

The Validity of the Malaysian Teachers' Global Competency Level Instrument Using Cohen Kappa, Content Validity Ratio and Content Validity Index Analyses

Siti Nurul Aqiedah Ibrahim^a & Mohd Effend Ewan Mohd Matore^{b,*}

^aFaculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

^bResearch Centre of Education Leadership & Policy, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract

Highly competent teachers are vital in developing globally competent individuals. There is a lack of empirical evidence supporting the validity aspect of the Malaysian Teachers' Global Competency Level Instrument. This study aims to evaluate face and content validity. A survey research design with a quantitative approach was conducted. It involves two experts for face validity and eight experts for content validity using purposive sampling techniques. For face validity, the two experts appointed were Malay language teachers with more than five years of teaching experience. To assess content validity, eight experts, including four professionals in measurement, evaluation, and global competency, and four field practitioners: teacher educators and outstanding teachers. The instrument consists of 73 items with four constructs: self-awareness, global awareness, attitudes & values, and skills. The analysis involved Cohen's Kappa for face validity, while content validity involved CVR and CVI. The results showed that face validity was (N=2, k=0.640), while for content validity, (N=8, CVI=0.95, and CVR=2 items refined). This instrument demonstrates strong validity as a measurement tool for the global competency level of Malaysian teachers. Further studies are recommended to be conducted: employing an advanced statistical analysis, like the Rasch Model, to enhance higher-quality items.

Keywords: content validity; face validity; global competency; teachers

Received: 15 March 2025

Revised: 20 April 2025

Accepted: 25 April 2025

1. Introduction

While globalization is progressing rapidly, educators face the challenge of preparing students to face an interconnected world. Teachers today play a significant part in equipping students for a fast-changing world (Mopara & Sanrattana, 2023). They must understand, appreciate, and respond to global issues in today's modern education. Cultural awareness, strong communication in different settings, and the ability to tackle global challenges using local solutions are among the skills needed to do this effectively. Nurturing these competencies among teachers is especially important, as it aligns with the Malaysian nation's vision of fostering a globally competitive and culturally harmonious society.

Technology competency, which is a part of global competency, is one of the key aspects in meeting the global needs of the 21st century. It has made it easier for us to learn from one another by bridging geographical gaps and expanding our opportunity to travel, interact with new people, experience different cultures, learn new languages, and exchange and develop ideas. It has never been quicker, simpler, or more inspiring for educators and learners to bring the outside world into the classroom. Technologies offer essential tools for 21st-century learning (Ramaila & Molwele, 2022). Furthermore, today's kids are undeniably inspired by technology. Today's kids are already accustomed to using digital devices as gateways to the outside world. When effectively utilised, these tools can offer kids essential connections to individuals, locations, and issues that extend well beyond their local community.

In the Malaysian context, the need for global competence, particularly in technology literacy, has become a national priority, in alignment with the Industrial Revolution 4.0, which has driven significant transformation in the education sector through the reformation of teaching and learning techniques. Not only students, teaching 4.0 encourages teachers

* Corresponding author.

E-mail address: effendi@ukm.edu.my

to integrate various types of technology in their classroom instruction so they can render it more interesting to new generations (Masdoki, Din & Mohd Matore, 2021). Globally competent individuals critically evaluate their own and others' perspectives, integrating insights grounded in factual evidence, scientific understanding, and historical context. This can communicate the distinctions between different points of view. One interesting approach to give such encounters is through digital technology, which has played a significant role in expanding chances for global exploration, instruction, and reflection. Such technologies are significantly beneficial for those with limited mobility or restricted access to travel opportunities.

Jiixin et al., (2024) highlight the need for defining and measuring global competence in different cultural contexts. Hence, this instrument is developed to measure Malaysian teachers' global competency in the local context. The debate about the reliability and validity aspect has gained fresh prominence with many arguing the usefulness of the items, especially from experts' perspectives. The instrument adapted from Western culture may not be fit with the local context. Besides, the changes to the items may need to be reassessed for local use. The application of certain analyses, such as Content Validity Ratio (CVR), Content Validity Index (CVI), and Cohen Kappa, will help to enhance the quality of the measurement items themselves.

1.1. Research Objectives

There are three objectives of this study, which are;

- a. Evaluating the face validity of the Malaysian Teacher's Global Competency Level Instrument by using Cohen's Kappa analysis.
- b1. Evaluating the content validity of the Malaysian Teacher's Global Competency Level Instrument by using Content Validity Ratio (CVR).
- b2. Evaluating the content validity of the Malaysian Teacher's Global Competency Level Instrument by using Content Validity Index (CVI) analysis.

