Educational Administration: Theory and Practice
2024, 30(5), 12917-12927
ISSN: 2148-2403
https://kuey.net/

Research Article

Face Validity Of Situational Judgement Test For Digital Leadership Using Content Validity Ratio

Nurhafizah Abdul Musid[1], Mohd Effendi Ewan Mohd Matore[2*], Aida Hanim A. Hamid[3]

[1]Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia Email: p113289@siswa. ukm.edu.my
[2*]Research Centre of Educational Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia Email: effendi@ukm.edu.my
[3]Research Centre of Educational Leadership and Policy, Faculty of Education Universiti Kebangsaan Malaysia, Malaysia Email: aidahanim@ukm.edu.my

Citation: Nurhafizah Abdul Musid, Mohd Effendi Ewan Mohd Matore, Aida Hanim A. Hamid (2024) Face Validity Of Situational Judgement Test For Digital Leadership Using Content Validity Ratio *Educational Administration: Theory and Practice*, *30*(5), 12917-12927
Doi: 10.53555/kuey.v30i5.5567

ARTICLE INFO

ABSTRACT

The post-COVID-19 situation has accelerated the growth of a different style of leadership, namely digital leadership. An instrument is needed to measure the level of digital leadership among teachers. Therefore, an instrument called the Situational Judgement Test (SJT) was developed in this study to measure the extent of digital leadership among teachers. Accordingly, a newly developed instrument was to be tested for validity, and one type of commonly assessed validity is face validity. To measure face validity, expert consensus is required and there are a few analyses commonly used for this purpose. This study aims to assess the face validity of SJT among digital teachers using an analysis called Content Validity Ratio (CVR) in order to analyse expert consensus. The reason for this is that analysing face validity using CVR is rarely discussed and there is very little and limited empirical evidence regarding the SJT. This study used a survey method based on a quantitative approach. A total of eight panels were involved using purposive sampling, comprising three primary school teachers, three secondary school teachers, and two other experts who hold a doctorate in education. Three scaled questionnaires were used in this study for face validity using CVR: inappropriate, appropriate but needs improvement, and appropriate. The analysis for this study is based on the CVR formula created using Microsoft Excel. Out of 45 items, only nine should be revised in terms of grammar, sentence structure, and word choice to ensure that the items in the SJT can be understood by the respondents of the actual study. The CVR analysis conducted in this study can provide empirical data on the suitability of an item in terms of language and paragraph structure. The CVR results also provide insight into the relevance of an item from a situational perspective and the selection of responses for this teacher's digital leadership SJT. Nevertheless, it is suggested that a future study be conducted with teachers from all types of schools in Malaysia with other types of validation such as the linguistic validation of SJT for digital leadership among teachers. In addition, other types of expert consensus can also be conducted such as the Fuzzy Delphi Method (FDM), Fleiss Kappa, or Cohen's Kappa to assess the face validity or even linguistic validation of SJT for digital leadership among teachers.

Keywords— *Situational Judgement Test, Digital Leadership, Face Validity, Content Validity Ratio, Teacher.*

1. INTRODUCTION

Assessment is the process of gathering and discussing information from various sources and activities to develop a deep understanding of what students know, understand, and can do with their knowledge based on their educational experiences (Tontus, 2020). An assessment tool is a tool we use to provide evidence of student learning, such as questionnaires and end-of-unit tests (Zhou, 2023). Meanwhile, various techniques can be used for competency assessment such as conceptual maps and/or mind maps, discussions, focus group

discussions, interviews, modelling, narrative methods, observation in action, portfolios, quizzes, simulation, and tests (Darling-Hammond et al., 2013; McConlogue, 2020). However, questionnaires and self-assessment measures are rather prevalent (Darling-Hammond et al., 2013; McConlogue, 2020). Every time an instrument is utilised, validity must be taken into account (Kane, 2016). As such, it is always necessary to establish at least some metrics of an instrument's validity for each new environment in which it is utilised, according to (Knekta et al., 2019).

Validity is the question of whether a measurement tool genuinely yields the intended results (Knekta et al., 2019). For a very long time, the idea of validity in educational assessments has been accepted and explored (Cizek, 2016; Cronbach & Meehl, 1955; Kane, 2016; Messick, 1995; Slaney, 2017). Instead of serving as an instrument's property, validity serves as a property of how that instrument is used in a given situation (Knekta et al., 2019). As defined by American Educational Research Association (AERA) et al. (2014), validity is the extent of the theoretical and evidential support to the planned application of the test score interpretations. Hence, a validated instrument for a particular purpose and population may not guarantee its effectiveness towards all purposes and populations (Knekta et al., 2019).

The usual instrument in social science studies, especially in the field of education, is a questionnaire with a Likert scale. However, the use of questionnaires in research has its limitations. One of the limitations relates to the design phase, where some studies have identified the need to check the consistency between the questions and response options (Pozzo & Borgobello, 2019). Therefore, in the present study, a consensus expert was used to assess the validity of situations and response options of the instrument used in this study, namely the Situational Judgement Test (SJT). Another limitation identified is that questions on perceptions and beliefs indicate the subjective nature of responses, which are often overlooked in some interpretations that relate responses to reality (Pozzo & Borgobello, 2019). Nonetheless, with the SJT, this limitation can be overcome because the SJT offers possible or frequent situations and answer options that are highly likely to occur in the lives of respondents or test takers.

