ISBN: 978-989-8704-52-8 © 2023

THE APPLICATION OF FUZZY DELPHI METHOD IN CONTENT VALIDITY ANALYSIS

Mohd Azli Jailani[1] and Chong Kim Loy[2]
[1]*Faculty of Social Sciences and Humanities, Open University Malaysia*
47301 Petaling Jaya, Selangor, Malaysia
[2]*UNITAR Graduate School, UNITAR International University*
47301 Petaling Jaya, Selangor, Malaysia

ABSTRACT

Validity is one of the essential criteria for an instrument to be used in counselling and educational learning instruction. Several components of validity requirements need to be validated before using the instrument. One of those components is content validity. As mentioned in the literature, content validity is a qualitative assessment. However, using qualitative assessment to analyse collective agreement among panel experts would create some issues. A few suggestions were made to find a seamless solution to overcome the problems. Among the suggestions is to use quantitative analysis rather than qualitative. Thus, this study aimed to explore an alternative approach to testing the content validity of an instrument. For this reason, a newly developed counselling needs assessment instrument was used to quantify the 16-panel experts' agreement. The instrument has 126 items, and the Fuzzy Delphi Method (FDM) was used to verify the panel experts' consensus on the instrument's content validity. The results have shown that out of 126 items, two items, item 62 and item 115, were discarded from the list as the items scored less than 0.5 of the defuzzification threshold value.

KEYWORDS

Validity, Content Validity, Learning and Instruction Assessment, Counselling Assessment, Fuzzy Delphi Method

1. INTRODUCTION

Testing has been regarded as a fundamental component in counselling services. This is because counsellors constitute a major group of test users (Anastasi, 1992; Leppma & Jones, 2018). The nature of counselling work that need to deal with different types of client's populations would require counsellors to use various types of instruments, which include measurements of cognitive and affective behaviour, self-administered inventories, computerised testing and several other approaches or methods. The same scenario in learning and instruction, the assessment provides educators with valuable information about students' knowledge, skills, and progress. It serves multiple purposes, including evaluating learning outcomes, identifying areas for improvement, and informing instructional decision-making.

To ensure the assessments can truly benefit clients in counselling and students in learning, the instrument itself need to meet certain types of requirements. One of the essential requirements is validity. According to Taherdoost (2016), validity simply can be defined as measure what it supposed to measure. Fraenkel and Wallen (2008, p.147) defined validity as "the appropriateness, meaningful, correctness and usefulness of the inferences a researcher makes." Without validity, a phenomenon such as intelligence cannot be explained and described by an intelligent test because the test did not measure what it supposed to measure.

In general, validity is divided into four different types of categories, firstly, face validity, secondly, content validity, thirdly, construct validity and fourthly, criterion validity (Taherdoost, 2016). Yaghmale (2003) delineated that measuring content validity of instruments are important because it gives confidence to the readers and researchers about the instruments. Muhamad Saiful Bahri Yusoff (2019) defined content validity as the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose.

Heale and Twycross (2015) mentioned that content validity is a subset of face validity. Therefore, Drost (2011), Yaghmale (2003), and Fraenkel and Wallen (2008) described content validity as a qualitative type of

validity rather than quantitative. The procedure of measuring content validity usually involves a group of panel experts who will be given an evaluation form which contains several open-ended questions for them to highlight their comments or responses about the items of an instrument. However, as explained by Rubio et al. (2003) and Ramlan Mustafa and Ghazali Darusalam (2018), this procedure depicted some limitations. Among the limitations are that experts' feedback is subjective; thus, it is prone to be biased, especially when senior experts deliberate their comment, the tendency to accept the comment would be almost certain as compared to a comment which is made by junior lecturers, even though their comment were profound.

The other limitation of measuring content validity via the qualitative method is that the feedback produced cannot be calculated since it is qualitative. Therefore, it is challenging to reduce errors or inconsistencies in the comments made by the panel experts (Ramlan Mustafa & Ghazali Darusalam, 2018).

2. PROBLEM STATEMENT

To overcome the problems mentioned above, a quantitative approach may be seen as the best approach to conduct content validity studies as an alternative to the conventional or qualitative approach.

