



Reliability and Validity of The Fuzzy Delphi Instrument of Islamic Cognitive Domains

Running Head: Reliability and Validity of Islamic Cognitive Domains

Hafizhah Zulkifli, Ahmad Munawar Ismail

Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.

Email: munawar@ukm.edu.my

Corresponding Author Email: hafizhah_zulkifli@ukm.edu.my

Nurul Asiah Fasehah Muhamad

Universiti Sains Islam Malaysia, Malaysia, Malaysia

Email: nurulasiahfasehah@usim.edu.my

Nur Hanani Hussin

Ministry of Education, Malaysia

Email: nurhanani.hussin@gmail.com

Sri Andayani Mahdi Yusuf

Sekolah Tinggi Agama Islam (STAI) Nusantara, Indonesia

Email: seridayani86@gmail.com

To Link this Article: <http://dx.doi.org/10.6007/IJARPED/v13-i2/21201>

DOI:10.6007/IJARPED/v13-i2/21201

Published Online: 14 May 2024

Abstract

This study aims to evaluate the content validity and reliability of the fuzzy Delphi instrument of Islamic Cognitive Domains. This study is based on the Development Design Research methods. The content validity of the instrument was analyzed using the Content Validity Index (CVI) by four experts. For the reliability, there were 30 respondents. The results of the analysis show that the overall value of content validity using CVI is 0.8. The value of Cronbach Alpha is 0.97. It means that the instrument has high validity and reliability.

Keywords: Content Validity Index, Fuzzy Delphi, Islamic Cognitive Domains, Instrument, Reliability, Validity

Introduction

Validity and Reliability of the instrument is essential in order to develop an instrument in achieving the appropriateness and accuracy of the items in the instrument. The instrument needs to be determined from various aspects, such as the theoretical and conceptual approach, to consider the desired data requirements collected before the validity test and

reliability are done (Omar et al., 2021). Not only that, the adjustment and construction of instrument items also need to be included in all constructs to be measured to help experts choose the best quality items (Salleh, 2018). If those aspects are met, then the research instrument will produce good results and meet the level of validity and reliability according to the appropriate value (Pallan, 2001)

Instrument validity refers to the ability of an instrument to measure an element that should be measured Clark & Watson (2019), while reliability is a process to test the consistency of the measured tool against the research instrument used (Pallant 2001; Othman et al. 2020). In general, the validity and reliability tests of the instrument need to be done to avoid misunderstandings about the interpretation of the item and to avoid the instrument failing to measure the objectives and research questions of the study.

There are many studies that examine the validity and reliability of instruments in various fields of study (Rahman et al., 2016; Sajari et al., 2023; Hashim et al., 2023; Matore 2017). Nevertheless, most teaching evaluation questionnaires have not presented sufficient evidence of validity Masuwai et al (2016), and the field of teacher education has been criticized for its lack of attention to validity and reliability issues in evaluation research (Grossman et al., 2008). Therefore, the determination of validity and reliability is to verify the instrument that affords to measure what it actually measures Masuwai et al (2016) and has certain qualities (Sürücü & Maslakci, 2020).

Next, there is issues on the face validity, that, there are past researchers who do not consider face validity as a valid component of content validity. This is because face validity is considered the most minimal and basic index of content validity (Sekaran & Bougie, 2011). Content validity itself requires in-depth statistical evaluation compared to face validity, which only requires intuitive evaluation (Hair et al., 2013). Even face validity is seen as non-psychometric in assessing validity because it is used as a surface-based procedure only. This procedure should be followed by content validity through expert evaluation (Gay et al., 2012).

Next, research on validity and reliability is still less especially research on content validity using the content validity index (CVI). Therefore, validation of Research Instruments Expert review is necessary to ensure construct validity as well as clarity of content (Kline, 2005). Since there is no statistical test to specifically evaluate content validity, researchers usually use a qualitative approach through expert evaluation (Kimberlin & Winterstein, 2008), followed by a quantitative approach using the CVI content validity index.

Content validity refers to the appropriateness of the items in the instrument to measure a construct (Nunally, 1967). A measurement tool that has valid and good content validity is a measurement that covers and represents all or almost all aspects of a construct, and all of the items are relevant to the construct or content that is to be measured (Kamaluddin & Nasir, 2019). Content validity can only be evaluated and determined by experts in the relevant field. Content validity needs to be checked on the measured tool that has been translated because it may contain items that do not fit the local context and socio-culture (Kamaluddin & Nasir, 2019).

The objective of this research is to evaluate the validity and reliability of the instruments of Islamic Cognitive Domain.