Nowadays, it is very difficult to find an instrument that measures the level of digital leadership among teachers. Most of the instruments for digital leadership focus on administrators such as principals or headmasters in the Malaysian context. Therefore, an initiative was taken in this study and the SJT was developed to assess teachers' digital leadership. Notably, no SJT has been virtually developed to measure digital leadership among teachers. Once the SJT has been developed, it should be tested for validity. One type of validity that is considered important for creating a high-quality instrument is face validity. Expert consensus is a common method for assessing validity. Therefore, this study aims to assess the face validity of SJT for digital teachers using an analysis called Content Validity Ratio (CVR) to analyse expert consensus.

2. LITERATURE REVIEW

In this part of the literature review, important works that relate and connect to the current field of study, such as validity, content validity ratio, SJT and digital leadership, are discussed and presented.

2.1 Validity

The validity of the measurement instrument to measure accurately without confounding it with any other characteristic is defined as "validity" (Sürücü & Maslakçi, 2020). Validity refers to the extent to which the measure actually measures what is intended to be measured (Rozali et al., 2022). In general, validity is determined by the meaningful and appropriate interpretation of data obtained from the measurement instrument as a result of analyses (Sürücü & Maslakçi, 2020). Additionally, validity helps the researcher determine whether a questionnaire is suitable for survey research (Rozali et al., 2022). However, in the present study, the Situational Judgement Test (SJT) is used as the instrument. In this context, validity refers to the degree to which the scale serves its intended use (Sürücü & Maslakçi, 2020).

Validity comprises four types: (i) content validity; (ii) criterion validity; (iii) construct validity; and (iv) face validity (Kuhlmann et al., 2022). The fourth measurement, known as face validity, is frequently used to assess how valid a test seems to methodological newcomers (Mosier, 1947). Face validity describes the tool's suitability, sensitivity, or relevance (Holden, 2010). Meanwhile, according to Taherdoost (2016), face validity is a measure of how closely a measure resembles a given construct. In addition, the definition of face validity by Martinez (2017) characterises a measurement of objects that can be seen as relevant to the topic at hand. Correspondingly, experts must give direct feedback regarding the instrument's presentation, applicability, and clarity in order for face validity to be established (Rashidi et al., 2022).

2.2 Content Validity Ratio

There are several analyses for measuring expert consensus, such as the Fuzzy Delphi Method (FDM), CVR, Fleiss Kappa, and Cohen Kappa. Each analysis has different requirements and conditions that must be met in

order to perform an analysis. In this study, the expert consensus analysis, CVR, is used to assess face validity. The CVR method requires experts to decide whether an item is necessary for measuring the construct measured by the instruments (Ayre & Scally, 2014; Lawshe, 1975; Zamanzadeh et al., 2015). CVR is used to determine the content validity of items through empirical measurements (Zamanzadeh et al., 2015). According to Zakaria et al. (2017) and Zamanzadeh et al. (2015), CVR is one of the commonly used quantitative validity measures to assess the content validity of specific questionnaires. It is more practical in terms of time and cost and is easy to administer and implement (Zamanzadeh et al., 2015).

The CVR is a method used to retain confidence in choosing an instrument's most crucial and accurate content (Zamanzadeh et al., 2015). However, the CVR can also be used to assess an instrument's face validity. For instance, the experts are required to clarify the degree of necessity of a given item to administer a construct in a group of items (Lawshe, 1975) by rating a score between 1 and 3, with the options "not necessary," "useful but not essential," and "essential," for each item, respectively. The range of the CVR is between 1 and -1, and a greater score implies that the experts are in accordance with the importance of a constituent in an instrument. CVR is calculated using CVR = (NeN/2)/(N/2) as the primary formula, such that Ne denotes the number of experts selecting "essential" and N denotes the number of experts in total. In reference to the Lawshe table, the CVR's numerical value is calculated.

2.3 Situational Judgement Test (SJT)

Research instruments are tools used to collect, quantitatively or qualitatively process, and then systematically organise different types of information in research (Sukmawati et al., 2023). Examples of quantitative research instruments include test instruments, non-test instruments, test inventories, and questionnaires (Sukmawati et al., 2023). The present study employs a quantitative approach and the typical instrument used in a social science study, namely a Likert scale questionnaire. Although the Likert scale questionnaire is most commonly employed in research, this study used a Situational Judgement Test (SJT) as the data collection instrument.

The SJT is an assessment format in which the test taker is presented with a series of scenarios depicting an interpersonal situation (Webster et al., 2020). Research that has used the SJT to collect data on digital leadership is rarely discussed. In this study, the SJT shows situations that the respondents, i.e., teachers, are confronted with in the context of digital leadership. The candidate is then required to evaluate several possible behavioural responses to each scenario shown or described (Webster et al., 2020). However, there is a need to understand the SJT's function in evaluating an individual's performance in a situation that is knowledge-based or behaviour-based (Abdul Musid et al., 2022). Therefore, the SJT used in this study is behaviour-based such that the decisions made by the respondents have a high correlation with their personality. The respondents must evaluate four possible actions for each situation. The response format for the SJT is the best single answer; therefore, only one answer is correct for each situation, and the respondents receive one point if they choose the correct answer.