One of the quantitative methods is the Fuzzy Delphi Method (FDM). Mohd Nazri Abdul Rahman et al. (2016) explained that FDM is not a new technique but an innovated technique from the traditional Delphi method. The traditional Delphi method is an expert opinion survey built with three features: anonymous response, iteration and controlled feedback, and statistical group response. However, this procedure requires researchers to conduct the survey several times, which could delay the research progress and increase the overall costs. To overcome the problems, FDM, as explained by Yu et al. (2010), was introduced to quickly get the consensus from the experts without going through many rounds of survey exercises. The method eliminates ambiguity caused by expert panels' differences in meaning and interpretation.

To identify whether FDM could deliberate information on content validity, a newly developed counselling needs assessment instrument was used to quantify the consensus of the panel experts over the instrument. The counselling needs assessment instrument blueprint, which consists of 126 items, was developed to assist school counsellors in determining the priorities of their guidance and counselling programmes or activities. In this study, FDM was deployed to verify whether the 126 items could assess students' counselling needs under the six components of needs, namely Academic needs, Emotional needs, Personal Development needs, Career needs, Peer Relationships and finally, Family needs.

2.1 Research Objective

To validate the content validity of a newly developed counselling needs assessment instrument through the Fuzzy Delphi method (FDM).

2.2 Research Question

What is the content validity status of a newly developed counselling needs assessment instrument as analysed through the Fuzzy Delphi method (FDM)?

3. LITERATURE REVIEW

Content validity is of significant importance in both counselling assessment and learning and instruction. It ensures that the content covered in assessments and instructional materials accurately represents the knowledge, skills, and competencies that are relevant and necessary for effective counselling and learning outcomes. This section explores the importance of content validity in counselling assessment and learning and instruction, supported by relevant citations.

Content validity is crucial in counselling assessment to ensure that the assessment tools accurately measure the constructs they intend to assess. In the context of counselling, content validity ensures that the assessment instruments cover the relevant areas of motivations, needs, personalities, interests, skills, and

attitudes necessary for effective counselling practice (Hays & Erford, 2014). Assessments with high content validity in counselling provide reliable and meaningful information about clients' psychological functioning, facilitating appropriate interventions and treatment planning.

In learning and instruction, content validity ensures that the instructional materials align with the desired learning outcomes. It ensures that the content covered in textbooks, lesson plans, and other instructional resources is relevant, accurate, and representative of the knowledge and skills that learners need to acquire (Messick, 1996). When instructional materials possess high content validity, they promote meaningful and effective learning experiences, allowing students to acquire the intended knowledge and skills.

Subject Matter Experts (SMEs) play a crucial role in establishing content validity. They possess in-depth knowledge and expertise in the specific domain being assessed. SMEs evaluate the relevance and appropriateness of the content to ensure it adequately represents the construct being measured (Downing, 2006). Their input and judgment are invaluable in determining the content's validity.

Content validity assessment is the most appropriately conducted through a qualitative approach (Brod, Pohlman & Waldman, 2014). The process began when a group of SMEs was selected, and they were asked to review the content thoroughly and provide their qualitative feedback based on the defined content domain. However, some experts, for example, Downing (2006), Messick (1996), Lynn (1986) and Merriam (2009), have deliberated that assessing content validity through qualitative research created some problems and challenges. Among those problems and difficulties are subjectivity, bias, limited generalisation, lack of standardisation and lack of consensus among experts.

To overcome the above-mentioned constraints. Kaufman and Gupta (1988) and Ishikawa (1993) have presented an application, which is known as the Fuzzy Delphi Method (FDM). The FDM approach rectifies the process by inserting some new mechanisms such as spectrum development, aggregation of experts' opinions, defuzzification, and reaching a consensus in just one circle of exercise (Ramlan Mustafa & Ghazali Darusalam, 2018). The advantages of FDM, such as handling ambiguity and uncertainty, capturing expert consensus, addressing vagueness and complexity, efficiency and cost-effectiveness, support for multi-criteria decision making, and flexibility in application, make it a valuable approach for decision-making processes that require flexibility, expert input, and consideration of multiple viewpoints and criteria.

4. METHODOLOGY

The process of reviewing the items by using FDM is illustrated with the following steps:

Step 1 – selection of panel experts: 16 panel experts were selected to verify the developed items in the counselling needs assessment instrument. The experts were identified based on Hsu and Sanford (2007) model of the expert panels selection, which emphasizes two different criteria: a) the experts must have been involved and practised the job, in this case counselling, and b) the length of occupational background must at least minimum 4 to 5 years of experience - following that the selected panel experts for this study were the school counsellors from the States of Selangor and Perak, who fulfilled the outlined criteria.

