & Arithmetic	Simplification Heuristics (Simplify the problem)					
3	Number & Arithmetic	Generic Heuristics (Make a systematic list)					

Figure 2 An example of the content validation form layout where items represent (measure) the domain

2.2 Step 2: Choosing Expert Review Panels

Individuals who are chosen to evaluate and criticise an evaluation method are typically chosen based on their knowledge of the subject under investigation (Yusoff, 2019). Considering this factor, the reviewers are selected based on their expertise and experience in mathematical thinking, mathematics education and problem-solving. In content validity determination, determining the number of experts needed has always been somewhat arbitrary. The number of experts is often determined by how many open and agreeable people the instrument's creator or user may find (Lynn, 1986). Although this practice is widespread, specific guidelines should be considered when selecting experts for content validity determination. Lynn (1986) suggested having a minimum of five experts based on past studies as they would be at least able to provide a sufficient level of control for chance agreement. However, researchers suggested there should be at least six content-validation experts and ideally not more than ten (Davis, 1992; Polit et al., 2006; Lynn, 1986). Therefore, considering this recommendation, this study has used seven experts as reviewers. Table 1 shows the position and background of the selected experts.

Table 1 List of Expert

No	Position	Country
1	Associate Professor	Malaysia
2	Associate Professor	Malaysia
3	Senior Lecturer	United Kingdom
4	Associate Professor	Malaysia
5	Professor	Indonesia
6	Senior Lecturer	Malaysia
7	Senior Lecturer	Australia

2.3 Step 3: Conducting Content Validation

The content validation is conducted non-face-to-face. An online content validation form is sent to the experts, and a clear description (Figure 1) and instruction (Figure 2) are provided to facilitate the content validation process (Step 4). The response rate and time might be challenging due to difficulty in getting a response on time and the risk of not getting a response from the experts.

2.4 Step 4: Reviewing and Scoring Items

In this stage, experts are required to objectively analyse the item at a point before assigning a score to each one. Experts are encouraged to submit written feedback or suggestions to enhance the items' relevance in assessing students' mathematical thinking. All suggestions are taken into account when refining the items.

Upon completion of reviewing the items, all experts are requested to provide a score on each item independently based on the relevant scale, which is promoted by Davis (1992) and appears to be frequently employed: 1- not relevant at all; 2-somewhat relevant (the item needs some revision); 3-quite relevant (needs minor revision); and 4-very relevant. The experts must submit their responses to the researchers once they have completed the scores for all items. The reminder notification was sent to the experts if they took a long time to review the items.

2.5 Step 5: Calculating the Content Validity

The content validity index (CVI), Kappa statistics, and content validity ratio (CVR) were used to calculate the content validity.

2.5.1 Calculating Content Validity Index and Kappa

There are two forms of determining CVI for the item (I-CVI) and CVI for scale (S-CVI). I-CVI is the proportion of content experts giving the item a relevant or very relevant rating. The I-CVI is determined by dividing the number of experts who gave a “3” or “4” rating by the total number of experts, yielding the percentage of experts who agreed on the item's importance. Based on the expert rate, the score of 3 or 4 is given a new score of 1, which indicated relevance, while scores of 1 and 2 were categorised as irrelevant items and given a new score of 0. The S-CVI is the "proportion of items rated as quite relevant or extremely relevant by raters concerned" (Waltz, 2005). Two methods of assessing the value of S-CVI are the average of I-CVI scores for all items on the scale (S-CVI/Ave) and the proportion of items on the scale that are rated as "3-quite relevant" or "4-very relevant" by all experts (S-CVI/UA). The average of the I-CVI scores for all items on the scale, or the average proportion relevance determined by all experts, is known as S-CVI/Ave. In contrast, S-CVI/UA is defined as the per cent of items on the scale that all experts rate as 3 or 4 on the relevant scale. When all experts agreed on the item, it received a Universal Agreement (UA) score of 1; otherwise, it received a score of 0. The formulas of CVI indices are illustrated in Eq. 1, Eq.2 and Eq 3.

$$I-CVI = \frac{\text{Agreed item (score "3" or "4")}}{\text{Number of expert}} \tag{1}$$

$$S-CVI/Ave = \frac{\text{Sum of I-CVI}}{\text{Number of Items}} \quad OR$$
$$S-CVI/Ave = \frac{\text{Sum of proportion relevance rating}}{\text{Number of experts}} \tag{2}$$

$$S-SVI/UA = \frac{\text{Sum of UA scores}}{\text{Number of items}} \tag{3}$$

Even though many researchers utilise CVI to measure content validity, it still has limitations. CVI does not consider the risk of exaggerated values because of the chance agreement. As a result, using both CVI and Kappa statistics to determine the level of agreement between content experts could provide quantifiable procedures (Zamanzadeh et al., 2017; Brennan & Hays, 1992). The Kappa statistic is an inter-rater agreement consensus index that is used in conjunction with CVI to ensure that expert agreement is not due to chance. The value of the Kappa statistic is determined based on Eq. (4) and Eq. (5).