1.2. Global Competency

According to the OECD (2018), global competence refers to a person's capacity to think critically and take the role of an open-minded global citizen. This includes embracing multiculturalism, understanding the diverse perspectives shaped by different backgrounds, and effectively engaging with individuals from various cultures to promote collective well-being and support sustainable development. Kopish (2016) emphasizes that globally competent teachers are those who bridge local and global challenges, even when working in small schools. Their influences are beyond the classroom, where they can develop the young generation into a global society that is informed, connected, and globally competent. In summary, a teacher's global competence refers to their ability to cultivate and shape students into open-minded individuals who respect cultural diversity, communicate effectively, plus collaborate with people from varied backgrounds to promote global sustainability. In the Malaysian context, global competence encompasses moral values, respect for diversity, productive cross-cultural communication, and the capacity to act on global issues. It extends above language proficiency, incorporating a comprehensive combination of cognitive, social, and emotional skills (Goh & Ooi 2020). Global competence in Malaysia can be defined as an individual's ability to engage meaningfully in an interconnected world, demonstrate tolerance to cultural diversity, communicate effectively, critical awareness of other cultures, and the ability to collaborate as a proactive global citizen.

1.3. The 21st Century Global Competency in Technology

In today's globalized world, achievement in schooling, job hunting, and civic life requires global competency in terms of technology literacy. Individuals with this competency are better prepared to face the problems and possibilities of the 21st century. Cross-cultural analyses and investigations into technology's role in skill development could further enrich understanding and inform strategies for enhancing educational practices in preparation for a globalized world (Herlinawati et al, 2024). The three categories of 21st Century Skills are learning skills, literacy skills, and life skills. Literacy skills encompass Information Literacy, Media Literacy, and Technology Literacy. As learning with new technology is done online, digital skills and new literacies are becoming increasingly popular. New technology-age learners must enhance new literacy abilities such as digital, technological, and Internet literacy as a means to meet the new demands of today's educational problems (Yeşilyurt, & Vezne, 2023).

Information literacy is the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning (ACRL, 2015). Understanding information means accurately comprehending its meaning and ramifications, which may be broad in scope. To produce information means to use the communication abilities outlined in the previous section to impart new knowledge, whether orally, in writing, or through diagrams. Information literacy is the foundation of all other skills since it enables a person to learn about the world around them. The 21st century is commonly referred to as the Information Age, therefore, the capacity to look past the torrent of misinformation is critical (Luqiu, 2019). Media literacy is described as understanding the many channels and media through which information is created and delivered. Communication skills are essential in this context, as individuals must be able to identify the source, intended audience, purpose, and situational context. Individuals who understand these four aspects will be able to make informed decisions about the relevance of the information. In the twenty-first century, media bias is prevalent; therefore, having media literacy in your arsenal is your best shot for avoiding disinformation and making educated judgments (Potter, 2024). Technology literacy can be simplified as knowing what can be done with the technology, how to use technology proficiently, and deciding on the type of technology to be used (Yeşilyurt, Vezne, 2023). As technology becomes an increasingly dominant component of 21st-century life, humans must comprehend its purpose, workings, and ramifications to avoid misuse and successfully solve issues. Additionally, technological literacy may promote creativity and progress.

1.4. Instrument Validity

Wang et al. (2023) suggested to perform a validity test of an instrument to guarantee the validity of the developed instrument and involved procedure. Omar, Hamzah, & Kee (2021) stated that validity is the measure of how well an instrument precisely assesses the specified attributes. This is supported by Anggraini et al. (2023). The validity can be proven if the items and instruments can provide values that match what is measured. Research can utilize multiple types of validity, including face validity, content validity, construct validity, convergent and discriminant validity, criterion validity, predictive validity, concurrent validity, cross-cultural validity, and simultaneous validity (Zhang & Garcia, 2023; Cohen et al., 2018; Bolarinwa, 2015).

Face validity assesses the clarity and purpose of a construct used for measurement. It is the simplest and most foundational type of validity (Zhang & Garcia, 2023). It is carried out by collecting responses from sample that share common traits with the actual study sample (Baco & Ishak, 2020). Ensuring that the developed items are appropriate, logically structured, and linguistically adapted with precise word selection is vital in establishing their face validity (Sabtu et al., 2024). According to Oluwatayo (2012), face validity can be achieved with the presence of experts in the studied field or individuals with no psychometric experience providing a quantitative assessment of the instrument's appropriateness for its intended purpose. However, as face validity lacks statistical data support, it should not be considered a measurement validity principle. Therefore, following face validity procedure, content validity testing should follow afterward (Cohen et al., 2018; Gay et al., 2012), as content validity is a more in-depth statistical evaluation compared to the intuitive assessment of face validity (Hair et al., 2013).