Literature Review

Cognitive Domains

Islamic Domain Cognitive is being developed to expand Bloom's taxonomy to fit the nature of Islamic Education. The Framework of Islamic Cognitive Domains provide a sample table of

specifications that would help Islamic Education Teachers to evaluate the students. The framework of the Islamic Cognitive domains have several level such as identifying, remembering, understanding, explaining, internalizing, reflecting, habitualization, and realizing, It is different from Bloom taxonomy which have six level of cognitive domains namely knowledge, comprehension, applying, analyzing, synthesis, Evaluating. Taxonomy comes from the Greek word "tassein" meaning classify, "nomos" means rule. So, taxonomy means a hierarchy of classification over principles or rules. The concept of Bloom's taxonomy was developed in 1956 by Benjamin S. Bloom with his best friend Krathwohl. Benjamin S. Bloom is from the United States. Originally, Bloom introduced two domains in Bloom's taxonomy: the cognitive domain and the affective domain. Subsequently, in 1966, Simpson added the psychomotor domain as a complement to the previous Bloom Taxonomy. So to conclude, there are three domains in Bloom's taxonomy: the cognitive domain, the affective domain, and the psychomotor domain (Nayef et al., 2013; Meng, 2005)

Meanwhile, the Anderson taxonomy revised Bloom taxonomy and replaced the six major categories from Bloom's which is replaced by verbs and several subcategories were reorganized. The knowledge category was renamed with the word "remembering" instead. Comprehension became "understanding" and synthesis was renamed to "creating" to better reflect the nature of thinking (Nayef et al., 2013).

There is also a new taxonomy introduced by Aripin et al (2020) called the holistic taxonomy framework derived from Syed Muhammad Naquib al-Attas school of thought, which have four hiererki namely information, thinking, deep thinking, and Pure submission to God. In addition, there are also Solo Taxonomy that stands for structure of the observed learning outcome. This solo taxonomy describes the level of complexity in a student's understanding of a subject in a five keys stages. It aims to helps make the learning outcomes visible, helps students to search for the next step in their learning and is really useful as a structure for feedback. The five stages were prestructural, unistructural, multistructural, relational, extended abstract.

Fuzzy Delphi Method

The Fuzzy Delphi Method (FDM) is one of the measurement methods that reforms the traditional Delphi method. It is an effective method to solve the uncertain and ambiguous context of the study Jamil & Noh (2021), solve the problem of unpredictability Zulkifli et al (2022), and structure the data collected (Siraj, 2023). This method will involve a group of experts agreeing with the decision, which will translate into a quantitative process empirically. Chang et al (2000) stressed that the objective of FDM is to process the ambiguity of predictive items and the content of the study by an expert panel selected based on the criteria determination.

FDM is commonly used in education studies, especially in aspects of evaluation and determination. Asra et al (2014) used FDM in the determination of M-Learning elements, while Noh et al (2015) used it to design a manual in learning psychology for secondary school. For Islamic studies, FDM was used by Habibah et al (2014) to determine the 'riadah ruhiyyah' for teacher professionalism in Islamic education and Zulkifli et al (2022) in designing the content of religious education learning among children with learning disabilities.

The researcher will develop a survey instrument for FDM in order to achieve agreement on the elements of knowledge taxonomy based on Islamic education with seven levels, starting with identifying, remembering, understanding, explaining, internalising, pondering, habitualization, and realization. The instrument is named as Islamic Cognitive Domains for implementation in the National Education System.

Validity

Validity can be considered as the degree to which a test measures what is supposed to measure. In research validity in data collection means that your findings truly represent the phenomenon you are claiming to measure. So we can say that it is in line with the word appropriate, correctness, meaningfulness and usefulness. There are many methods and techniques can be used to test the validity and reliability of the scales used in quantitative research. For instance, there were content validity, concurrent validity, predictive validity, criterion related validity, construct validity and face validity. In this study, the researcher used content validity to measure the Islamic cognitive domain.

According to Fox (1994), content validity refers to whether the content of the questions or items of the instrument being measured is truly representative and accurate. The expert expressions contained in the measuring instrument represent the phenomenon intended to be measured (Bollen, 1989). It is debatable that in one test, some questions in the questionnaire are not relevant to the intended subject. Content validity can be a trivial issue if the questionnaire contains the right questions to describe the construct being assessed. In other aspects, content validity shows the meaning of the measurement coverage of a concept (Babbie, 2007).

There are situations when the overlapping contents of an item need to be verified by experts. Experts are those who have experience in the field of education. According to Hambleton and Patsula (1999), the validity of external expert panels must be based on the following criteria: (1) specializes in the language, knowledge, and culture of a subject; (2) involving more than one panels and translators from various perspectives and referencing the item. The panel was given the opportunity to provide insight, modify, and provide suggestions on the items.