$$Pc = \left[\frac{N!}{A!(N-A)!}\right] \times 0.5^{N} \tag{4}$$

P_c = Probability of chance agreement

N = Number of experts

A = Number of expert who agreed

$$K = \frac{(I-CVI) - P_C}{1 - P_C} \tag{5}$$

2.5.2 Calculating Content Validity Ratio

The CVR method represents the value of the proportional level of expert agreement in rating an item as essential (Lawshe, 1975). CVR is computed using Lawshe Test. CVR's approach is to uphold whether an item is required to operate a construct in a set of items. The item was rated "1" as essential and "0" as not essential. The value of CVR is determined by using Eq. (6). The individual value of CVR is compared to the CVR critical table proposed by Ayre and Scally (2014) to determine the importance of each item in the instruments.

$$CVR = \frac{\left(N_e - \frac{N}{2}\right)}{\frac{N}{2}} \qquad (6)$$

N_e = Number of expert indicating "essential"

N = Total number of expert

3.0 FINDINGS

The key findings from both mathematical thinking tests (Set 1 and Set 2) were represented in terms of analysis content validity index (CVI), Kappa Coefficient and content validity ratio (CVR).

3.1 Determining Content Validity Index and Kappa Coefficient

Researchers commonly compute two types of CVIs (Lynn, 1986). The content validity of individual items (I-CVI) is the first type, while the content validity of the total scale is the second (S-CVI). Based on the value of I-CVI, the researchers have determined the score of the Kappa coefficient. Tables 2 and 3 indicate the CVI and Kappa coefficient evaluation for Mathematical Thinking Tests. The values of I-CVI for all items in both tests ranged from 0.875 to 1.000 and are classified as appropriate to assess students' mathematical thinking. Based on the assessments, the overall S-CVI for the 10-item scale was 0.957, indicating that the items had strong content validity for evaluating students' mathematical thinking. Apart from assessing the elimination of the items using CVI, the Kappa coefficient was used to evaluate all of the items. The findings (refer to Tables 2 and 3) revealed that all items are considered excellent (Polit & Beck, 2006; Zamanzadeh et al., 2014), with the scores of the Kappa coefficient ranging from 0.849 to 1.000.

Table 2 The Relevance Ratings on the Item by Experts for Mathematical Thinking Test 1 (CVI)

Item	E 1	E 2	E 3	E 4	E 5	E 6	E 7	Expert in Agreement	I-CVI	UA	Pc	Kappa
Item 1	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 2	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 3	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 4	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 5	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 6	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 7	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 8	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 9	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 10	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Proportion relevance	1	1	1	0.7	1	1	1	S-CVI/Ave (Method 1)	0.957	0.7		

Table 3 The Relevance Ratings on the Item by Experts for Mathematical Thinking Test 2 (CVI)

Item	E 1	E 2	E 3	E 4	E 5	E 6	E 7	Expert in Agreement	I-CVI	UA	Pc	Kappa
Item 1	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 2	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 3	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 4	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 5	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 6	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 7	1	1	1	0	1	1	1	6	0.857	0	0.0547	0.849
Item 8	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 9	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Item 10	1	1	1	1	1	1	1	7	1.000	1	0.0078	1.000
Proportion relevance	1	1	1	0.7	1	1	1	S-CVI/Ave (Method 1)	0.957	0.7		

3.2 Determining Content Validity Ratio

Table 4 and 5 show the evaluation of all experts to the items. Based on the calculation of CVR values, all things considered critical must be incorporated into the mathematical thinking test. All items are indicated range of 0.714 to 1.000.

Table 4 The Relevance Ratings on the Item by Experts for Mathematical Thinking Test 1 (CVR)

Item	E 1	E 2	E 3	E 4	E 5	E 6	E 7	Ne	CVR
Item 1	1	1	1	1	1	1	1	7	1.000
Item 2	1	1	1	1	1	1	1	7	1.000
Item 3	1	1	1	1	1	1	1	7	1.000
Item 4	1	1	1	0	1	1	1	6	0.714
Item 5	1	1	1	0	1	1	1	6	0.714
Item 6	1	1	1	1	1	1	1	7	1.000
Item 7	1	1	1	0	1	1	1	6	0.714
Item 8	1	1	1	1	1	1	1	7	1.000
Item 9	1	1	1	1	1	1	1	7	1.000
Item 10	1	1	1	1	1	1	1	7	1.000

