KICSS2023

Kedah International Conference on Social Science and Humanities
UiTM Kedah (Online), Malaysia, 21-22 June 2023:
2nd International Conference on Business, Finance, Management and Economics
(BIZFAME)

Expert Evaluation on the Content Validity of the Novice Teachers' Assessment Inventory (InPGN)

Muslihaza Musikin[1], Mohd Effendi a.k.a. Ewan Mohd Matore[2*], Norazlinda Mohamad[3]
*Corresponding Author

[1] SMK Sg Ramal, Bangi, Malaysia
[2*] Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia
[3] Europasia Engineering Services Sdn. Bhd, Petaling Jaya, Malaysia

muslihazamusikin@gmail.com, effendi@ukm.edu.my

Abstract
This study aims to evaluate the validity of the Novice Teachers' Assessment Inventory (InPGN) utilizing expert consensus perspectives and the Content Validity Index (CVI). A total of 97 items involved nine experts from assessment literacy, psychometrics, and language field by using purposive sampling. Validity tests found that InPGN has a CVI value of 0.98, indicating a very high level of instrument suitability. An implication of this remains expected to provide an excellent measurement tool in helping novice teachers improve their professional development in a more authentic assessment dimension.

Keywords: content validity; assessment literacy; novice teacher; expert concensus

1.0 Introduction
Teacher assessment literacy is a professional requirement in formulating teacher competency standards across various aspects. It also influences educational policies in North America, Europe, Australia, and New Zealand that emphasize continuous assessment literacy (formative and summative). Malaysian context requires competent teachers to perform holistic assessment activities (Curriculum Development Division (BPK), 2018).

However, measuring existing teacher assessment literacy still needs more evidence in terms of test content reliability and validity, internal consistency dependability, score stability, and their correlations with student accomplishment (Gotch & French, 2014), generally to all teachers. The existing assessment literacy instrument has gone chiefly through various validity techniques but has yet to be discussed in expert-based assessment.

On a global basis, various tools have been created to assess teacher assessment literacy, such as the Teacher Assessment Literacy Questionnaire (TALQ), Measurement Literacy Questionnaire, Classroom Assessment Literacy Inventory (CALI), and Assessment Practices Inventory (API). These instruments were constructed based on the Standards for Teacher Competence in Educational Assessment of Students (STCEAS) developed by the American Federation of Teachers, the National Council on Measurement in Education, and the National Council on Measurement in Education in 1990. However, the planned measurement of teacher assessment literacy still needs more evidence regarding instrument content reliability and validity, internal consistency reliability, score stability, and their correlations with student accomplishment (Gotch & French, 2014).

Therefore, developing the Novice Teachers' Assessment Inventory (InPGN) as a competency measurement tool involving the concepts of knowledge, skills, practices, and integrity of assessment requires expert consensus to strengthen the assessment competency of novice teachers. InPGN adapted from the Teacher Assessment Inventory by Abu Mansor (2017) and Classroom Assessment Practices and Skills (CAPS) by Koloi-Keaikitse (2012) to measure the competency of novice teachers through four assessment dimensions, namely knowledge, skills, practice, and integrity. This article aims to assess the validity of the Novice Teachers'

Assessment Inventory (InPGN) through expert consensus perspectives using the Content Validity Index (CVI) method. The objective is explicitly to calculate the score of the InPGN items based on three aspects being evaluated from CVI: relevance, clarity, and simplicity.

2.0 Literature Review

2.1 Competency Assessment of Novice Teachers
Assessment literacy involves the competency of teachers to interpret assessment information critically and meaningfully (Hopfenbeck, 2019). Assessment activities carried out by teachers include planning and administering tests, building items, preparing the scoring rubrics, analyzing evidence, applying information, recording assessment data, and making a report (Deluca et al., 2016). All of these assessment activities require teachers who are competent in assessment literacy. The researchers focus on two primary components of assessment literacy: conceptual knowledge (Willis et al., 2013) and skills in applying assessment knowledge (Gotch & French, 2014).