The satisfaction of an appropriate and valid instrument agreement is necessary to achieve item stability. Any item identified as inappropriate will be dropped. Meanwhile, unclear items need to be corrected, rearranged, or even dropped. Kline (2005) also suggested that content validity should go through the process of a panel of experts. For this study, the instrument's content validity was measured using the value of the Content Validity Index (CVI), which is the level of agreement among a panel of experts (Lynn, 1986). For this study, one instrument was developed to determine the validity of the Islamic Cognitive Domain the implementation in the national education system.

Evaluation of content validity according to expert opinion is a form of statistical analysis based on the content validity of whether the items in the measuring instrument should be on the scale or not, and it is calculated according to the formula by (Lawshe, 1975). According to Lawshe (1975), each statement in the pool of items created is presented to experts to obtain their opinions. The content validity index (CVI) of the instrument is determined by using the value level, which involves the level of agreement between experts (Lynn, 1986). Dichotomous values of agree and disagree were used to assess content validity (Vargas & Luis 2008, as referenced in Sangoseni et al., 2013).

The CVI is different from the content validity ratio (CVR). CVR is the analysis based on scores from the experts. Experts score these statements as "Appropriate", "Appropriate but Should Be Corrected" and "Subtracted". If half of the experts expressed their opinion on the statement in the measuring instrument as 'Appropriate', then $CVR = 0$, if more than half of them stated "Appropriate", then $CVR > 0$, and if less than half of the experts stated "Appropriate" then $CVR < 0$. If the CVR is 0 (zero) or negative, that expression must be subtracted from the measuring instrument (Yeşilyurt & Çapraz, 2018).

Reliability

Reliability is aligned with consistency, accuracy, predictability, equivalence, and replicability. There were three types of reliability such as test-retest, the equivalent form, and internal consistency. The internal consistency can be divided into three for example the split half procedure, the Kuder Richardson procedure and the Cronbach alpha procedure (Tasir & Abu, 2003).

Methodology***Research Design***

The researcher conducted this study using a survey method. The validity of a questionnaire instrument means to what extent an instrument that will be used can test what it should be tested or to what extent the specified indicator has or can meet the item construct (Darusalram & Hussin, 2021). An instrument is said to have high validity if the degree of its ability to measure what it is supposed to measure is high (Konting, 1990). The instrument validity of the content aspect is an important aspect of ensuring that the test items used are truly valid. The validity of a measurement tool or instrument refers to the extent to which the instrument can measure the required aspects (Konting, 2000).

Fraenkel and Wallen (1996) have divided instrument validity into three techniques: content-related evidence of validity, criterion-related evidence of validity, and construct-related evidence of validity. According to Palaniappan (2009), there are several types of instrument validity that are important and can be used in pilot tests, such as face validity, content validity, construct validity, convergent validity, discriminant validity (discriminant validity), criterion-related validity, predictive validity, and concurrent validity.

For the Islamic Cognitive Domain Instrument, the researcher focused on content validity. According to Pallant (2001); Chua (2005); Cresswell (2007), content validity refers to the ability and capability of a research instrument to contain information that includes the component or field to be studied. Content validity is the extent to which the items in the instrument are relevant and represent the targeted construct (Sekaran & Bougie, 2016).

Sample and Data Collection

The content validity technique uses a group of experts (expert judgement or intelligent judgement) to prove the accuracy of item content when evaluating the instrument. Cohan et al (2005) explained that the content of the instrument must be able to demonstrate the validity and comprehensive nature of the domain to be studied. The researcher chose to carry out content validity for this instrument based on the recommendation of Creswell (2007), who stated that the technique of a group of experts (expert judgement or intelligent judgement) needs to be used to prove the accuracy of the item content when evaluating an instrument, or the researcher can refer to a number of experts in their field to confirm the items contained in the study instrument.

Dimopoulus and Pantis (2003) and Makki, Khalick, and Boujoude (2003) stated that at least three field experts are required for item content verification. Mullen (2003) explained that a group of experts are those who are trained in a specific field. Akbari and Yazdanmehr (2014) explained three characteristics of experts:

- Working in their field or profession for more than 5 years.
- Have specific experience.
- The individual is directly involved in the relevant study.

Therefore, content validity is important to ensure that the items used are appropriate and representative of the area being measured (Wiersma, 2000). For that purpose, the researcher has prepared a validation form to be submitted to the selected panels. The selected experts are knowledgeable in the fields of Islamic education and curriculum development. This panel of experts agreed to participate in the validation procedure.

Therefore, the expert content validity procedure begins with the researcher asking for consent from the experts involved. After approval, the researcher gave several documents to the experts, such as an appointment letter, a review account form, and a study synopsis via email. Some selection criteria are determined in the selection panel as follows: Have at least 5 years of teaching experience in the field of Islamic education curriculum and Expert in the field of Islamic education and curriculum development.