To improve the perception of the developed instruments, content validity is often used in validity studies as empirical evidence. It assesses how fit alignment of the instrument's items with the applied measurement scale (Amatan et al., 2021), effectively measures the intended construct (Kamaluddin et al., 2017), and guarantees that the measurement encompasses all dimensions of a construct (Bolarinwa, 2015). It is validated based on the level of conformity among experts regarding the instrument's items (Zainal et al., 2020). Expert validation can be performed through a panel of construct theorists (Bhattacharyya et al., 2017) or content professionals (Anizar et al., 2018). The review obtained from this process is essential in improving and optimizing the questionnaire (Darusalam & Hussin, 2021; Oluwatayo, 2012). Mohamat el al. (2022) reported that content validity testing that involves multiple experts offers advantages, as it promotes greater objectivity and helps prevent imbalance or bias.

The expert panel's consensus on validity will assess whether the developed items effectively reflect the four constructs of the Malaysian Teachers' Global Competency Instrument in measuring teachers' global competency levels. Additionally, experts will provide suggestions for improving measurement items or recommend their removal. They will also have the opportunity to propose more suitable replacements for existing items. Only items that achieve a high level of consensus based on Cohen's Kappa, CVR, and CVI analysis will be retained.

1.5. Problem Statement

In summary, sustainable global competency among teachers is crucial in shaping students who can compete at the global level (Jiaxin et al., 2024), ultimately becoming proactive global citizens. Although comprehensive studies on the global competency of Malaysian educators are limited, current discoveries highlight the demand for action to ensure that Malaysian teachers achieve top-tier of global competency. Research on the global competency of in-service teachers remains scarce, both internationally and in Malaysia. This lack of research makes it challenging for stakeholders to accurately assess teachers' global competency levels, particularly within the country. Although there is a rising awareness of the significance of global competence in 21st-century education, instruments to assess teachers' readiness to integrate a global perspective into their teaching practice are still limited (Jiaxin et al., 2024). Therefore, the development of a valid and reliable instrument, that is adapted to the educational context in Malaysia, is an urgent need. Such an instrument can not only provide insight into teachers' strengths and weaknesses but can also help formulate professional development programs and policy initiatives to strengthen global education in Malaysia.

Hence, this research focuses on filling that gap by assessing the validity aspect of the instrument developed to measure the global competence level of Malaysian teachers. Based on the recognized global competence framework and adapted to the Malaysian cultural and educational landscape, this instrument aims to provide an evidence-based approach to assess and improve global competence among educators. By systematically exploring the main dimensions of global competence and validating this instrument through empirical methods, this study contributes to the development of global education knowledge and offers practical solutions to improve the quality of teachers in Malaysia.

2. Methods

2.1 Research Design and Instrument

The research design of this research is a survey study that applies a quantitative method by applying a questionnaire form of the Malaysian Teachers' Global Competence Instrument. Hence, this instrument is tested by examining its face and content validity. In developing this instrument, a literature review process is employed to identify the definition and constructs of global competency. Hence, determining operational definitions and conceptualization for each construct and selecting items that are appropriate and meet the construct's requirements. The items in this instrument were adapted from the Global Competence Rubric by Parmigiani et al. (2022), Parmigiani et al. (2023), and the PISA Global Competence Questionnaire, OECD (2018). As a result of this process, 73 items were systematically selected, where 32 items were from the Global Competence Rubric and 41 items were from the PISA Global Competence Questionnaire. Subsequently, they were classified according to the construct of Self-Awareness (16 items), Global Awareness (12 items), Attitudes & Values (21 items), and Skills (24 items).

2.2 Expert Sampling and Criteria

The selection of a qualified and reliable board of experts to check the content validity of the instrument is a critical step in affirming that each item in the instrument is relevant and appropriate to the studied construct. These panels usually consist of individuals with profound expertise in the relevant field, covering academics, professionals, or experienced practitioners. Among the main criteria in the appointment of experts are mastery in the field, experience in instrument development, and diversity of perspectives. This is to ensure a holistic and inclusive assessment. Ayre and Scally (2014) suggested the ideal number of expert panels is between five and ten people to reach a consensus without compromising the range of opinions. The appointment of a high-quality expert panel is important in ensuring the reliability and validity of the instrument for use in wider research. Two experts for face validity and eight experts, namely four field experts and four field experts for content validity experts, were involved in this study. The appointment criteria for face validity experts are (1) having experience in teaching Malay for more than five years and (2) having at least a degree in Malay language education. Meanwhile, the criteria for the appointment of content validity experts are (1) having experience in their areas of mastery for more than five years, (2) specializing within the discipline of education or measurement and evaluation, and (3) having at least a master's level of education.

2.3 Face and Content Validity Procedure

The expert review form developed and utilized in this study contains a description of the study and guidelines for filling out the expert validation form. The description or information of the study that is highlighted is the purpose and

objective of the study, while the guidelines cover the criteria measured for validity, an explanation of the measurement scale, and a description of the acknowledgment of the expert review form. It also includes guidelines for the panel to rate each item according to validity criteria. For face validity, the criterion measured is readability using a 2-point Likert scale (1 and 2), namely Scale 1 for Yes and Scale 2 for No. The expert validation form for content validity evaluates accuracy and appropriateness criteria using a 4-point Likert scale that represents the level of agreement on content validity attributed to Yaghmaie (2003); Scale 1 for Not Suitable, 2 for Not Suitable, 3 for Suitable, 4 for Very Suitable. Experts were contacted via email to obtain their consent to be part of the expert panel for this study. Once they agreed, an appointment letter as a face or content validity expert and the expert validity form were also sent via email. The expert panel then returned the validity form after reviewing, evaluating, and digitally signing it via email. Face validity data was analysed using Cohen's Kappa analysis, while content validity data was analysed through the Content Validity Ratio (CVR) and Content Validity Index (CVI).