The aspects of knowledge and skills in assessment literacy are highlighted by Stiggins (1991) as a competency component through the Iceberg Model. This model has been a guideline for constructing Teacher Competency Standards in Malaysia. In addition, various initiatives were also undertaken to strengthen the assessment concept among novice teachers (Pastore & Andrade, 2019). However, the effectiveness of assessment in the classroom remains inaccessible when novice teachers cannot adapt skills to the existing knowledge. Existing assessment-related tools, such as Suah's (2012) Teacher Assessment Practise Inventory, focus primarily on knowledge and assessment abilities. Hence, InPGN has been developed by adapting the existing instruments to meet the above requirements.

2.2 Content Validity in Assessment
There are various validity techniques used by past researchers to obtain quality and valid research instruments (Polit et al., 2007). Assessment through the validity index will assist in the process of InPGN validity because these criteria can verify the statistical relationship values between instrument items that suit the purpose of the instruments. Content validity assessment is the first step in the instrument creation process (Lynn, 1986). An expert panel assessment is required in this procedure to verify that the items are in line with the constructs studied (Kim et al., 2018). A panel of experts will assess the instruments based on the constructs represented in this study. Assessments will be carried out under four scales of agreement adapted from Yaghmaie (2003).

The construction of InPGN is a result of the adaptation of two assessment-related instruments, the Teacher Assessment Inventory by Abu Mansor (2017) and the Classroom Assessment Practices and Skills (CAPS) by Koloi-Keaitse (2012). Generally, the Teacher Assessment Inventory has undergone the overall content validity process. Regarding the Assessment Knowledge Test relative to the Teacher Assessment Inventory, past researchers have only conducted content validity through the Test Specification Table. In contrast, content validity was only acquired through a review by three experts from literacy, psychometrics, and language for the constructs of practice and assessment integrity. Past researchers have yet to perform any validity analysis to obtain expert consensus, although it is imperative to get quality measurement instruments in the context of instrument development (Polit et al., 2007). The Content Validity Index (CVI) is a popular validity method (Almanasreh et al., 2019) by past researchers (Aksah et al., 2023; Mohamad et al., 2022; Roebianto et al., 2023; Roy et al., 2023) to obtain validity analysis of the developed instruments. CVI can be analyzed through the formula CVI = n/N, where n is the number of several expert panel consensus, and N is the number of experts.

3.0 Methodology
A survey strategy was used in this investigation, which took a quantitative approach through questionnaires written and online (email) to measure the competency of novice teachers in Malaysia. A quantitative method was selected due to its ability to analyze numerical data using statistical procedures besides generalizing the findings (Creswell, 2014). Questionnaires were chosen as a research tool due to their ability to obtain standard data from a comprehensively large research sample and being the most simple and effective method (Fraenkel & Wallen, 2009). The respondents were selected using a purposive sampling technique, and they fulfilled the expert panel criteria outlined by the researcher.

The questionnaire instrument used in this study was adapted from the Teacher Assessment Inventory by Abu Mansor (2017) and Classroom Assessment Practices and Skills (CAPS) by Koloi-Keaikitse (2012). The questionnaires consist of two parts, namely Section A, which is related to the demographics of the expert panel, and Section B, which is related to validity analysis for the Novice Teachers' Assessment Inventory or *Inventori Pentaksiran Guru Novis* (InPGN). InPGN was modified according to the context of the study, comprising five sections with three demographic-related items in Section A, followed by 25 items in Section B to test the level of assessment knowledge, 29 items in Section C to measure assessment skills, 28 items in Section D to measure the level of assessment practices, and 15 items in Section E regarding the level of assessment integrity. Meanwhile, Section C was translated into Malay Language according to the suitability of the study location.

The Content Validity Index (CVI) approach with a four-point Likert scale was used to assess content validity to obtain the scale of consensus among experts towards InPGN. Lynn (1986) outlined two phases in instrument development: the i) development phase and the ii) judgment-quantifying phase.

InPGN consists of 3 sections: Section A about demographics, Section B tests the level of assessment knowledge, Section C measures assessment skills, Section D measures the level of assessment practices, and Section E measures the level of assessment integrity. The Section C construct was translated into Malay Language according to the suitability of the study location. For the second

phase – expert validation, the judgment-quantifying phase involves content validity for each item in InPGN. In this phase, a panel of experts will determine the level of construct relevance for each item used. InPGN was given to nine experts for assessment based on relevance, clarity, and simplicity, where the items would be assessed using a four-point ordinal Likert scale.