2.4 Digital Leadership

There are some leadership styles in education such as transformational, transactional, innovative, charismatic, and culturally responsive and distributive leadership. However, the post-COVID-19 situation has accelerated the growth of another leadership style, namely digital leadership. In the literature, digital leadership is referred to as a combination of transformational leadership understanding and digital technology competence, which is a more comprehensive form (De Waal et al., 2016). Digital leadership is also referred to as e-leadership and technology leadership. Notably, digital leaders are leaders who have new and creative ideas in the digital domain, able to draw stakeholders' attention to the digital environment in the context of organisational processes and activities, motivate them, maintain organisational continuity by staying in touch with members in the digital domain, and develop digital strategies that could give the organisation a competitive advantage (Balci et al., 2022).

The measurement of a variable depends on the constructs used by the researcher. According to the literature, a variety of constructs can be used to measure digital leadership. In this study, nine constructs of digital leadership were used, all of which were taken from two sources, namely Sheninger (2019) and the International Society for Technology in Education (2017). Specifically, there are seven constructs introduced by Sheninger (2019), namely (i) student engagement, learning and outcomes, (ii) learning environment and spaces, (iii) professional growth and learning, (iv) communication, (v) public relations, (vi) branding, and (vii) opportunities. The two additional constructs are (i) empowered professionals and (ii) learning catalysts, as proposed by the International Society for Technology in Education (2017).

To summarise, it can be deduced that face validity is one of the types of validity. The aim of face validity in this study is to determine whether an item is appropriate and correct in terms of grammar, sentence structure, and word choice. Indeed, face validity is crucial for the respondents to understand the SJT. The analysis chosen to

conduct face validity in this study is the CVR and the critical value of CVR proposed by Lawshe (1975) is based on the number of expert panels. In this study, an instrument called SJT was used, which differs from the Likert-type questionnaire commonly used in social science studies. An item in the SJT consists of situations and choices for the answers. The SJT response format used is contingent on the best single answer. The SJT used in this study deals with digital leadership, which consists of nine constructs, and each construct is very important for measuring teachers' digital leadership.

3. METHODOLOGY

This study used a survey method that is contingent on a quantitative approach. In the present study, the CVR technique was employed to evaluate the face validity of an instrument called the Situational Judgement Test (SJT). The SJT is a procedure that consists of difficult circumstances related to work and various sets of intended actions (Lievens & Motowidlo, 2016; Muck, 2013; Oostrom et al., 2015; Patterson et al., 2016). In this study, the SJT was used to measure teachers' digital leadership. Usually, the survey instrument via a questionnaire is used and measured using a Likert scale (Aithal & Aithal, 2020). However, the present study contributes to the aspect of measurement by analysing the face validity of another type of instrument, namely the SJT, using the CVR. Face validity discussions involving CVR are lacking, as CVR has always been used to analyse content validity (Madadizadeh & Bahariniya, 2023; Rashid et al., 2021). Therefore, this is another contribution to the aspect of measurement.

The research design involves a survey using a quantitative approach. Surveys are increasingly popular in research that aims to collect and analyse the opinions of different groups of people at specific points in time (Zimba & Gasparyan, 2023). This research design is appropriate for this study, which aims to collect the opinions of a group of teachers and lecturers on the items of the SJT for teachers' digital leadership. The respondents will determine the appropriateness of the items in terms of sentence clarity, spelling, and word choice, as this is the goal of face validity in a study. Data collection was done by distributing the questionnaire via email to the face validity panels, and the data collection process was conducted over a two-week period.

3.1 Sampling and Participants

The present study utilised purposive sampling in the selection of relevant panellists, including primary school teachers, secondary school teachers, a lecturer from the Institute of Teacher Education (IPG), and a lecturer from Institut Aminuddin Baki. The total number of panel experts involved is eight. The study employed primary and secondary school teachers because the SJT being promoted was to be answered by primary and secondary school teachers in the actual study. As recommended by Nevo (1985), one of the recommended rater types involves the group being designated to use a particular instrument. The two lecturers selected for the panel have a doctorate; thus, they have sufficient knowledge and skills to evaluate the SJT. This is because face validity is also defined as the extent of the ability of the items assessed on an instrument to correspond to the constructs and objectives of the intended study (Hardesty & Bearden, 2004; Nevo, 1985). Table 1 below presents the details of the respective experts.

TABLE I. DETAILS OF EXPERTS

No.	Panel Experts	Position	Institution
1	A001	Teacher	Secondary school
2	A002	Teacher	Primary school
3	A003	Lecturer	Institute of Teacher Education
4	A004	Teacher	Secondary school
5	A005	Lecturer	Institut Aminuddin Baki
6	A006	Teacher	Secondary school
7	A007	Teacher	Primary school
8	A008	Teacher	Primary school

3.2 Instrument

The instrument consists of nine constructs on the content of digital leadership: (1) student engagement, learning, and outcomes; (2) learning environment and spaces; (3) professional growth and learning; (4) communication; (5) public relations; (6) branding; (7) opportunities (8) empowered professionals; (9) learning catalyst. The situations in the first construct (student engagement, learning, and outcomes) refer to student behaviours that teachers encounter during the teaching and learning process. The second construct, learning environment and spaces, refers to dilemmas related to information communication and technology facilities.