Table 5 The Relevance Ratings on the Item by Experts for Mathematical Thinking Test 2 (CVR)

Item	E 1	E 2	E 3	E 4	E 5	E 6	E 7	Ne	CVR
Item 1	1	1	1	1	1	1	1	7	1.000
Item 2	1	1	1	1	1	1	1	7	1.000
Item 3	1	1	1	1	1	1	1	7	1.000
Item 4	1	1	1	0	1	1	1	6	0.714
Item 5	1	1	1	0	1	1	1	6	0.714
Item 6	1	1	1	1	1	1	1	7	1.000
Item 7	1	1	1	0	1	1	1	6	0.714
Item 8	1	1	1	1	1	1	1	7	1.000
Item 9	1	1	1	1	1	1	1	7	1.000
Item 10	1	1	1	1	1	1	1	7	1.000

4. DISCUSSION AND CONCLUSION

Content validity is a necessary next component of the research process to validate the research instrument. In this context, the instrument is a mathematical thinking test. Content validity of the instrument is a crucial way to pinpoint problem areas, reduce language error, determine the appropriateness aspect to be measured, and determine whether respondents are interpreting questions correctly by ensuring that the order of questions is not influencing the way a respondent answers and ensuring that the time given is suitable. Even though it is almost impossible to design a perfect instrument, many considerations still need to be highlighted to develop a good instrument.

Based on the results of I-CVI, each item in both tests is categorised as an appropriate item with scores between 0.857 to 1.000. Based on recommendations by Lynn (1986) and Polit and Beck (2006), with the number of experts from 6 to 8 people, the score of CVI must be at least 0.83. Considering the overall content validity of the item, the values of S-CVI/Ave also show very high validity with a score of 0.957. Past scholars stated that the value of S-CVI must be 0.8 or higher as an indication of acceptable validity (Almanasreh et al., 2019; Zamanzadeh et al., 2014;

Davis, 1992; Grant & Davis, 1997; Polit & Beck, 2004),. The content validity of the instruments is adequate, according to the S-CVI/Ave. The value of S-CVI for these instruments also could be considered at a congruity level (Waltz et al., 2005). According to Almanasreh et al. (2019), Polit and Beck (2006) and Zamanzadeh et al. (2014), values over 0.74, between 0.6 and 0.74, and between 0.4 and 0.59 are classified as excellent, good, and fair, respectively. Based on these categories, all items in these instruments have achieved excellent levels. High CVR scores suggest that members agree on the importance of a particular item in the instrument (Ayre & Scally, 2014). In this article, the CVR value for each item ranged from 0.714 to 1.000. These values reflected the high percentage of panellists rating an item as "essential" for the instruments.

In conclusion, the quantification of content validity using CVI (I-CVI & S-CVI), Kappa coefficient, and CVR revealed that the items in the instruments had excellent content validity. All the constructed items are maintained as they have a relatively high agreement value among experts. A future study can verify that the instrument's reliability is examined to improve the assessment instrument's application.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Education, Universiti Teknologi MARA for the unconditional support given while completing the study.

AUTHOR CONTRIBUTION STATEMENT

NAMN wrote the conceptualization and produced the original draft. PS and GN reviewed and edited the writing. NSR wrote the methodology section and contributed to the visualization. AHMH proofread and made changes to the writing drafts. All authors read and approved the final manuscript.

DECLARATION OF CONFLICTING INTERESTS

The authors declare that there is no conflict of interest.

FUNDING

The authors received no financial support for this article's research, authorship, and/or publication.

REFERENCES

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066.

Ayre, C. & Scally, A. J. (2014). Critical Values for Lawshe's Content Validity Ratio, *Measurement and Evaluation in Counseling and Development*, *47*(1), 79-86. DOI: 10.1177/0748175613513808.

Berenson, S. B., & Carter, G. S. (1995). Changing assessment practices in science and mathematics. *School Science and Mathematics*, *95*(4), 182-186. http://dx.doi.org/10.1111/j.1949-8594.1995.tb15759.x

Brennan, P. F & Hays, B. J. (1992). The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. *Research in Nursing & Health*, *15*(2), 153-158.

Chua, Y. P. (2012). *Mastering Research Methods*. Mc-Graw Hill (Malaysia) Sdn. Bhd.

Cook D. A. & Beckman T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine, 119*(2), 7–16. https://doi.org/10.1016/j.amjmed.2005.10.036

Cohen, I. S., & Fowler, J. V. (1998). Create assessments that do it all. *Mathematics Teaching in the Middle School*, *4*(1), 44-47.

Davis L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, 5(4), 194–197.