Step 2 – determining linguistic scale: Each of the panel expert was given a set questionnaire which contained 126 items for them to verify based on the following triangular fuzzy and linguistic scales (Table 1):

Table 1. 5 points linguistic scale and the fuzzy scales

Linguistic scale (5 points)	Fuzzy scale
1 = Highly inappropriate	(0.0, 0.1, 0.2)
2 = Inappropriate	(0.1, 0.2, 0.4)
3 = Moderately appropriate	(0.2, 0.4, 0.6)
4 = Appropriate	(0.4, 0.6, 0.8)
5 = Highly appropriate	(0.6, 0.8, 1.0)

The panel experts were asked to rate whether the developed items could really assess students' counselling needs by circling the items with 5 different linguistic scales as mentioned above. Figure 1 below illustrates the sample of the survey form.

Item	As a student…	Score	My current condition…	Score
1.	As a student, we need to understand the different types of feelings we experience such as happy, sad and angry.	1 2 3 4 5 6	I am able to understand the different types of feelings I experience such as happy, sad and angry.	1 2 3 4 5 6

Score for item appropriateness:				
Highly appropriate	Appropriate	Moderately appropriate	Inappropriate	Highly inappropriate
5	4	3	2	1

Note:
……
……

Figure 1. Sample of the survey form

Step 3 – determining the Threshold "d" value: The Threshold "d" value is important in determining the levels of agreement among the expert panels upon the developed items. The data from this exercise were entered onto a Microsoft Excel worksheet and analysed with the following formula:

$$d(m,n) = \sqrt{\frac{1}{3}[(m1 + n1)2 + (m2 - n2)2 + (m3 - n3)2]}$$

Ramlan Mustafa and Ghazali Darusalam (2018) explained if the threshold "d" value is lower or equivalent to 0.2 (≤ 0.2) it can be said the panel experts have achieved their consensus on the items.

Step 4 – determining group consensus percentage: the data from the Microsoft Excel worksheet file were analysed again to quantify the value of group consensus percentage.

As deliberated by Zanariah Ahmad et al. (2014) the group consensus percentage must exceed 75 percent (> 75 percent) for the next process to be proceeded. If the group consensus percentage is lower than 75 percent it means the items need to be removed or the process need to be redone again.

Step 5 – determining the ∝-cut value (defuzzification value): to determine the ∝-cut value, the data were analysed with the following formula:

$$Amax = 1/3 * [(m1 + m2 + m3)]$$

The formula was used to rank and discard the items which have ∝-cut less than of 0.5. The items with values below than 0.5 (< 0.5) will have to be removed from the list as it did not achieve the levels of agreement or consensus from all panel of experts (Ramlan Mustafa & Ghazali Darusalam, 2018).

5. RESULTS AND DISCUSSION

The feedbacks from the 16 panel experts were entered onto the Microsoft Excel worksheet file. The data were analysed to determine the threshold "d" value. From the data sheet it was found that the overall scores of the "d" value for 126-items is 0.2, which means it surpasses the requirement (≤ 0.2) to proceed with the next procedure.

As deliberated earlier, Step 4 is conducted to determine the group consensus percentage. The consensus percentage from the panel experts must exceed the minimum requirement of 75 percent before it can be proceeded to the next process. From the results, it was found that the overall percentage score for this process is 91.35 percent, which means the group agreement has exceeded the minimum percentage value of 75 percent. The decision whether to retain or discard some items is determined by the final process, that is determining or identifying the ∝-cut value or defuzzification value.

For step 5 (defuzzification value), a formula (i – A = 1/3 * [m1 + m2 + m3]) was used to find the ∝-cut value of each item. According to the formula, If the score value is 0.5 and below (< 0.5), the item should need to be removed from the list. The ∝-cut value also would help the researcher to rank the items according to its priority and importance as per reviewed by the panel experts.

From the data, it has shown that all items have scored the ∝-cut values more than 0.5 (< 0.5), except for item 62 and item 115, which the score points are 0.4. The items were ranked according to the agreement made by the panel experts. Table 2 explains the details.