The third construct is professional growth and learning, which is about cultivating the use of information communication and technology as well as sharing knowledge with colleagues. The fourth construct, communication, includes situations such as effective communication skills and methods of information transfer. Subsequently, for the fifth construct, the dilemmas teachers face in public relations relate to effective information dissemination to parents and announcing school programmes to the community. The sixth construct, branding, encompasses situations involving publicising the school in the local area and engaging with parents.

Opportunity is the seventh construct of digital leadership, which highlights the opportunity to promote students' digital technology skills and improve their learning environment. The eighth construct is empowered professionals, which is about sharing knowledge, skills, and experience in educational technology and endeavouring to keep up with developments and improve digital technology skills. The final construct is the learning catalyst. The dilemmas in the learning catalyst involve the ways to improve student engagement in the teaching and learning process, as well as how to create documentation for sharing with other teachers.

In this study, a CVR questionnaire consisting of 45 items and three scales was used: (1) inappropriate, (2) appropriate but in need of improvement, and (3) appropriate. The *inappropriate* scale means that the item is not suitable for measuring the digital leadership skills of teachers. Meanwhile, the *appropriate but needs improvement* scale refers to an item that is suitable for assessing digital leadership; however, this item needs to be modified in terms of the wording and content of the situations or response options. Finally, the *appropriate* scale indicates that the item is suitable and perfect for the situations or response options. The experts on the panel were given questionnaires to complete according to the three approval scales. Once completed, the instruments were then returned to the researcher via email.

3.3 Analysis of Data

Essentially, the research data analysis is contingent on the CVR formula created using Microsoft Excel. The formula for CVR calculation is as follows:

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

Where

ne = number of expert panels that rated the item as essential
N = total number of expert panels involved

According to Lawshe (1975), CVR values range from -1 to +1. A value of +1 means that the element evaluated by the expert panel is important to the validity of the content. If the CVR value is <0, then less than half of the expert panel rated the item as essential. A CVR value of 0 means that part of the expert group rated the item as unimportant, and another part rated it as essential. However, CVR>0 indicates that more than half of the expert panel rated the item as essential. If the CVR value is 1, then all experts agreed to rate the item as essential. However, in this study, the essential scale was converted to an appropriateness scale, as mentioned in the previous subtopic. Based on the feedback gathered from the eight-member expert panel, the minimum value of 0.693 at a significance level of 0.05 for a two-sided test was used to calculate the CVR (Wilson et al., 2012). However, the critical value of the CVR in this study is 0.750, according to the value suggested by Lawshe (1975), with a number of eight experts.

4. Results and Discussion

Table 2 shows the CVR values for 45 items in the SJT for digital leadership among teachers. Table 2 shows that nine items had a CVR value of 0.500, 22 items had a CVR value of 0.750, and 14 items had a CVR value of 1.000. As explained in the previous subsection, the critical value of the CVR is 0.750, as there were eight experts. Therefore, 36 items were accepted as items in the SJT for digital leadership among teachers. The remaining nine items should be revised as they did not reach the critical value. These items need to be reviewed in terms of words used in the sentence, spelling, choice of answers, and incomplete sentences. Most importantly, the items must be revised so that they can be included in the SJT.

TABLE II. CVR Values and Item Status

No.	Construct	No. of Items	CVR Value	Item Status
1	Student engagement, learning, and outcomes	1	0.750	Accepted
		2	1.000	Accepted
		3	1.000	Accepted
		4	1.000	Accepted
		5	0.750	Accepted
2	Learning environment and spaces	1	0.750	Accepted
		2	0.750	Accepted
		3	0.500	Revised
		4	1.000	Accepted
		5	0.750	Accepted
3	Professional growth and learning	1	0.750	Accepted
		2	0.750	Accepted
		3	0.500	Revised
		4	0.750	Accepted
		5	0.750	Accepted
4	Communication	1	0.750	Accepted
		2	0.750	Accepted
		3	1.000	Accepted
		4	0.750	Accepted
		5	0.500	Revised
5	Public relations	1	0.750	Accepted
		2	0.500	Revised
		3	1.000	Accepted
		4	0.750	Accepted
		5	1.000	Accepted
6	Branding	1	0.500	Revised
		2	0.750	Accepted
		3	0.750	Accepted
		4	1.000	Accepted
		5	1.000	Accepted
7	Opportunity	1	1.000	Accepted
		2	0.750	Accepted
		3	0.750	Accepted
		4	1.000	Accepted
		5	1.000	Accepted
8	Empowered professionals	1	0.750	Accepted
		2	0.750	Accepted
		3	0.500	Revised
		4	1.000	Accepted
		5	0.750	Accepted
9	Learning catalyst	1	0.500	Revised
		2	0.750	Accepted
		3	0.500	Revised
		4	0.500	Revised
		5	1.000	Accepted

All five items for the construct of *student engagement, learning, and outcomes* were accepted. However, the panel provided some constructive comments on the items, such as the response options for Item 1, Item 3, and Item 4, which required improvement. This is because some response options for these items were not suitable. The panel also suggested other response options to replace the current response options; therefore, these

suggestions were considered in this study. Meanwhile, the comment for Item 3 relates to the term between *Pengajaran dan Pembelajaran* (PdP) and *Pengajaran dan Pemudahcaraan* (PdPc). For this study, we decided to use the term *Pengajaran dan Pembelajaran* (PdP), as this term is frequently used in the official documents of the Ministry. In addition, the panel suggested that some terms be explained in detail first. The reason for this is that some teachers were not familiar with the terms. One of the experts also opined that the wording of the situation is incomplete. Thus, the expert suggested that an elaboration be made on how to guide the students to use the computer.