A total of four experts were met to confirm each construct and item used in this study. The five experts, consisting of experts from Islamic education and curriculum development, were selected to answer the questionnaire. The details of the experts are as in Table 1;

Table 1

The Panel Profiles of the Questionnaire Item Suitability

Post	Organization/Institution	Field of Expertise	Experience (Years)
Panel 1; Lecturer in the Faculty of Education	The International Islamic University College (KUIS) Bandar Seri Putra, 43000 Kajang, Selangor, Malaysia	i. Islamic Education ii. Education Technology iii. Education Pedagogy iv. Teaching Methodology	>10 year
Panel 2: Associate Professor	Associate Professor (Islamic Education) at International Islamic College University of Selangor	i. Islamic Education Pedagogy ii. Research Methodology (Islamic Education)	>10 years
Panel 3: Curriculum Officer	Ministry of Education	Curriculum Development	>10 years
Panel 4: Lecturer	Faculty of Education (FP) Islamic University Selangor	i. Islamic Education Teaching ii. Islamic Education Curriculum	>10 years

Meanwhile, for reliability, there were 30 respondents answered the questionnaire.

Instrument

This instrument consists of eight constructs, namely identifying, remembering, understanding, explaining, internalising, pondering, habitualization, and realization were given to the respondents. This research uses a Delphi Fuzzy Questionnaire with 66 items. This instrument is built based on literature highlights, previous studies, and document analysis. Next, the items constructed through conceptualization and operationalization according to

the context must be verified in terms of content to ensure that the items represent the theoretical domain of the construct as per the content validity and face validity procedures (Rungtusanatham, 1998).

Analyzing Data

To ensure that the content of the instrument has high integrity and validity, analysis such as the Test Determination Table (JPU), Content Validity Ratio (CVR), and Content Validation Index (CVI) is used as a guideline to prepare items according to the specified criteria. For this instrument, the researcher chose to use content validity using experts, namely CVI analysis.

Thus, the researcher has distributed a set of instrument evaluations (Content Validity Table), which consists of eight constructs, namely identifying, remembering, understanding, explaining, internalising, pondering, habitualization, and realization. The sub constructs, measurement scales, constructed item statements, scales of expert agreement, and a comment column for each questionnaire item statement administered to the expert. A three-point scale is used to assess the importance of each item, namely: (3) Not necessary—Scale 3; (2) Useful but not essential—Scale 2; (1) Essential (very important)—Scale 1 (Lawshe, 1975).

The questionnaire instrument has gone through a process of research, recommendations, and comments from experts. Overall comments and suggestions from the experts are recorded in the review statement form. Then, to determine the value of the content validity agreement reached between experts, the Content Validation Index (Content Validation Index, CVI) is used. CVI takes into account the average rating and degree of suitability given by experts. There is a view that suggests 0.78 and above for three experts and more (Polit & Beck, 2006). The average rating for each construct and its degree of fit will be determined through the calculation of CVI by each expert. A good CVI value is ≥ 0.80 (Davis, 1992). In summary, the number of experts and the CVI value are as shown in Table 2. Further, the formula for the CVI count is as follows:

$$\text{Content Validity Ratio} = \frac{\text{Total score of every expert}}{\text{Total of score}} \\ \text{Average Index} = \frac{\text{Total of CVI}}{\text{Number of expert}}$$

Source: Polit & Beck (2006)

Table 2

Number of Experts and the Implications of Acceptance of CVI Score Deductions

Number of Panel	CVI Acceptance Value	Suggested Resources
Two panels	At least 0.80	Davis (1992)
Two until five panels	Should be 1	Polit & Beck (2006), Polit et al., (2007)
At least six panels	At least 0.83	Polit & Beck (2006), Polit et al., (2007)
Six to eight panels	At least 0.83	Lynn (1986)
At least nine panels	At least 0.78	Lynn (1986)

Reliability

For the reliability of Islamic Cognitive Domain Instrument, The data obtained in this study has been analyzed using the Statistical Package for Social Science (SPSS) version 22.0. Reliability is often utilized interchangeably with stability and internal consistency (Creswell, 2010; Pallant, 2001; Sekaran, 1992). When assessing the internal consistency of a construct, Cronbach's Alpha value is frequently employed (Cronbach, 1946; Norusis, 1977). It is common practice to use a Cronbach's Alpha value greater than 0.60 as an indicator of an instrument's reliability (Konting, 1990; Pallant, 2001). The researcher utilized Cronbach's Alpha values to examine the questionnaire's reliability in light of the above explanation. Table 3 shows the Interpretation of Alpha Coefficient.