Table 2. The ∝-cut values and the items rankings

Item	Score value			Item	Score value		
	Fuzzy evaluation	Average fuzzy number	Rank		Fuzzy evaluation	Average fuzzy number	Rank
1	11	0.7	25	55	8.8	0.6	120
2	10.6	0.7	66	56	9.8	0.6	105
3	10.8	0.7	60	57	10.2	0.6	93
4	10.2	0.6	97	58	11.2	0.7	19
5	10.6	0.7	66	59	9.2	0.6	116
6	9.8	0.6	110	60	10.2	0.6	97
7	11	0.7	25	61	10.4	0.7	86
8	11.6	0.7	6	**62**	**6.9**	**0.4**	**126**
9	11.4	0.7	15	63	8.6	0.5	123
10	11.6	0.7	6	64	10	0.6	102
11	9.8	0.6	105	65	10.6	0.7	66
12	12.2	0.8	1	66	10.8	0.7	44
13	10.8	0.7	44	67	11.2	0.7	23
14	10.8	0.7	44	68	10.6	0.7	66
15	10.6	0.7	66	69	11	0.7	25
16	11.6	0.7	13	70	9.2	0.6	116
17	11.8	0.7	3	71	9.4	0.6	113
18	10.8	0.7	60	72	10.8	0.7	60
19	10.4	0.7	86	73	10.8	0.7	60
20	9.8	0.6	110	74	10.6	0.7	66
21	10.8	0.7	44	75	10.2	0.6	97
22	11	0.7	25	76	7.9	0.5	124
23	10.8	0.7	60	77	8.6	0.5	122
24	10.8	0.7	60	78	10	0.6	103
25	10.8	0.7	44	79	9.4	0.6	118
26	11	0.7	25	80	9.2	0.6	116
27	9.8	0.6	110	81	11	0.7	25
28	10.8	0.7	44	82	10.4	0.7	86
29	11.6	0.7	13	83	9.8	0.6	105
30	12.2	0.8	1	84	11	0.7	25
31	11.6	0.7	6	85	11.2	0.7	23
32	10.2	0.6	94	86	10	0.6	103
33	10.6	0.7	66	87	10.6	0.7	66
34	10.8	0.7	44	88	10.8	0.7	44
35	10.2	0.6	97	89	10.8	0.7	44
36	8.8	0.6	120	90	10.8	0.7	44
37	10.8	0.7	44	91	9.8	0.6	105
38	11	0.7	25	92	10.6	0.7	66
39	11	0.7	25	93	10.6	0.7	66
40	9.4	0.6	113	94	10.8	0.7	44
41	10.2	0.6	94	95	10.6	0.7	66
42	10.6	0.7	66	96	11.2	0.7	19
43	11.4	0.7	15	97	11	0.7	25
44	10.2	0.6	94	98	9.2	0.6	116
45	10.6	0.7	66	99	10.4	0.7	86
46	10.6	0.7	66	100	10.8	0.7	44
47	11.8	0.7	3	101	11	0.7	25
48	11.4	0.7	15	102	10.2	0.6	97
49	10.8	0.7	44	103	10.6	0.7	66
50	10.8	0.7	44	104	10.4	0.7	86
51	11.6	0.7	6	105	11	0.7	25
52	10.6	0.7	66	106	10.6	0.7	66
53	9.8	0.6	105	107	10.6	0.7	66
54	11.2	0.7	19	108	10.4	0.7	86

Item	Score value		
	Fuzzy evaluation	Average fuzzy number	Rank
109	10.8	0.7	44
110	11	0.7	25
111	11	0.7	25
112	10.4	0.7	86
113	10.6	0.7	66
114	11	0.7	25
115	**7.1**	**0.4**	**125**
116	11.8	0.7	3
117	11.6	0.7	6
118	11.6	0.7	6
119	11.6	0.7	6
120	11.4	0.7	15
121	11	0.7	25
122	11.2	0.7	19
123	11	0.7	25
124	11	0.7	25
125	10.6	0.7	66
126	11	0.7	25

Finally, item 62 and item 115 were removed from the list as the items have scored the $\propto$-cut of less than 0.5, and the items were ranked 125 for item 62 and 126 for item 115 by the panel of experts.

6. CONCLUSION

In summary, the findings have shown that this study managed to answer the research question: "What do the panel experts say about the 126-items counselling needs assessment instrument through the Fuzzy Delphi method (FDM) of analysis?"