The second construct is *learning environment and spaces*. Out of the five items proposed in this construct, one must be revised, namely Item 3. This is because three response options were labelled as inappropriate; therefore, a major correction was needed for this item. The response choices offered must be relevant. This is because candidates are often given the opportunity to respond to an explanation of a hypothetical job-related event by evaluating several answers according to how likely they are and which is most appropriate under the circumstances (Mumford, 2015). Other points must also be improvised, including the structure and choice of words that should be changed. For example, the word 'membimbing' should be used to replace 'memberi tunjuk ajar.' One of the experts opined that 'membimbing' sounds more appropriate and should be used in the situation in Item 5. Meanwhile, one of the response options in Item 5 must be improvised because the sentence was too vague and should be clarified.

The third construct is *professional growth and learning*. The results showed that Item 3 must be revised. The panel suggested improving the verbs used in the response options to make the items simple and easy to understand. Some of the response options must be changed because the two response options for Item 3 were similar. For the remaining items, the response choices should be revised to make the meaning of the sentence clearer. If the meaning of the sentence is unclear or vague, then the test takers may misunderstand the verse in question. As a result, they may interpret the sentence according to their own understanding rather than choosing an answer that describes their respective actions or the correct action in the measured situation. As Patterson et al. (2016) explained, the response specifications for SJTs are categorized into either the knowledge-based category, i.e., the best answer and/or the most optimal response, or the behaviour-based category, i.e., the most likely taken action in a given situation.

Next, *communication* is the fourth construct of digital leadership. According to the results, Item 5 must be revised. More specifically, the verb should be changed and a preposition should be added to one of the response options to further improve the sentence. Additionally, one term in one of the response options, 'excellent teacher,' should be changed to 'expert.' Since communication is the purpose of writing (Prior, 2006), it is imperative that writers first identify their audience in their work so that they can thoughtfully evaluate the syntactic complexity of their writing as well as the language used to express their ideas (Hayes & Flower, 1980). Another comment given by the panel involves the response options for Item 5. One of the experts opined that the response options should be improvised because some of them were not relevant to the given situation. Therefore, the suggestions for the situation and response options for Item 5 were accepted and revised by the researchers.

The fifth construct of digital leadership is *public relations*. Based on the results, Item 2 must be revised with some changes regarding the verb used in the situation. In addition, the word used to represent the dilemma in the situation was deemed inappropriate; thus, the panel suggested deleting the word to show dilemma. In SJT scenarios, the candidate must make a series of complicated decisions about how to deal with certain difficult interpersonal problems (Patterson et al., 2016). Other comments were that some response options should be worded more clearly because they were confusing. Since the developed SJT is intended for both primary and secondary school teachers, the panel suggested revising the response options to make it suitable for both types of schools. Therefore, the sentence should be reviewed and improved to make it easier for the respondents to understand the response options.

Another construct of digital leadership is *branding*. Of the five items proposed for this construct, one must be revised based on the opinion of the panel, namely Item 1. In the panel's opinion, one word should be deleted in the given situation because it occurred twice in one sentence, which makes the meaning of the verse incomprehensible. The sentence structure for Items 2 and 3 also required improvement. As pointed out by Hayes (2004), writing requires meticulous evaluation of ideas throughout pages, paragraphs, and even sentences to ensure coherency in a way that serves the writing purpose. However, no comments were given for Items 4 and 5. Nevertheless, one of the experts commented on the correct spelling of soft copy, whether it encompasses one word or two words. During the review process, both one-word and two-word spellings were accepted; however, the most commonly used spelling includes two words. Therefore, the researchers decided to retain the two-word spelling, i.e., soft copy.

The seventh construct of digital leadership is *opportunity*. All items in this construct were accepted by the panel; however, some comments were made by the panel to improve the items. For example, the fourth

response option in Item 2 should be more precise. As stated by Philippakos (2019), the sentence structure is cognitively and linguistically demanding. Therefore, it was revised to make it more precise. Other comments involved the second and third response options in Item 3. The panel somewhat agreed that these response options seemed similar. Therefore, the panel suggested changing one of the response options to 'online courses.' However, after careful consideration, the researchers decided to retain the two response options. The reason for this is that both possible answers suggest different approaches to solving the situation. There were no comments from the panel on the remaining items.