2.4 Cohen's Kappa

Face validity data was analysed using Cohen's analysis via the GraphPad website (2024). Cohen's Kappa (k) is an agreement measure used to evaluate the level of conformity between two raters (also known as judges or observers) (Dettori & Norvell, 2020). It is a statistical measure to evaluate the extent of consensus among experts in assigning the same category, exceeding expected agreement by chance (Rau & Shih, 2021). Cohen's Kappa values range from -1 to 1, enabling the assessment of inter-rater reliability beyond chance. Perfect agreement (k = 1) is rarely achieved, but values approaching 1 indicate excellent agreement (Sabharwal, 2021). According to Landis and Koch (1977), the interpretation of Cohen's Kappa can be categorized into six agreement levels. A kappa value ranging from 0.81-1.00 indicates excellent agreement, while values between 0.61 and 0.80 reflect good agreement. Moderate agreement is presented by the value 0.41-0.60, and fairly weak agreement falls between 0.21-0.40. A Kappa value between 0.00-0.20 is considered weak and any value below 0.00 is interpreted as very weak agreement.

2.5 Content Validity Ratio (CVR)

CVR is the ratio of content validity for items, representing the ratio of respondents who rated items on scales 3 and 4 for each validity aspect to the total number of experts involved in content validation. The CVR formula is as follows:

$$CVR = \frac{ne - \frac{N}{2}}{N/2} \quad (1)$$

Where:

ne = number of experts who rated the item 3 or 4

N = total number of experts

The obtained CVR value ranges from -1 to +1, with values closer to +1 indicating that most experts agree the item is essential and should be included in the instrument. The CVR value is then compared to the critical CVR threshold to determine whether the item should be retained or removed (Almanasreh et al., 2019; Lawshe, 1975). According to Lawshe's (1975) criteria, items must achieve a CVR value of at least 0.56 to be considered for retention in the instrument.

2.6 Content Validity Index (CVI)

The CVI analysis is conducted after obtaining the CVR values (Lawshe, 1975). CVI refers to the average validity score for all constructed items (Lindell & Brandt, 1999). This study employs two levels of CVI which are the Item-CVI (i-CVI) and the Scale-CVI (s-CVI). The i-CVI is obtained by dividing the number of experts who rated the item as 3 or 4 by the total number of experts. Meanwhile, the s-CVI is determined by averaging all i-CVI scores and dividing them by the number of total items within each construct. The CVI formulas are as follows:

$$i - CVI = \frac{Nc}{n} \quad (1)$$

$$s - CVI/Ave = \frac{\text{Total score } i - CVI}{\text{No. of expert}} \quad (2)$$

$$s - CVI/UA = \frac{\text{Total score UA}}{\text{No. of item}} \quad (3)$$

where:

Nc = number of experts rating the item 3 or 4

n = total number of experts

The i-CVI and s-CVI values obtained in this study will be compared with the recommended thresholds of i-CVI ≥ 0.78 (Lynn, 1986) and s-CVI ≥ 0.90 (Polit et al., 2007; Oktari et al., 2020). If the i-CVI value exceeds 0.78 and the s-CVI value exceeds 0.90, the item is accepted and retained in the instrument; otherwise, it will undergo revision or be removed from the instrument.

3. Results

3.1 Objective 1 - Measuring the face validity of the Malaysian Teacher's Global Competency Level Instrument by using Cohen Kappa analysis.

Out of 73 items, 69 items received agreement from both experts. The Cohen's Kappa value of k = 0.640 indicates that the agreement level between the two evaluators is at a good level on the agreement scale. The results of Cohen's Kappa analysis using the template from the GraphPad website are shown in Table 1.

Table 1. Cohen's Kappa Analysis Findings

Total observed agreement	69	94.524%
Expected total agreement by chance	61.9	84.76%
Cohen's Kappa value	0.640	

3.2 Objective 2a- CVR Analysis for Content Validity

Lawshe (1975) suggested setting the critical CVR value at 0.75, referring to the involvement of eight experts. Items with CVR values lower than the critical threshold of 0.75 are deemed non-essential and will be excluded from the instrument. Based on Table 2, 36 items achieved full agreement (CVR=1.00) from the eight expert panel members. The total number of retained items is 71, while two items, KM15 and KG9, were removed due to CVR values below 0.75, scoring 0.625 and 0.50, respectively.