Table 3

Interpretation of Alpha Coefficient

Alpha Coefficient	Interpretation
>0.90	Very highly reliable
0.80-0.90	Highly reliable
0.70-0.79	Reliable
1.60-0.69	Marginally/minimally reliable
>0.60	Unacceptably low reliable

Source: Cohen et al (2018)

Findings

Table 4 shows that the content validity index (CVI) for Islamic Cognitive Domains based on the four experts' scores was 0.80. it shows that the cognitive domains instrument has high validity and can be used in the research.

Table 4

CVI Score

Construct	Expert 1	Expert 2	Expert 3	Expert 4	Average of CVI
Islamic Domain Cognitive	68/80	63/80	62/80	66/80	3.23/4
	0.85	0.78	0.77	0.83	0.80

After the CVI steps, the researcher managed to revise the instrument based on the expert comments. Finally, there were 14 items were deleted from the total of 80 items. There were 66 items remaining as shown in table 5.

Table 5

Item deleted and remains of Islamic cognitive domain instrument

No	Construct	No of original item	No of Dropped items	No of item after CVI
1.	Identifying	10	1	9
2.	Remembering,	10	3	7
3.	Understanding	10	2	8
4.	Explaining	10	3	7
5.	Internalizing	10	1	9
6.	Pondering	10	2	8
7.	Habitualization	10	1	9
8.	Realization	10	1	9
Total				66 items

Table 6 shows the reliability of Islamic Cognitive Domain has good internal consistency with a Cronbach alpha coefficient reported for 66 items were 0.9.

Table 6

Reliability

Croanbach Alpha	N of Items
.97	66

Discussion

From the findings, we see that the content validity ratio has a high value that validates the Islamic cognitive domain instrument. Therefore the instrument can be used for the actual research. These findings is in line with Junli et al (2023); Iram et al (2023); Sulung et al (2023); Pandian et al (2023); Sajari (2023) that have cvi above .70. All the research is to validate the instrument. However there are also validation that used CVI to validate modules such as research from Tarmizi and Janan (2022); Sidek et al (2022), and Aulia and Hardeli (2022) which have high score is above .70. Next, the reliability of the Islamic cognitive Domain also have high score is aligned with the research from (Hashim et al., 2023; AlFarisa et al., 2023). The reason of this research is above .70 of CVI score and the reliability is high score is because of the development of the instrument has followed the procedure of the development instrument based on the appropriate model.

Conclusion

Validity and reliability is a prerequisites in the research. It is the researcher's obligation to increase the validity and reliability to make sure the instrument can be used properly. The researcher needs to be careful of the factors that influence the validity and reliability for instance the instructions are not clear, the sentences are difficult to understand, ambiguity, not enough time, the questions are difficult levels, poorly constructed items, the number of items is not enough, and the order of items is not correct.

The significant of this study was to determine the quality and accuracy of the data instrument and to make sure the instrument could be used by another researcher to test it either in different location or in same conditions. For instance, if another researcher would like to test the instrument to form and onwards, they do not hesitate to used the instruments as the instrument that had high reliability.

Next, the significant of this study is to make sure that the instrument measures what intends to be measured. For instance, the content of the instrument is for taxonomy in Islam. So, this instrument is valid for the subject of Islamic education only and for secondary school because it was developed according to the Curriculum and Assessment Standard Document. The validity of the instrument is trying to tell the truth of research findings.

Furthermore, the contributions of this study can be divided into two categories such as theoretical contributions and empirical contextual contributions. For the theoretical contributions, this research gives a new idea for expanding Bloom's taxonomy. This must be further researched as reliability and validity confirm the uses of Islamic cognitive taxonomy. Next, for the empirical contextual contributions, it is hoped that this instrument with high reliability and validity helps to enhance the trustworthiness and generalisability of research findings and minimize the errors in the instrument. Eventually, the instrument can be used readily, and further research can be conducted in various ways such as examining the significant difference between males and females, location and other demographic factors, or other analyses involving statistical techniques.

Recommendations

Future research can use the content validity ratio (CVR) analysis or else. Next upcoming research can use reliability from the Rasch model and EFA which the former can detect item and person reliability.

Limitations

This research limits the validity of instruments that used the content validity index only. The reliability in this research used Cronbach alpha coefficients on the instrument.

Ethics Statements

This is a survey so the researchers have an approval letter from university.

Acknowledgements

The authors thank the Ministry of Higher Education in giving the opportunity to support this research using Fundamental of Research Grant Scheme (FRGS/1/2022/SSI07/UKM/01/1). Also, thank you to Universiti Kebangsaan Malaysia for providing the space and opportunities to conduct this study

Conflict of Interest

There is no conflict of interest.

Funding

Fundamental of Research Grant Scheme (FRGS/1/2022/SSI07/UKM/01/1).