The eighth construct is *empowered professionals*. Based on the results, Item 3 must be revised. Some experts stated that the first response option was vague and should be related to the situation, while the second and third response options seemed irrelevant as response options. The corrections to be made in the other items involved replacing words with prepositions and checking the nouns used. As explained by Hayes and Flower (1980) and Hudson (2016), both writing clarity and reader comprehension can be influenced by grammar. Another point on which the panel made a comment involved the fourth response option in Item 1. The comment refers to the use of the preposition 'mengenai' (regarding). Alternatively, it was suggested by the panel that the preposition be changed to 'tentang' (about). The suggestion was then accepted by the researchers and the fourth response option in Item 1 was, therefore, changed.

The final construct of digital leadership is *learning catalyst*. For this construct, three items must be revised, namely Item 1, Item 3, and Item 4. For Item 1, one of the response options seemed somewhat irrelevant. Thus, corrections were made to this particular response option to make it more appropriate and relevant. Meanwhile, some experts thought that the situation for Item 3 was somewhat lengthy and should be revised. Consequently, the situation was shortened to save time for the respondents to read and understand. Furthermore, one of the response options seemed to contradict the given situation. However, after a careful review of the situation, the researchers decided to retain the response option. The same comment was made about Item 4, which was deemed a bit lengthy, and one of the response options should be replaced by using the cloud to store teaching materials. The risk of data loss becomes almost irrelevant with the use of cloud storage (Akingbade, 2016). Therefore, the situation in Item 4 was shortened and the response option was changed according to the suggestions of the panel of experts.

5. CONCLUSIONS

This study was conducted to assess the face validity of the SJT for measuring digital leadership among teachers. The face validity of an instrument is important because it determines whether an item is appropriate and correct in terms of grammar, sentence structure, and word choice. Face validity is also important to ensure that the respondents or test takers understand all items in the SJT and are able to answer them without confusion. The expert consensus analysis chosen for this study is CVR, which has provided solid empirical evidence confirming the validity of the SJT for digital leadership. The results have shown that 36 items of the SJT were accepted for measuring digital leadership. In the meantime, nine items must be revised further. However, it should be noted that the experts made very few suggestions regarding the 36 accepted items. To summarise, some suggestions were accepted and some were rejected after careful consideration by the researchers. In this paper, explicit explanations of the CVR were given, starting with the formula used, the scale based on the CVR value, and the critical CVR value for eight panels. This paper has also explained another dimension of the CVR associated with the SJT, as CVR analysis is normally used for Likert-scale questionnaires. The results showed that CVR analysis is also suitable for face validity and not only content validity.

6. IMPLICATIONS AND LIMITATIONS

The research findings have some important contributions. This study contributes greatly to the aspect of digital leadership and measurement. One leadership style that has been attracting a lot of attention recently is digital leadership, as most industries are moving towards digitalisation. Therefore, this study provides nine relevant constructs for measuring digital leadership among teachers. From the perspective of measurement, there is almost no SJT developed for measuring digital leadership among teachers; hence, this study provides an instrument that measures digital leadership among teachers, namely the SJT, with questionnaires being the usual instrument in social science studies. The use of SJT as an instrument in social sciences, especially in education, is considered new. In addition, conducting CVR analysis is rather challenging because an SJT item is longer than that of a Likert-scale questionnaire. This is because an item in the SJT consists of one situation and four response options. Therefore, the experts spent more time evaluating one situation and four response options for each item in the SJT. Moreover, the CVR analysis performed in this study can provide empirical data on the suitability of an item in terms of language and paragraph structure. The CVR results also provide information about the relevance of an item from a situational perspective and the selection of responses for the SJT on teachers' digital leadership. However, this study has one limitation, as it only includes teachers from certain types of schools. Thus, it would be more interesting if feedback could be obtained from teachers of all

types of schools in Malaysia. Accordingly, it is suggested that a future study involving teachers from all types of schools in Malaysia is conducted with another type of validation, such as linguistic validation of the SJT for teachers' digital leadership. In addition, other types of expert consensus can also be employed such as the FDM, Fleiss Kappa, or Cohen's Kappa to assess the face validity or linguistic validation of the SJT for digital leadership among teachers.

CONFLICT OF INTEREST

The authors certify that none of their known personal or financial conflicts might have an impact on the work in this article.

CONTRIBUTIONS OF AUTHORS

Conceptualization, Nurhafizah Abdul Musid; design, Mohd Effendi @ Ewan Mohd Matore; drafting manuscript, Nurhafizah Abdul Musid and Mohd Effendi @ Ewan Mohd Matore; writing, Nurhafizah Abdul Musid and Mohd Effendi @ Ewan Mohd Matore; critical version of manuscript, Mohd Effendi @ Ewan Mohd Matore and Aida Hanim A. Hamid; supervision, Mohd Effendi @ Ewan Mohd Matore and Aida Hanim A. Hamid. The published version of the manuscript has been carefully read and agreed upon by the authors.

FUNDING

The funding for the study was provided by the Faculty of Education, National University of Malaysia (UKM), under GP-2021-K021854 (Publication Reward Grant), GG-2022-020 (Research Fund of FPEND) and TAP-K021854 (Tabung Agihan Penyelidikan).