Table 2. Average CVR values for constructs Self-Awareness (KD), Global Awareness (KG), Attitudes & Values (SN), and Skills (KM) in terms of appropriateness and accuracy

CVR	Item	Total	Decision
1.00	KD1, KD2, KD5, KD6, KD10, KD11, KD12, KD13, KD16, KG1,KG2, KG3, KG4,KG5, SN3, SN6, SN8, SN11, SN13,SN15, SN17, SN18,SN19, SN20,SN21,KM1, KM3,KM4, KM8, KM9, KM11, KM13, KM14, KM17, KM21, KM24	36	Retained
0.875	KD3, KD4, KD7, KD8, KD14, KD15, SN1, SN5, SN9, SN12, SN16, KM2, KM7, KM10, KM12, KM16, KM18, KM19, KM23	19	Retained

CVR	Item	Total	Decision
0.75	KD9, KG6, KG7, KG8, KG10, KG11, KG12, SN2, SN4, SN7, SN10,SN14, KM5, KM6, KM20, KM22	16	Retained
0.625	KM15	1	Removed
0.50	KG9	1	Removed

3.3 Objective 2b- CVI Analysis for Content Validity

The CVI analysis results indicate that the i-CVI values for all items and the s-CVI values for each construct exceeded the required thresholds of 0.78 for i-CVI and 0.90 for s-CVI. A total of 35 items achieved a perfect i-CVI value of 1.00, while the lowest recorded i-CVI value was 0.81 for item KM15. Therefore, no items were removed based on the CVI analysis findings.

Table 3: i-CVI and S-CVI values for constructs Self-Awareness (KD), Global Awareness (KG), Attitudes & Values (SN), and Skills (KM)

Self-Awareness (KD)		Global-Awareness (KG)		Attitudes & Values (SN)		Skills (KM)	
Item	i-CVI	Item	i-CVI	Item	i-CVI	Item	i-CVI
KD1	1.00	KG1	1.00	SN1	0.94	KM1	1.00
KD2	1.00	KG2	1.00	SN2	0.94	KM2	0.94
KD3	0.94	KG3	1.00	SN3	1.00	KM3	1.00
KD4	0.94	KG4	1.00	SN4	0.88	KM4	1.00
KD5	1.00	KG5	1.00	SN5	0.94	KM5	0.88
KD6	1.00	KG6	0.88	SN6	1.00	KM6	0.88
KD7	0.94	KG7	0.88	SN7	0.94	KM7	0.94
KD8	0.94	KG8	0.88	SN8	1.00	KM8	1.00
KD9	0.88	KG9	0.88	SN9	0.94	KM9	1.00
KD10	1.00	KG10	0.88	SN10	0.88	KM10	0.94
KD11	1.00	KG11	0.88	SN11	1.00	KM11	1.00
KD12	0.94	KG12	0.88	SN12	0.94	KM12	0.94
KD13	1.00			SN13	1.00	KM13	1.00
KD14	0.94			SN14	0.88	KM14	1.00
KD15	0.94			SN15	1.00	KM15	0.81
KD16	1.00			SN16	0.94	KM16	0.94
				SN17	1.00	KM17	1.00

Self-Awareness (KD)		Global-Awareness (KG)		Attitudes & Values (SN)		Skills (KM)	
Item	i-CVI	Item	i-CVI	Item	i-CVI	Item	i-CVI
				SN18	1.00	KM18	0.94
				SN19	1.00	KM19	0.94
				SN20	1.00	KM20	0.88
				SN21	1.00	KM21	1.00
						KM22	0.88
						KM23	0.94
						KM24	1.00
S-CVI	0.96	S-CVI	0.93	S-CVI	0.96	S-CVI	0.95

These findings establish that the majority of the items in this instrument are valid and can be retained for further use in measuring the global competency level of Malaysian teachers.

4. Discussion

By incorporating Cohen Kappa analysis for face validity and CVR and CVI analysis for content validity, this research focuses on validating the Malaysian Teachers’ Global Competency Level Instrument. The result obtained contributes to the reliability of this instrument in measuring the global competency level of Malaysian teachers.

For the first objective, the result gained by Cohen's Kappa analysis for face validity showed that the kappa value was at a good level, in accordance to Landis and Koch (1977). Most of the items in this instrument were rated consistently by both experts and are considered appropriate in terms of face validity. Overall, this finding shows that almost all the items in this instrument meet the criteria of face validity from the aspect of readability. However, the value of Cohen’s Kappa suggests there potentially may be some errors. As face validity is arguably the least reliable type of validity (Taherdoost, 2016), it is advisable to perform a content validity analysis to further evaluate this instrument.