References

- Akbari, R., & Yazdanmehr, E. (2014). A Critical Analysis of the Selection Criteria of Expert Teachers in ELT. *Theory & Practice in Language Studies*, 4(8).
- Alfarisa F, Hidayah, M., Anikoh, Nisa, Tadzkiya, S. (2023). Validity and Reliability Test on Non-Test Instruments of Learning Motivation of Grade IV Students. *Jurnal Hurriah: Jurnal Evaluasi Pendidikan dan Penelitian*. 4 (2), PP 257-265

- Aripin, M.A, R. Hamzah, P. Setya, M. H. M. Hisham, M. I. Mohd Ishar. (2020). Unveiling a new Taxonomy in education field. *International Journal of Evaluation and Research in Education*. 9(3). 524-530. DOI: 10.11591/ijere.v9i3.20458
- Artini Ramlie @ H. & Hussin, Z. (2014). Profesionalisme Perguruan Pendidikan Islam berasaskan 'Riadhah Ruhyyah': Satu analisa keperluan. *The Online Journal of Islamic Education*, 2(1), 7-24.
- Asra, M. R. T. L. A., Saedah, S., & Siti Aisyah, H. (2014). Implementation Model of M-learning Based Discovery Learning on Teacher Education. In *International Conference on Global Trends in Academic Research*, (Vol 1). 335-353.
- Aulia, A., & Hardeli. (2022). Validity of E-Module Based on Problem Based Learning Integrated Demonstration Video and Science Literacy. *IJIS Edu: Indonesian Journal of Integrated Science Education*, 4(1), 45-52.
- Babbie, E. R., Halley, F., & Zaino, J. (2007). *Adventures in social research: data analysis using SPSS 14.0 and 15.0 for Windows*. Thousand Oaks, CA: Pine Forge Press.
- Bollen, K. A. (1989). The consequences of measurement error. *Structural Equations with Latent Variables*, John Wiley & Sons, Inc., 151-178. <https://doi.org/10.1177/1038416217724516>.
- Chang, P. T., Huang, L. C., & Lin, H. J. (2000). The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human resources. *Fuzzy sets and systems*, 112(3), 511-520.
- Chua. (2005). *Kaedah dan Statistik Penyelidikan*. Mc. Graw Hill Education, Malaysia.
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427.
- Creswell, J. W. (2007). *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, 2nd ed. Thousand Oaks, CA, Sage.
- Cronbach, L. J. (1946). A case study of the Splithalf Reliability Coefficient. *Journal of educational psychology*, 37(8), 473
- Davis, L. L. (1992). Instrument Review. Getting the Most from a Panel of Experts. *Applied Nursing research*, 194-197.
- Darusalam, G. & Hussin, S. (2021). *Metodologi Penyelidikan dalam Pendidikan: Amalan dan Analisis Kajian*, hlm. Edisi Ket. Kuala Lumpur: Penerbit Universiti Malaya. doi:10.14425/9789674881733
- Dimopoulos, D.I., & Pantis, J.D. (2003). Knowledge and Attitudes Regarding Sea Turtles in Elementary Students on Zakynthos, Greece. *The Journal of Environmental Education*. 34 (3), 30-38.
- Fox, R. (1994). Validating lecturer effectiveness questionnaires in accounting. *Accounting Education*, 3(3), 249-258.
- Fraenkel, J.R & Wallen, E.N. (1996). *How to Design and Evaluate Research*. United State of America: Mc Graw-Hill, Inc.
- Gay, L.R., Mills, G.E. and Airasian, P.W. (2012) *Educational Research: Competencies for Analysis and Application*. 10th Edition, Pearson, Upper Saddle River.
- Grossman, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research*, 64, 405–408. doi:10.1016/j.jpsychores.2008.02.001
- Habibah Ramlie, Zaharah Hussin, Mohd Ridhuan Mohd Jamil, Ahmad Arifin Sapar, Saedah Siraj, & Nurul Rabihah Mat Noh. (2014). Aplikasi Teknik Fuzzy Delphi Terhadap Keperluan Aspek “ Riadhah Ruhyyah ” Untuk Profesionalisme Perguruan Pendidikan Islam. *The Online Journal of Islamic Education*, 2(2), 53–72.