The expert panel selection is based on three criteria outlined by Davis (1992) and Grant and Davis (1997): academic qualifications, expertise, and experience in the respective fields. The InPGN expert panel comprises four experts from public universities in the field of assessment literacy, a professional practitioner from the Malaysian Examination Board, two experts in psychometrics, and two experts in language and linguistics. The expert panel assessed InPGN using the same validity form with a four-point ordinal Likert scale adapted from Yaghmaie (2003), as shown in Table 1.

Table 1. CVI Assessment Criteria based on a Four-Point Likert Scale

Criteria	Score	Interpretations
Relevance	1	Irrelevant.
	2	The item needs to be reviewed.
	3	Relevant but needs a slight revision.
	4	Very relevant.
Clarity	1	Unclear.
	2	The item needs to be reviewed.
	3	It's clear but needs a slight revision.
	4	Very clear.
Simplicity	1	Not simple.
	2	The item needs to be reviewed.
	3	It's simple but needs a slight revision.
	4	Very simple.

(Source:) Yaghmaie (2003).

The CVI for each component and the overall instrument is calculated using a percentage technique. CVI is determined for each item based on the number of experts that rated the item as relevant (scoring 3 or 4). As a result, the number of experts who rated the item as relevant was divided by the total number of experts (Polit & Beck, 2006). First, the total number of experts who scored 3 or 4 on the assessment criteria (relevance, clarity, and simplicity) would be determined. The proportion of each item with a CVI value between 0.78 and 1.00 determines the instrument's content validity. On the other hand, items with a CVI value of less than 0.78 should be reviewed (Polit et al., 2007).

4.0 Findings and Discussions

The findings indicated that all items have gone through the CVI score calculation process in terms of relevance, simplicity, and clarity. Eighty-one (81) of ninety-seven (97) items were validated for the simplicity criterion. Therefore, CVI = 81/97 = 0.84 denotes that 84% of the items yielded CVI = 0.78 for both requirements. Eight-six (86) out of ninety-seven (97) items were also relevant, with CVI = 86/97 = 0.89, representing 89% of the clarity criterion. The relevance criterion showed the highest validity value with CVI = 87/97 = 0.90, validating eighty-seven (87) out of ninety-seven (97) items. It suggests that 90% of the items were valid for the relevance criterion. CVI was calculated after I-CVI had been determined. CVI for the instrument denotes the number of items assessed as accurate, CVI = 0.98. The construct of the Assessment Knowledge Test consists of 25 objective multiple-choice items. Figure 1 shows a total of 21 items in the aspect of relevance recorded a CVI value of 1.
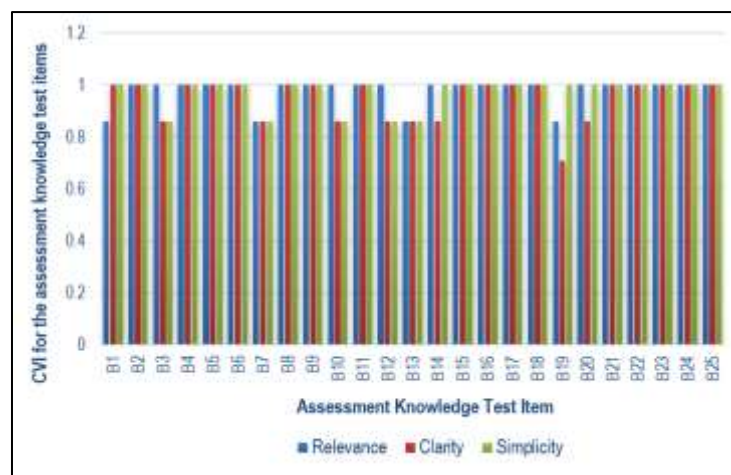

Fig. 1: CVI for the Assessment Knowledge Test items
(Source:) adapted from Abu Mansor (2017) and Koloi-Keaikitse (2012)