ACKNOWLEDGEMENTS

I am indebted to both supervisors, Associate Professor Ts Dr. Mohd Effendi @ Ewan Mohd Matore and Dr. Aida Hanim A. Hamid, for their commitment to this research. I would also like to thank the University Research Group (KPU) Educational Evaluation UKM for their excellent support and cooperation throughout the entire research process.

REFERENCES

1. Abdul Musid, N., Mohd Matore, M. E. E., & A. Hamid. A. H. (2022). The Prospective of Situational Judgement Test in Assessing Individual Performance. *International Journal of Academic Research in Progressive Education and Development*, *11*(3), 1810–1819. https://doi.org/DOI:10.6007/IJARPED/v11-i3/15401
2. Aithal, A., & Aithal, P. S. (2020). Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-based Statistical Approach. *International Journal of Management, Technology, and Social Sciences*, *5*(2), 233–251.
3. Akingbade, L. O. (2016). Cloud Storage Problems, Benefits and Solutions Provided by Data De-duplication. *International Journal of Engineering Science and Innovative Technology*, *5*(6), 70–77.
4. American Educational Research Association (AERA), American Psychological Association (APA), & National Council for Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
5. 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
6. Balci, E. V., Tiryaki, S., Demir, Y., & Baloğlu, E. (2022). Digital Leadership on Twitter: The Digital Leadership Roles of Sports Journalists on Twitter. *International Journal of Organizational Leadership*, *11*(Special Issue 2022), 21–35. https://doi.org/10.33844/ijol.2022.60337
7. Cizek, G. J. (2016). Validating Test Score Meaning and Defending Test Score Use: Different Aims, Different Methods. *Assessment in Education: Principles, Policy and Practice*, *23*(2), 212–225. https://doi.org/10.1080/0969594X.2015.1063479
8. Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957
9. Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Jamal, J., Aber, J. L., Baker, E., Bennett, R., Gordon, E., Haertel, E., Hakuta, K., Ho, A., Lee Linn, R., Pearson, P. D., Popham, J., Resnick, L., Schoenfeld, A. H., Shavelson, R., Shepard, L. A., … Steele, C. M. (2013). *Criteria for High-Quality Assessment*. https://edpolicy.stanford.edu/%0Alibrary/publications/847
10. De Waal, B., Van Outvorst, F., & Ravesteyn, P. (2016). The Objective-Subjective Dichotomy of Technology Revisited. *12 Th European Conference on Management, Leadership and Governance*, *November*, 52–61.

11. Hardesty, D. M., & Bearden, W. O. (2004). The Use of Expert Judges in Scale Development Implications for Improving Face Validity of Measures of Unobservable Constructs. *Journal of Business Research*, *57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8
12. Hayes, J. R. (2004). What Triggers Revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision, Cognition, and Instructional Processes* (pp. 9–20). Kluwer Academic Publishers.
13. Hayes, J. R., & Flower, L. (1980). Identifying the Organization of Writing Processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Lawrence Erlbaum.
14. Holden, R. R. (2010). Face Validity. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (Edisi ke-4, pp. 1–2). John Wiley & Sons. https://doi.org/10.1007/978-1-4419-1153-7_200212
15. Hudson, R. (2016). Grammar Instruction. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (Edisi ke-2, pp. 288–300). Guilford.
16. International Society for Technology in Education. (2017). *ISTE Standards for Educators*. https://www.wcupa.edu/education-socialWork/assessmentAccreditation/documents/ISTE_Standards_For_Educators.pdf
17. Kane, M. T. (2016). Explicating Validity. *Assessment in Education: Principles, Policy and Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192
18. Knekta, E., Runyon, C., & Eddy, S. (2019). One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. *CBE Life Sciences Education*, *18*(1), 1–17. https://doi.org/10.1187/cbe.18-04-0064
19. Kuhlmann, L., Teo, K., Olesen, S. S., Phillips, A. E., Faghih, M., Tuck, N., Afghani, E., Singh, V. K., Yadav, D., Windsor, J. A., & Drewes, A. M. (2022). Development of the Comprehensive Pain Assessment Tool Short Form for Chronic Pancreatitis: Validity and Reliability Testing. *Clinical Gastroenterology and Hepatology*, *20*(4), 770–783. https://doi.org/10.1016/j.cgh.2021.05.055
20. Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/https://doi.org/10.1111/j.1744-6570.1975.tb01393.x
21. Lievens, F., & Motowidlo, S. J. (2016). Situational Judgment Tests: From Measures of Situational Judgment to Measures of General Domain Knowledge. *Industrial and Organizational Psychology*, *9*(1), 3–22. https://doi.org/10.1017/iop.2015.71
22. Madadizadeh, F., & Bahariniya, S. (2023). Tutorial on How to Calculating Content Validity of Scales in Medical Research. *Perioperative Care and Operating Room Management*, *31*(2), 100315.
23. Martinez, L. S. (2017). Validity, Face and Content. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc. https://doi.org/10.4135/9781483381411.n651
24. McConlogue, T. (2020). *Assessment and Feedback in Higher Education: A Guide for Teachers*. UCL Press.
25. Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences from Person's Responses and Performances as Scientific Inquiry into Score Meaning. *American Psychologist*, *50*(9), 741–749. http://psycnet.apa.org/journals/amp/50/9/741.pdf&uid=1996-10004-001&db=PA
26. Mosier, C. I. (1947). A Critical Examination of the Concepts of Face Validity. *Educational and Psychological Measurement*, *7*(2), 191–205.
27. Muck, P. M. (2013). Entwicklung Von Situational Judgement Tests. *Zeitschrift Für Arbeits- Und Organisations Psychologie A&O*, *57*(4), 185–205.
28. Mumford, S. (2015). *The Situational Judgment Test: Cognition, Constructs and Criterion Validity*. University of Sheffield.
29. Nevo, B. (1985). Face Validity Revisited. *Journal of Educational Measurement*, *22*(4), 287–293. https://doi.org/https://psycnet.apa.org/doi/10.1111/j.1745-3984.1985.tb01065.x
30. Oostrom, J. K., De Soete, B., & Lievens, F. (2015). Situational Judgment Testing: A Review and Some New Developments. In I. Nikolaou & J. K. Oostrom (Eds.), *Employee Recruitment, Selection and Assessment: Contemporary Issues for Theory and Practice* (pp. 172–189). Psychology Press.
31. Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational Judgement Tests in Medical Education and Training: Research, Theory and Practice: AMEE Guide No.100. *Medical Teacher*, *3*(1), 3–17.
32. Philippakos, Z. A. T. (2019). Sentence Construction: Supporting Elementary Students' Editing Skills. *The Language and Literacy Spectrum*, *29*(1), 1–22.
33. Pozzo, M. I., & Borgobello, A. (2019). Using Questionnaires in Research on Universities: Analysis of Experiences from a Situated Perspective. *REIRE Revista d'Innovació i Recerca En Educació*, *12*(2), 1–16. https://doi.org/10.1344/reire2019.12.227010
34. Prior, P. (2006). A Sociocultural Theory of Writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 54–66). Guilford Press.
35. Rashid, S. N. M., Osman, N., & Mohamad. L. Z. (2021). Content Validity of Drug Addiction Recovery Test Instruments Using Content Validity Ratio (CVR) Method. *Journal of Drug Delivery and Therapeutics*, *11*(4), 24–29. https://doi.org/10.22270/jddt.v11i4.4949
36. Rashidi, S. N., Majid, F. A., & Hashim, H. (2022). Building Soft-Employability Skills (SES-KIT): Reliability, Face Validity and Content Validity Testing. *International Journal of Academic Research in*