- Hair, J. F., Celsi, M. W., Orin, D. J., & Bush, R. P. (2013). *Essentials of Marketing research*. 3rd Ed. New York: McGraw Hill.
- Hambleton, R. K., Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Applied Testing Technology Journal*. [www.testpublishers.org/jbook.htm]
- Hashim, N. A. E., Surat, S., & Zulkifli, H. (2023). Kesahan dan Kebolehpercayaan Instrumen Pelaksanaan Pengajaran dan Pembelajaran dalam Talian Guru Pendidikan Islam Sekolah Rendah. *International Journal of Civilizational Studies and Human Sciences*. 6(2): 124-140.
- Imran, I., Khan, R. A., & Aslam, K. (2023, November). The Development and Validation of Student Engagement Questionnaire. In *Proceedings* (Vol. 37, No. 4, pp. 26-30).
- Junli, A. Isa, S. N. I., Ibrahim, F. S., Abd Razak, S. N. R., & Gartika, N. (2023). Development and Validation of Questionnaire For Knowledge, Attitudes And Self-Efficacy On Adult Cardiopulmonary Resuscitation Among Healthcare Providers. 151-158 <https://doi.org/10.22452/jummec.sp2023no2.17>
- Kamaluddin, M. R., & Nasir, R. (2019). Kesahan kandungan dan muka. In Mohammad Rahim Kamaluddin & Wan Shahrazad Wan Sulaiman (Eds.) *Teknik Kesahan dan kebolehpercayaan alat ujian psikologi* (pp.51-66). Penerbit Universiti Kebangsaan Malaysia.
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and Reliability of Measurement Instrument used in Research. *American Journal Health-system Pharmacy*, 65 (23), 2276-2284.
- Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. California: Sage Publication.
- Konting, M. J. (2020). *Kaedah Penyelidikan Pendidikan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, 28(4), 563-575.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-386. <https://doi.org/10.1097/00006199-198611000-00017>.
- Mahamud, S., & Othman, N. (2020). Kesahan dan Kebolehpercayaan Instrumen Kecerdasan Menghadapi Cabaran Teknologi Menggunakan Model Rasch. *Akademika*, 90 (Khas 3), 29-41.
- Makki, M.H., M., Abd-el-Khalick, f., & Boujaoude, S. (2003). Lebanese Secondary School Students' Environmental Knowledge and Attitudes. *Environmental Education research*, 9(1), 21-33.
- Masuwai, A., Tajudin, N. A. M., & Saad, N. S. (2016). Evaluating the Face and Content Validity of a Teaching and Learning Guiding Principles Instrument (TLGPI): A Perspective Study of Malaysian Teacher Educators. *Geografia: Malaysian Journal of Society and Space*, 12(3), 11-21. <https://doi.org/10.1063/1.4983911>
- Mat Saberi, N. A., Radzi, M. N. M., & Loy, C. L. (2023). Assessing content validity of preschool teacher's creative teaching instrument for children's language arts: A Study based on CFTI (Creativity Fostering Teacher Index) using the Content Validity Ratio (CVR). *Jurnal Pendidikan Bitara UPSI*, 16, 147-156. <https://doi.org/10.37134/bitara.vol16.sp2.14.2023>
- Meng, E. A. (2005). *Penilaian Tahap Kecekapan (PTK): Kompetensi Khusus (Fungsi) Pengurusan P7P (Kertas 701/02)*. Selangor: Penerbit Fajar Bakti Sdn. Bhd