Only four items, B1, B7, B13, and B19, recorded a CVI value of 0.86. Meanwhile, in clarity, seven items were recorded to have a CVI value of 0.86, such as B3, B7, B10, B12, B13, B14, and B20, while B19 recorded a CVI value of 0.71. Finally, in simplicity, five

items recorded a CVI value of 0.86, such as B3, B7, B10, B12, and B13, while the rest recorded a CVI value of 1. One item recorded a CVI value of < 0.78, which is item B19 (CVI = 0.71) under the criterion of clarity. Item B19 is related to the knowledge of scoring and grading under the assessment knowledge construct that requires teachers to make decisions after conducting an item analysis. The expert panel questioned using "or" to remove and revise test items after performing the item analysis. In conclusion, all items for the Assessment Knowledge Test construct, except for B19, were accepted with a slight improvement as recommended by the expert panel. For the construct of assessment skills, a total of 29 items, as seen in Figure 2, were assessed by the expert panel based on three aspects of assessment.
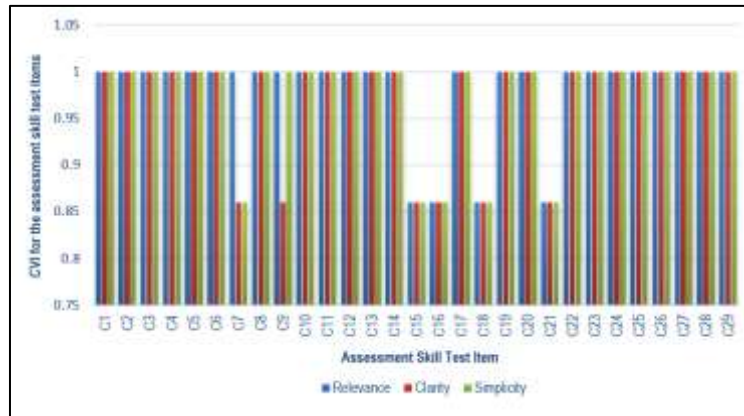

Fig. 2: CVI for the Assessment Skill Test items
(Source:) adapted from Abu Mansor (2017) and Koloi-Keaikitse (2012)

Regarding relevance, only four items, C15, C16, C18, and C21, recorded a CVI value of 0.86, while the rest recorded a CVI value of 1. Meanwhile, regarding clarity, six items recorded a CVI value of 0.86, such as C7, C9, C15, C16, C18, and C21. It shows that the items were accepted with a slight improvement. The final aspect assessed was simplicity, which recorded five items with a CVI value of 0.86, such as C7, C15, C16, C18, and C2, whereas the rest recorded a CVI value of 1. Overall, the expert consensus was obtained with a CVI value of > 0.78 for all 29 items in the three aspects of the expert validity assessment. Based on Figure 3, all assessment practice test items in the three assessment aspects recorded a CVI value of 1 except for item D28. In all assessment aspects, this item recorded a CVI value of 0.86. However, the item was accepted with improvement in the sentence structure. The item was stated as "stating assessment criteria" and had been modified to "stating assessment criteria in producing project work."
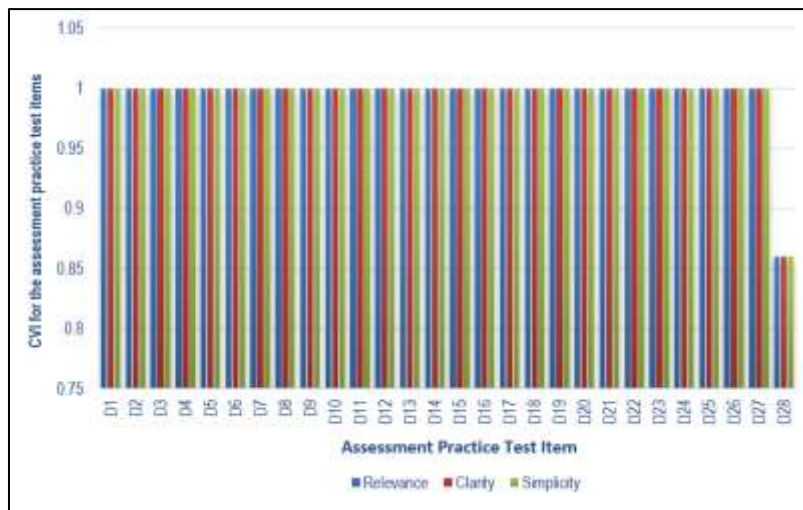