A further evaluation for this instrument as aligned with the second objective was conducted by using CVR analysis. The result shows that 71 items out of 73 items were accepted to be retained in this instrument with CVR values ranging from 0.75 to 1.00, where this shows a prominent share of expert agreement in determining the prerequisites of those items in the instrument (Ayre & Scally 2014). Two items that recorded a CVR value of less than 0.75 and also the items will be modified and improved according to the recommendations and comments from the expert panel. The items in this instrument are of high quality and deemed essential by the experts.

For the third objective, findings from the i-CVI analysis for each item in this research instrument ranged from 0.81 to 1.00, beyond the recommended baseline of 0.78 (Polit et al., 2007). While the s-CVI values for all four constructs ranged from 0.93 to 0.96, surpassing the established threshold of 0.90 for scale validity (Lynn, 1986). These results show that i-CVI and s-CVI values have met the designated parameters, and all of them maintain construct validity. These confirm that this instrument is valid and appropriate for assessing Malaysian teachers’ global competency. However, it is suggested to further analyze the items in this instrument by measuring their psychometric traits using the Rasch Measurement Model (Kamaruddin & Mohd Matore, 2021; Mohd Noh & Mohd Matore, 2020).

This validated instrument provides a valuable tool for informing teacher training and educational policy on teachers’ global competency level. The education authorities can design targeted professional development programs aligned with the Global Competence Framework (OECD, 2018) by identifying specific strengths and weaknesses in teachers’ awareness, intercultural sensitivity, and engagement with global issues. Such programs could ensure that teachers are better equipped to nurture global competencies in students, contributing to a more globally aware and culturally responsive education system. This can support the national education goals in developing globally competent citizens.

In this study, educational technology plays a pivotal role in the administration and analysis of face and content validity. Expert feedbacks were collected via email, while the consensus indices, such as Cohen's Kappa, CVR, and CVI were calculated using the Graphpad statistical platform and Microsoft Excel. These tools enabled efficient communication, data organization, and statistical analysis, which streamlined the validation process and minimized human errors. The integration of these tools mirrors the technology or digital competencies that underpin the construct of global competence itself, particularly in the context of digitally enhanced communication and intercultural engagement (Redecker, 2017; OECD, 2018). In addition, more advance analysis maybe needed to strengthen the quality of the items by evaluating using reflective measurement model consists of reflective indicator loadings, internal consistency reliability, convergent validity, and discriminant validity (Subhaktiyasa, 2024).

5. Conclusion

This study effectively collected consensus from language experts, professional specialists, and field experts for all 73 developed items. For Cohen's Kappa validity, the obtained value falls within the "good" category, indicating a high rate of agreement between the two language experts. Meanwhile, the strong CVR value demonstrates consistent agreement among expert evaluators. In terms of content validity, the study found a high level of consensus from the eight experts. These findings may help to enhance a better understanding of the necessity of conducting both face and content validity analysis to ensure the development of comprehensive, reliable, and accurate instrument tools. It also shows how educational technology can enhance the validation process and enable the integration of global competency assessment into digital platforms for teacher development. An implication of this is the possibility that this instrument will be used widely in assessing Malaysian teachers' global competency. These findings also suggest the potential for this instrument to be adapted for use in other educational contexts, making it useful for cross-cultural validation and international study.. However, the current study has only examined face validity by involving two language experts. It is advisable to have more language experts to ensure the high quality of items in terms of readability and language style through Fleiss' Kappa analysis. Moreover, the reader should bear in mind that the study is based on subjective expert judgement, which may introduce bias, and the use of online tools may limit the depth of the discussion compared to in-person validation panels. However, the instrument can be considered valid in content and ready for use in a pilot study after refinement of all items, particularly those that did not meet the CVR threshold, based on feedback and suggestions from expert evaluators. All 73 items will be measured using the Rasch measurement model to ensure that only items meeting key assumptions are used in the study phase.

Acknowledgements

The author sincerely appreciates the contribution of the expert panels in this study. This study received approval from the Educational Policy Planning and Research Division of the Ministry of Education of Malaysia. It was conducted without specific funding from public, commercial or non-profit organisation. This study was funded by the Faculty of Education, Universiti Kebangsaan Malaysia (UKM) with a Publication Reward Grant (GP-2021-K021854).

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>
- Amatan, M. A., Han, C. G. K., & Pang, V. (2021). Kesahan kandungan soal selidik faktor konteks, input dan proses terhadap penerimaan pelaksanaan elemen pendidikan STEM dalam pengajaran dan pembelajaran guru menggunakan nisbah kesahan kandungan (CVR). *International Journal of Advanced Research in Future Ready Learning and Education*, 23(1), 10–22. <https://akademiabaru.com/submit/index.php/frle/article/view/3923>
- Anggraini, F. D. P., Aprianti, A., Muthoharoh, N. A., IPermatasari, I., & Azalia, J. L. (2023). Validity and reliability questionnaire test of knowledge, attitudes, and behavior on dengue fever prevention. *International Journal on Health and Medical Sciences*, 1(2), 46–54.