- Salleh, M. S. (2018). Pembinaan Konstruksi Instrumen Penilaian Kendiri Akhlak Guru Pendidikan Islam Menurut Pandangan Pakar. *UMRAN-International Journal of Islamic and Civilizational Studies*, 5(3), 1-11.
- Jamil, M. M. R., & Noh, M. R. (2020). *Kepelbagaian metodologi dalam penyelidikan reka bentuk dan pembangunan*. Selangor: Qaisar Prestige Resources.
- Matore, M. M. E @ E., & Idris, H., Abdul Rahman, N., & Khairani, A. Z. (2017). Kesahan Kandungan Pakar Instrumen Ikbar bagi Pengukuran AQ menggunakan Nisbah Kesahan Kandungan. *Proceeding of International Conference on Global Education V (ICGE V)*, Padang, Indonesia (10 – 11 April 2017). *Global Education, Common Wealth and Cultural Adversity*. pp. 979 – 997
- Mullen, P. M. (2003) "Delphi: myths and reality", *Journal of Health Organization and Management*, Vol. 17 No. 1, pp. 37-52. <https://doi.org/10.1108/14777260310469319>.
- Murray, T. J., Pipino, L. L., and Gigch, J. P. (1985). A pilot study of fuzzy set modification of Delphi. *Hum. Syst. Manag.* 5, 76–80. doi: 10.3233/HSM-1985-5111
- Nayef, E. G., Yaacob, N. R. N., & Ismail, H. N. (2013). Taxonomies of educational objective domain. *International journal of academic research in business and social sciences*, 3(9), 165.
- Norusis, M. J. (1977). *SPSS professional statistic 7.5*. North Michigan Avenue, Chicago
- Nunnally, J.C. (1967). *Psychometric Theory*. McGraw-Hill
- Omar, A., Hamzah, S. A., & Kee, C. P. (2021). Kesahan dan Kebolehpercayaan Instrumen Batas Etnik di Sekolah Kebangsaan Zon Utara Semenanjung Malaysia. *Sains Humanika* 18(3): 253–269.
- Othman, N., Had, Ghazali, N. H. C. M., Zabit, M. N. M., & Abdul, N. (2020). Penerokaan Dan Pengesahan Instrumen Kompetensi Matematik Dalam Kalangan Guru Matematik Sekolah Menengah. *Jurnal Penyelidikan Dedikasi*, 18(2), 14–29.
- Pandian, V., Awang, M., Ishak, R., & Ming, G. K. (2023). Validity and Reliability of Organizational Trust Instrument. *International Journal of Academic Research in Progressive Education and Development*, 12(2), 84–96
- Palaniappan, A. K. (2009). *Penyelidikan dan SPSS (PASW)*. Petaling Jaya: Pearson Malaysia Sdn. Bhd.
- Pallant, J. (2001). *SPSS survival manual - a step by step guide to data analysis using SPSS for windows (version10)*. Buckingham Open University Press.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on Research Methods Is the CVI An Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing and Health*. 30: 459–467.
- Polit, D. F., Beck, C. T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Research in Nursing & Health*, 29, 489–497
- Rahman, N. A., Masuwai, A., Tajudin, N. M., Tek, O. E., & Adnan, M. (2016). Validation of Teaching Andlearning Guiding Principles instrument for Malaysian Higher Learning Institutions. *Malaysian Journal of Learning and Instruction*, 13(2), 125-146.
- Rungtusanatham, M. (1998). Let's Not Overlook Content Validity. *Decision Line*, 10-13.
- Sajari, F. F., Zulkifli, H., & Surat, S. (2023). Kesahan dan Kebolehpercayaan Instrumen Kediaan GPI Melaksanakan Model Tasmik j-QAFSecara dalam Talian. *Journal of Quran Sunnah Education and Special Needs*.
- Sangasoseni, O., Hellman, M., & Hill, C. (2013). Development and validation of a questionnaire to assess the effect of online learning on behaviors, attitudes, and clinical practices of

- physical therapists in the United States regarding evidence-based clinical practice. *The Internet Journal of Applied Sciences and Practice*, 11(2), 1-12.
- Sekaran, U., & Bougie, R. (2011). *Research Methods for Business. Skill Building Approach*. 5th ed. New Delhi: John Wiley & Sons Ltd.
- Sekaran, U., & Bougie, R. (2016). *Research Methods for Business. Skill Building Approach*, hlm. seventh ed. John Wiley & Sons Ltd.
- Sidek, S. F., Yatim, M. H. M., Said, C. S. (2022). The Design, and Validation of a tool to Measure Content Validity of a Computational Thinking Game-Based Learning Module for Tertiary Educational Students, *International Multidisciplinary Research Journal*. 1-9.
- Siraj, S. (2023). *Buku penyelidikan reka bentuk dan Pembangunan*. Kedah: Institut Pendidikan Guru Kampus Darulaman.
- Sulong, F., Norrahim, N. M., Ibrahim, N. S., Razak, N. M., Adnan, N. A. M., Mansyur, M., & Nor, N. M. (2023). Validity and Reliability: Questionnaires for Understanding And Interpretation Towards The Malaysian Healthier Choice Logo: Received 2023-07-09; Accepted 2023-09-15; Published 2023-09-20. *Journal of Health and Translational Medicine (Jummec)*, 317-324.
- Sürücü, L., & Maslakci, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, 8(3), 2694-2726.
- Tasir, Z., & Abu M. S. (2003). *Analisis data berkomputer SPSS 11.5 for Windows*. UTM Venton Publishing.
- Vargas, D. D., & Luis, M. A. V. (2008). Development and validation of a scale of attitudes towards alcohol, alcoholism and alcoholics. *Revista latino-americana de enfermagem*, 16(5), 895-902.
- Wiersma, W. (2000). *Research Methods In Education: An Introduction*. Boston: Allyn and Bacon.
- Yeşilyurt, S., & Çapraz, C. (2018). A roadmap for content validity used in scale development studies. *Journal of Erzincan University Faculty of Education*, 3, 251–264. <https://doi.org/10.17556/erziefd.297741>.
- Zulkifli, H., Rashid, S. M. M., Mohamed, S., Toran, H., Raus, N. M., Pisol, M. I. M., & Suratman, M. N. (2022). Designing the content of religious education learning in creating sustainability among children with learning disabilities: A fuzzy delphi analysis. *Frontiers in Psychology*, (Volume 13) <https://doi.org/10.3389/fpsyg.2022.1036806>