Progressive Education and Development, *11*(3), 638–649. https://doi.org/10.6007/ijarped/v11-i3/14546
37. Rozali, M. Z., Puteh, S., Yunus, F. A. N., Hamdan, N. H., & Latif, H. F. M. (2022). Reliability and Validity of Instrument on Academic Enhancement Support for Student-Athlete Using Rasch Measurement Model. *Asian Journal of University Education*, *18*(1), 290–299. https://doi.org/10.24191/ajue.v18i1.17199
38. Sheninger, E. (2019). *Digital Leadership: Changing Paradigms for Changing Times* (Edisi ke-2). SAGE Publications Inc.
39. Slaney, K. (2017). Construct Validity: Developments and Debates. In *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions* (pp. 83–109). Palgrave Macmillan.
40. Sukmawati, Sudarmin, & Salmia. (2023). Development of Quality Instruments and Data Collection Techniques. *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar*, *6*(1), 119–124. https://doi.org/10.55215/jppguseda.v6i1.7527
41. Sürücü, L., & Maslakçi, A. (2020). Validity and Reliability in Quantitative Research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726. https://doi.org/10.15295/bmij.v8i3.1540
42. 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*, *5*(3), 28–36. https://doi.org/10.2139/ssrn.3205040
43. Tontus, H. O. (2020). Concept of Assessment and Evaluation. In *Glossary of Assessment & Evaluation in Higher Education* (pp. 11–17). Livre de Lyon.
44. Webster, E. S., Paton, L. W., Crampton, P. E. S., & Tiffin, P. A. (2020). Situational Judgement Test Validity for Selection: A Systematic Review and Meta-Analysis. *Medical Education*, *54*(10), 888–902. https://doi.org/10.1111/medu.14201
45. Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe ' s Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/0748175612440286
46. Zakaria, M. N., Lau, Y. J., Husain, W. S. W., Aw, C. L., Othman, N. A. N., Salim, R., Abdullah, B., Zainun, Z., & Mohamad, W. N. W. (2017). Development and Psychometric Validation of a New Tinnitus Questionnaire for Clinical Use. *Auditory and Vestibular Research*, *26*(2), 71–85. http://avr.tums.ac.ir/index.php/avr/article/view/178
47. Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. (2015). Design and Implementation Content Validity Study: Development of an Instrument for Measuring Patient-Centered Communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017
48. Zhou, Y. H. (2023). Tools in Classroom-Based Formative Assessment. *Open Access Library Journal*, *10*(e10826), 1–9. https://doi.org/10.4236/oalib.1110826
49. Zimba, O., & Gasparyan, A. Y. (2023). Designing, Conducting, and Reporting Survey Studies: A Primer for Researchers. *Journal of Korean Medical Science*, *38*(48), 1–11. https://doi.org/10.3346/jkms.2023.38.e403