- Anizar, Gani, A., Khaldun, I., & Bahi, M. (2018). The development of a module with Microsoft Excel-based interactive media on the topic of buffer solution. *Journal of Physics: Conference Series*, 1088, 1-6. <https://doi.org/10.1088/1742-6596/1088/1/012119>
- Association of College and Research Libraries. (2015). Framework for Information Literacy for Higher Education. American Library Association. <https://www.ala.org/acrl/standards/ilframework>
- Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, 47(1), 79–86. <https://doi.org/10.1177/0748175613513808>
- Baco, S., & Ishak, M. Z. (2021). Kesahan dan kebolehpercayaan instrumen penilaian sendiri pembelajaran ungkapan algebra tingkatan dua. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 6(11), 127–137. <https://doi.org/10.47405/mjssh.v6i11.1136>
- Bhattacharyya, S., Kaur, R., Kaur, S., & Ali, S. A. (2017). Validity and reliability of a questionnaire: A literature review. *Chronicles of Dental Research*, 6(2), 17–24.
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science research. *The Nigerian Postgraduate Medical Journal*, 22(4), 195–201. <https://doi.org/10.4103/1117-1936.173959>
- Cohen, L., Manion, L. & Morrison, K. 2018. *Research Methods in Education*. 6th ed. New York: Routledge.
- Darusalam, G., & Hussin, S. (2021). Metodologi penyelidikan dalam pendidikan: Amalan dan analisis kajian. Penerbit Universiti Malaya.
- Dettori, J. R., & Norvell, D. C. (2020). Kappa and beyond: Is there agreement? *Global Spine Journal*, 10(4), 499–501.
- Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Merrill Prentice Hall.
- Goh, P. S. C., & Ooi, P. K. (2020). Developing global competence in Malaysian universities: Challenges and possible solutions. *International Journal of Higher Education*, 8(4), 256–265.
- GraphPad Software. (2025). Quantify agreement with kappa. Retrieved from <https://www.graphpad.com/quickcalcs/kappa1/>
- Hair, J. F., Celsi, M. W., Ortinau, D. J., & Bush, R. P. (2013). *Essentials of marketing research* (3rd ed.). McGraw Hill.
- Herlinawati H., Marwa M., Ismail N., Junaidi, Liza L.O., Situmorang D.D.B (2024). The integration of 21st century skills in the curriculum of education. *Heliyon*. 10(15): e35148. doi: 10.1016/j.heliyon.2024.e35148.
- Jiixin, G., Huijuan, Z., & Hasan, M. H. (2024). Global competence in higher education: A ten-year systematic literature review. *Frontiers in Education*, 9, Article 1404782. <https://doi.org/10.3389/feduc.2024.1404782>
- Kamaluddin, M. R., Nasir, R., Wan Sulaiman, W. S., Khairudin, R., & Ahmad Zamani, Z. (2017). *Validity and psychometric properties of Malay translated religious orientation scale-revised among Malaysian adult samples*. *Akademika*, 87(2), 133–144.
- Kamaruddin, M., & Mohd Matore, M. E. E. (2021). Development and Validation of Psychometric Properties of the 10 IB Learner Profile Instrument (10IBLP-I): A Combination of the Rasch and Classical Measurement Model. *International Journal of Environmental Research and Public Health*, 18(12), 6455. <https://doi.org/10.3390/ijerph18126455>