Through the process, the agreement of the panel experts was quantified and summarized quantitatively. This has proven the earlier assumption that the time of completing the process of getting the panel experts consensus can be shorten via FDM. This exercise also is seen to be practical in eliminating confusions that were driven from the conventional method, which is more qualitative in nature. FDM standardizes the process especially through its linguistic scale for not to limit the feedback with just the normal responses such as 'good' or 'very good' but with the fuzzy scales. FDM also allows for a more comprehensive and nuanced evaluation of the relevance and representativeness of content in the context of counseling and learning and instruction

In conclusion, the utilisation of FDM in assessing content validity holds significant implications for counseling and learning and instruction. The FDM, which combines the Delphi technique with fuzzy logic, offers a systematic and rigorous approach to gather expert opinions and reach consensus in situations characterised by uncertainty and ambiguity.

REFERENCES

Anastasi, A. (1992). *What counsellors should know about the use and interpretation of psychological tests*. Journal of counselling and development, 70, 610 – 615.

Arash Habibi, Farzad Firouzi Jahantigh & Azam Sarafrazi. (2015). *Fuzzy Delphi Technique for Forecasting and Screening Items*. Asian Journal of Research in Business Economics and Management, 5(2), 130-143.

Downing, S. M. (2006). Twelve Steps for Effective Test Development. In S. M. Downing & T. M. Haladyna (Eds.), Handbook of test development (pp. 3–25). Lawrence Erlbaum Associates Publishers.

Drost, E. A. (2011). *Validity and reliability in social science research*. Education Research and Perspectives, 38 (1), 105-124.

Fraenkel, Jack R, & Norman E. Wallen. (2008). *How to Design and Evaluate Research in Education*. Boston: McGraw-Hill.

Hays, D. G., & Erford, B. T. (2014). Developing Multicultural Counseling Competence: A Systems Approach (3rd ed.). Pearson.

Heale, R., & Twycross. (2015). *Validity and reliability in quantitative studies*. Evidence-Based Nursing, 18(3), 66 – 67.

Hsu, C., & Sandford, B. (2007). *The Delphi technique: Making sense of consensus*. Practical Assessment, Research and Evaluation, 12(10), 1 – 8.

Hsu, Y.L. Lee, C.H. & Kreng, V.B. (2010). *The application of Fuzzy Delphi Method and Fuzzy AHP in lubricant regenerative technology selection*. Expert Systems with Applications, 37 (2010), 419–425.

Leppma, M., & Jones, K.D. (2018). *Multiple Assessment Methods and Sources in Counselling: Ethical Considerations*. Retrieved from http://www.counseling.org/knowledge-center/vistas

Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382–385.

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation. John Wiley & Sons.

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13(3), 241–256.

Mohd Nazri Abdul Rahman, Mariani Md Nor, Noraziah Ahmad Nadzim, Nor Mashitah Mohd Radzi & Nurnadia Moktar. (2016). *Application of fuzzy delphi approach in designing homeschooling education for early childhood Islamic education*. International Journal of Academic Research in Business and Social Sciences, 6(12), 837 – 846.

Muhamad Saiful Bahri Yusoff. (2019). *ABC of content validation and content validity index calculation*. Education in Medicine Journal, 11(2), 49 – 54.

Ramlan Mustafa, & Ghazali Darusalam. (2018). *Applikasi Kaedah Fuzzy Delphi dalam penyelidikan sains sosial*. Kuala Lumpur: Penerbitan Universiti Malaya.

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*. National Association of Social Workers, Inc. 94 – 104.

Taherdoost, H. (2016). *Validity and reliability of the research instrument; How to test the validation of a questionnaire/survey in a research*. International Journal of Academic Research in Management (IJARM), 5, 28-36.

Tee, Q.T. Anizah Mohamed, & Norlidah Alias, (n.d). Application of Fuzzy Delphi Method in Exploring Facebook as a Tool for Writing Therapy Among Young Adults. The Malaysian Online Journal of Educational Technology, 2(3), 1-10.

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

Yu, L.H., Cheng, H.L., & Kreng, V.B. (2010). *The application of Fuzzy Delphi method and Fuzzy AHP in lubricant regenerative technology selection*. Experts Systems with Applications, 37, 419 – 425.

Zanariah Ahmad, Mohamad Muhidin Patahol Wasli, Mohd Salihin Hafizi Mohd Fauzi, Mohd Ridhuan Mohd Jamil & Saedah Siraj. (2014). *Fuzzy Delphi analysis for future environmental education using interactive animation*. 2nd International Seminar Teaching Excellence and Innovation, Universiti Malaya.