Drijvers, P., Kodde-Buitenhuis, H. & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathhematics, 102,* 435–456 https://doi.org/10.1007/s10649-019-09905-7.

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health, 20(3), 269–274.*

Hatfield, M. M., Edwards, N. T., Bitter, G. G., & Morrow, J. (2003). *Mathematics methods for elementary and middle school teachers* (4th ed.). John Wiley & Sons.

Haynes S.N., Richard D., & Kubany, E.S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*. 7(3):238. https://doi.org/10.1037//1040-3590.7.3.238

Lake, B (2015). Using True or False Questions for Student Engagement in Your Course. https://teachonline.asu.edu/2015/11/using-true-or-false-questions-in-your-course/

Lawshe, C. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-386

NCTM. (2000). *Principles and Standard for School Mathematics*.

Polit, D., & Beck, C. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique And Recommendations. *Research in Nursing & Health*, 29, 489-49

Polit, D., Beck, C. and Owen, S. V. (2007). Focus on Research Methods: Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing Health*, 30, 459–467

Rosli, R., Goldsby, D. & Capraro, M. M. (2013). Assessing Students' Mathematical Problem-Solving and Problem-Posing Skills. *Asian Social Science*, 9(16), 54-60.

Sabri, M., Retnawati, H., & Fitriatunisyah. (2019). The implementation of authentic assessment in mathematics learning. *Journal of Physics: Conference Series*, *1200*(1). https://doi.org/10.1088/1742-6596/1200/1/012006.

Singh, P., Han, C. T., Nasir, N. A. M., Ramly, M. A. Bin, & Hoon, T. S. (2016). Factors Contributing to Students' Poor Performance in a Mathematics Preparatory Program. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings*. https://doi.org/10.1007/978-981-287-664-5_28

Singh, P., Hoo, T. S., Nasir, N. A. M., Han, C. T, Rasid, N. S., Hoong, J. B. Z. (2021). Obstacles Faced by Students in Making Sense of Fractions. *The European Journal of Social and Behavioural Science, 1*, 3282-3299.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2009). *Elementary and middle school mathematics: Teaching developmentally* (7th ed.). Allyn & Bacon/Merill.

Waltz C. F., Strickland O. L. and Lenz E. R. (2005). *Measurement in Nursing and Health Research*. Springer Publishing Company.

Watt, H. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics, 58*(1), 21-44.
Yusoff, M. S. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, 49-54.
Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, *1*, 163-171.

AUTHOR BIOGRAPHIES

Nurul Akmal Md Nasir (M.Sc., B.Ed.) is a Senior Lecturer at Universiti Teknologi MARA Puncak Alam Campus, Selangor. He received a Master of Science (Mathematics) from Universiti Teknologi Malaysia, Skudai, Johor in 2012. Currently, he is a Ph.D. student in the Faculty of Education, Universiti Teknologi MARA in the area of mathematical thinking. He has authored and co-authored a few indexed journals which were indexed by Scopus and Springer and MyCite. His research interest includes mathematical thinking, problem-solving, cognition and metacognition of students at the school and university level.

Parmjit Singh (Ph.D., M.Ed., B.Sc.) is a professor of Mathematics Education at the Faculty of Education, Universiti Teknologi MARA Puncak Alam Campus, Selangor. He has co-authored over 80 publications including number sense, problem-solving, mathematical thinking, and assessments. The current research interests in Professor Maltech's group include (1) online assessment among tertiary-level students; (2) innovation in teaching and learning mathematics; and (3) mathematical thinking among undergraduate students.

Geethanjali Narayanan (Ph.D., M.Ed., B.Sc.) is a Senior Lecturer at Universiti Teknologi MARA Puncak Alam Campus, Selangor. She has published many papers in refereed journals and chapters in books, with areas of interest including Rasch Analysis, mathematics education and the development of teachers' pedagogy.

Aisyah Hani Mohd.Habali (M.Ed., B.Ed.) is a lecturer at Centre of Foundation Studies, Universiti Teknologi MARA, Cawangan Selangor, Kampus Dengkil. She received both her Bachelor and Master's degree of Teaching English Language as Second Language (TESL) from Universiti Teknologi MARA. The current research interests are applied linguistics, sociolinguistics, pedagogy and teaching writing.

Nor Syazwani Rasid (M.Sc., B.Ed.) is a Senior Lecturer at Universiti Teknologi MARA Puncak Alam Campus, Selangor. She received a Bachelor of Education Science (Mathematics) and for Master of Science (Mathematics) from Universiti Teknologi MARA and Universiti Teknologi Malaysia. She is one of the researchers who focuses on assessment in mathematics and innovation in teaching and learning mathematics.