- Kopish, M. (2016). Preparing globally competent teacher candidates through cross-cultural experiential learning. *Journal of Social Studies Education Research*, 7(2), 75–108.
- Landis, J.R. & Koch, G.G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33(1): 159-174.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Lindell, M. K., & Brandt, C. J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of CVI, T, rWG(J)g and rWG(J)g indexes. *Journal of Applied Psychology*, 84(4), 640–647. <https://doi.org/10.1037/0021-9010.84.4.640>
- Luqiu, L. R. (2019). In the age of misinformation: The importance of information literacy. In T. T. C. Tan & M. Y. Fee (Eds.), *Information experience in theory and design* (pp. 199–205). Springer. https://doi.org/10.1007/978-981-13-6681-9_15
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386.
- Mohd Matore, M.E.E, Zainal, M.A., Mohd Noh, M.F., Khairani, A.Z., & Abd Razak, N. (2021) The Development and Psychometric Assessment of Malaysian Youth Adversity Quotient Instrument (MY-AQi) by Combining Rasch Model and Confirmatory Factor Analysis. *IEEE Access*, 9, 13314–13329
- Masdoki, M., Din, R. & Mohd Matore M.E.E (2021) Teaching 4.0 competency in higher learning institutions: A systematic mapping review. *International Journal of Learning, Teaching and Educational Research*. 20 (10), 217–231
- Mohamat, R., Sumintono, B., & Abd Hamid, H. R. (2022). Analisis kesahan kandungan instrumen kompetensi guru untuk melaksanakan pentaksiran bilik darjah menggunakan Model Rasch pelbagai faset. *Jurnal Pendidikan Malaysia*, 47(1), 1–14. <https://doi.org/10.17576/JPEN-2022-47.01-01>
- Mohd Noh, M.F & Mohd Matore M.E.E (2022) Rater severity differences in English language as a second language speaking assessment based on rating experience, training experience, and teaching experience through many-faceted Rasch measurement analysis. *Frontiers in Psychology*, 13:941084. doi: 10.3389/fpsyg.2022.941084
- Mopara, R., & Sanrattana, W. (2023). Developing teachers to develop students' 21st century skills. *World Journal of Education*, 13(3), 94–104. <https://doi.org/10.5430/wje.v13n3p94>
- Oktari, R. S., Comfort, L. K., Syamsidik, & Dwitama, P. (2020). Measuring coastal cities' resilience toward coastal hazards: Instrument development and validation. *Progress in Disaster Science*, 5, Article 100065.
- Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, 2(2), 391–400.
- Omar, A., Hamzah, S. A., & Kee, C. P. (2021). Kesahan dan kebolehpercayaan instrumen batas etnik di sekolah kebangsaan Zon Utara Semenanjung Malaysia. *e-BANGI: Jurnal Sains Sosial dan Kemanusiaan*, 18(3), 253–269.
- Organization for Economic Cooperation and Development (OECD). (2018). PISA 2018 global competence. Retrieved from <https://www.oecd.org/en/topics/sub-issues/global-competence/pisa-2018-global-competence.html>
- Parmigiani, D., Jones, S., Silvaggio, C., Nicchia, E., Ambrosini, A., Pario, M., et al. (2022). Assessing global competence within teacher education programs: How to design and create a set of rubrics with a modified Delphi method. *SAGE Open*, 12(4), Article 21582440221128794. <https://doi.org/10.1177/21582440221128794>

- Parmigiani, D., Bar Nir, A., Ferguson-Patrick, K., Forkosh-Baruch, A., Heddy, E., Impedovo, M. A., et al. (2023). Assessing the development of global competence in teacher education programmes: Internal consistency and reliability of a set of rubrics. *Higher Education Pedagogies*, 8(1), Article 2216190. <https://doi.org/10.1080/23752696.2023.2216190>
- 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 and Health*, 30(4), 459–467.
- Potter, W. (2022). Analysis of definitions of media literacy. *Journal of Media Literacy Education*, 14(2), 27–43. <https://doi.org/10.23860/JMLE-2022-14-2-3>
- Ramaila, S., & Molwele, A. J. (2022). The role of technology integration in the development of 21st century skills and competencies in life sciences teaching and learning. *International Journal of Higher Education*, 11(5), 9–17. <https://doi.org/10.5430/ijhe.v11n5p9>
- Rau, G., & Shih, Y. S. (2021). Evaluation of Cohen’s kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes*, 53, Article 101021. <https://doi.org/10.1016/j.jeap.2021.101021>
- Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. Publications Office of the European Union. <https://doi.org/10.2760/159770>
- Sabharwal, C. L. (2021). Cohen’s kappa statistic and new kappa statistic for measuring and interpreting inter-rater agreement. *International Journal of Research in Engineering and Science*, 9(7), 23–28.
- Sabtu, S. H., Mohd Matore, M. E. E., & Maat, S. M. (2025). Enhancing face validity in teaching quality with Six Sigma approach for sustainable development goals (SDG). *SDGsReview*, 5, e03335. <https://doi.org/10.47172/2965-730X.SDGsReview.v5.n02.pe03335>
- Subhaktiyasa, P.G. (2024). PLS-SEM for Multivariate Analysis: A Practical Guide to Educational Research using SmartPLS. *EduLine: Journal of Education and Learning Innovation*, 4(3), 353-365. <https://doi.org/10.35877/454RI.eduline2861>
- Taherdoost, Hamed. (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*, 5(3), 28-36. <https://doi.org/10.2139/ssrn.3205040>.
- Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324–1337. <https://doi.org/10.1080/0144929X.2022.2072768>
- Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1),25-27.
- Yeşilyurt, E., Vezne, R. (2023). Digital literacy, technological literacy, and internet literacy as predictors of attitude toward applying computer-supported education. *Education and Information Technologies*. 28, 9885–9911. <https://doi.org/10.1007/s10639-022-11311-1>
- Zainal, M. A., Mohd Matore, M. E. E., W Musa, W. N., & Hashim, N. H. (2020). Kesahan kandungan instrumen pengukuran tingkah laku inovatif guru menggunakan kaedah nisbah kesahan kandungan (CVR). *Akademika*, 90, 43-54. <https://ejournals.ukm.my/akademika/article/view/42052>
- Zhang, Z., & Garcia, L. (2023). Examining dimensionality and validity of the academic integrity survey instrument. *Journal of Education and Development*, 7(1), 46–54. <https://doi.org/10.20849/jed.v7i1.1326>