Fig. 3: CVI for the Assessment Practice Test items
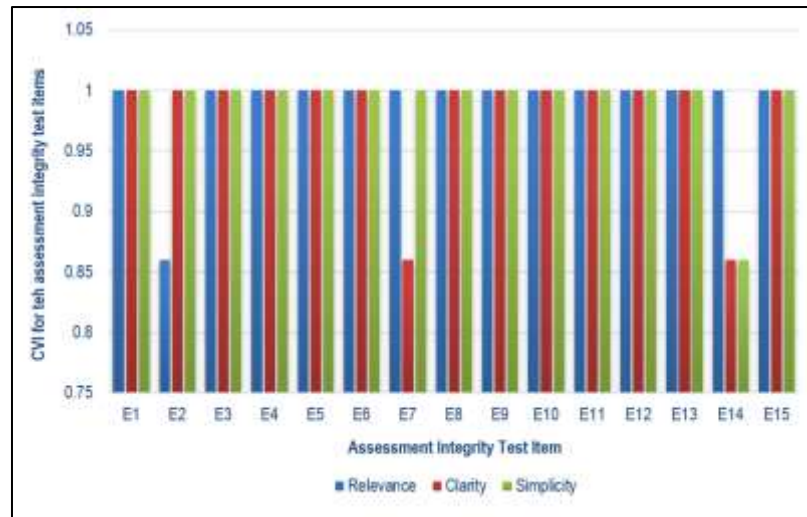*(Source:) adapted from Abu Mansor (2017) and Koloi-Keaikitse (2012)*

Fig. 4: CVI for the Assessment Integrity Test items
(Source:) adapted from Abu Mansor (2017) and Koloi-Keaikitse (2012)

Based on Figure 4, item E2 in the aspect of relevance, items E7 and E14 in the part of clarity, and item E14 in the element of simplicity all recorded a CVI value of 0.86. The remaining items recorded a CVI value of 1. Therefore, only a small number of items from the assessment integrity construct should improved according to the recommendations given by the expert panel. The researcher agrees with Polit dan Beck (2006) on consensus expectations, which cannot reach 100 percent considering the bias factors of the expert panel. The CVI value of the clarity criterion is high (CVI = 0.97), represented by 97% of the items assessed as "very clear" by the expert panel. As for the simplicity criterion's CVI value, all items showed a higher value (CVI = 0.98) than the clarity criterion. The result showed that 98% of the items were assessed as "very simple." The relevance criterion showed the highest CVI value (CVI = 0.99), represented by 99% of the overall items evaluated as "very relevant" by the expert panel. All items with a CVI value of 0.78 were accepted with a slight improvement as recommended by the expert panel.

The CVI value of InPGN has the highest score, with 0.98 (CVI = 0.78). It shows that InPGN has a very high level of item relevance through expert consensus. The number of experts involved also plays a role in determining the level of agreement on the instrument. However, some limitations have been identified in the implementation of CVI, such as a bias factor that may exist among the panel of experts. Experts should be selected more carefully to ensure that the validity assessment can be carried out by genuinely eligible and appropriate (Rubio et al., 2003).

5.0 Conclusion

Based on the findings of this study, some items need to be improved regarding several technical errors and sentence phrases. Competency measurement should be precise for each construct represented by the items to obtain a high-reliability value. The constructs of the adopted items have not only been valid for use in previous studies, but they are also appropriate in the context of assessment and the local culture of the research population. Finally, significant vital limitations need to be considered, such as the number of experts and the diversity of the experts. A variety of expert criteria can provide holistic results and may increase the number of experts to enhance the fairness of the judgment and avoid bias. Quantitative analysis findings for InPGN give only one type of evidence of content validity through CVI. Since content validity is among the critical stages in selecting the research instruments, additional analyses, such as content validity ratio and coefficient, are required. InPGN, which has been translated, modified, and gone through a very detailed content validity process, is capable of becoming a survey tool for future researchers, particularly in measuring the assessment competency of novice teachers.

Acknowledgments

Paper Contribution to Related Field of Study

This research extends our knowledge of research methods by using the Content Validity Index (CVI) and applying expert consensus for assessment literacy among novice teachers. This work contributes to the existing understanding of assessment literacy by providing solutions to the complexity of assessment by catering to ratings from different experts.

References

Abu Mansor, N. (2017). *Pengukuran pengetahuan, amalan dan integriti pentaksiran bilik darjah dalam kalangan guru Reka Bentuk dan Teknologi.* [Master's thesis, Universiti Teknologi Malaysia]. Universiti Teknologi Malaysia

Aksah, H., Che Ani, A. I., Johar, S., & Husain, S. H. (2023). Evaluating the Content Validity: Development of An Instrument for Measuring Functional Building Performance. *International Journal of Global Optimization and Its Application*, 2(1), 12–19. https://doi.org/10.56225/ijgoia.v2i1.161

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. http://dx.doi.org/10.1016/j.sapharm.2018.03.066.

Curriculum Development Division (BPK)(2018). *Panduan Pelaksanaan Pentaksiran Bilik Darjah. Kementerian Pendidikan Malaysia* (1st ed.). Curriculum Development Division. http://jpnperak.moe.gov.my/ppdbagandatuk/attachments/article/3555/Panduan Pelaksanaan Pentaksiran Bilik Darjah 2018.pdf.

Creswell, J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* (4th ed.). Thousand Oaks, CA.

Davis, L.L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, 5(4): 194–197. https://doi.org/10.1016/S0897-1897(05)80008-4

Deluca, C., Lapointe-mcewan, D. & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4): 248–266. https://doi.org/10.1080/10627197.2016.1236677

Distefano, M., Pryer, M., & Erffmeyer, R. (1983). Application of content validity methods to the development of a Job-Related Performance rating criterion. *Personnel Psychology*, 621–632. http://onlinelibrary.wiley.com/doi/10.1111/j.1744-6570.1983.tb02239.x/abstract

Fraenkel, J.R. & Wallen, N.E. (2009). *How to design and evaluate research in education.* (7th ed.). McGraw-Hill.

Gotch, C.M. & French, B.F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2): 14–18. https://doi.org/10.1111/emip.12030

Grant, J.S. & Davis, L.L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20(3): 269–274. https://doi.org/10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

Hopfenbeck, T.N. (2019). Assessment literacy across contexts and competencies. *Assessment in Education: Principles, Policy and Practice,* 26(2): 119–122. https://doi.org/10.1080/0969594X.2019.1601357. https://doi.org/10.1080/0969594X.2019.1601357

Koloi-Keaikitse, S. (2012). *Classroom assessment practices: A survey of Botswana primary and secondary school teachers.* [Doctoral dissertation, Ball State University]. https://cardinalscholar.bsu.edu/bitstream/handle/123456789/195866/Koloi-KeaikitseS_2012-2_BODY.pdf;sequence=1

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6): 382–386. https://doi.org/10.1097/00006199-198611000-00017

Mohamad Tarmizi, M. H., & Janan, D. (2022). Content Validity of The Reading Skills Teaching Module Through a Foreign Language Teaching Approach. *Malaysian Journal of Social Sciences and Humanities (MJSSH),* 7(12), e001990. https://doi.org/10.47405/mjssh.v7i12.1990

Pastore, S. & Andrade, H.L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84: 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Polit, D.F. & Beck, C.T. (2006). The content validity index: Are you sure you know what is being reported? Critique and recommendations. *Research in Nursing & Health*, 31(4): 341–354.

Polit, D.F., Beck, C.T. & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4): 459–467. https://doi.org/10.1002/nur.20199

Roebianto, A., Savitri, S. I., Aulia, I., Suciyana, A., & Mubarokah, L. (2023). Content Validity: Definition and Procedure of Content Validation in Psychological Research. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, 30(1), 5–18. https://doi.org/10.4473/TPM30.1.1

Roy, R., Sukumar, G. M., Philip, M., & Gopalakrishna, G. (2023). Face, content, criterion, and construct validity assessment of a newly developed tool to assess and classify work–related stress (TAWS– 16). *PLoS ONE*, 18(1 January), 1–11. https://doi.org/10.1371/journal.pone.0280189

Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S. & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Sosial Work Research*, 27(2): 94–104. https://doi.org/10.1093/swr/27.2.94

Stiggins, R.J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7): 534–539.

Suah See Ling. (2012). *Analisis Model Literasi dan amalan pentaksiran guru sekolah serta kajian tentang jurang antara keduanya.* [Doctoral thesis, Universiti Sains Malaysia]. School of Educational Studies, Universiti Sains Malaysia.

Willis, J., Adie, L. & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher*, 40(2): 241–256. https://doi.org/10.1007/s13384-013-0089-9

Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1): 25–27.