DOI Link: https://doi.org/10.61586/MR6Lq
Vol.65, Issue.10, Part.1, Oct 2025, PP.87-101

# Analysis of Validity and Reliability of the visual, auditory and kinaesthetic skills Instrument

**Mohd Fikri Basiron[1], Hafizhah Zulkifli[2*]**

[1]Faculty of Education, National University of Malaysia. Bangi, Selangor, Malaysia
[2]Faculty of Education, National University of Malaysia. Bangi, Selangor, Malaysia
* Corresponding Author: hafizhah_zulkifli@ukm.edu.my

Received Feb 2025 Accepted Sep 2025 Published Oct 2025

**ABSTRACT**
As the educational environment continues to shift toward digitalisation, students are required to develop a diverse range of skills to keep pace with global advancements. This study aims to determine the validity and reliability of a diagnostic instrument designed to assess visual, auditory, and kinaesthetic learning skills. The evaluation process ensured that all items were aligned with the study objectives and conceptually appropriate. Five experts assessed the instrument in terms of content validity, linguistic accuracy, and item construction, while 38 fifth-year students participated as the sample for reliability testing. The Content Validity Ratio (CVR) was used to measure expert consensus on item acceptability, where higher values indicated strong agreement. Given the dichotomous nature of the items, the Kuder–Richardson Formula 20 (KR-20) was applied to determine reliability. The results revealed a reliability coefficient of 0.70, indicating satisfactory internal consistency. Overall, the findings confirm that the diagnostic test is a valid and reliable tool for assessing students' learning styles and can be further refined for broader application in future studies.
**Keywords:** Pre-Post Test Instruments, Validity, Reliability, Content Validity Ratio (CVR), Kuder Richardson-20 (KR-20)

## 1. INTRODUCTION

A research instrument is a means for gathering, analysing, assessing, and interpreting data about a certain situation or topic (Elangovan and Sundaravel, 2021). Furthermore, the research instrument serves as an assessment procedure, with the goal of drawing general conclusions based on the data (Verawati et al., 2020). Campbell and Stanley (1963) described research instruments as modifications to measuring methodologies used to produce a variety of grades, accomplishments, or study outcomes. Creswell (2018) distinguishes between two types of research instruments: quantitative and qualitative research designs. This is because quantitative studies often utilise numerical scales and closed questions, whereas qualitative studies use word interpretation and open questions to evaluate the study hypothesis (Creswell, 2018). As a result, the research instrument serves as a testing tool for the research hypothesis, allowing the data to be generalised, and interpreted using descriptive or inferential analysis. The next step in producing excellent instruments is to ensure their validity and dependability.

According to Tobón and Luna-Nemecio (2021), every instrument developed must go through a validation and reliability procedure before it can be utilised successfully. Validity is the process of evaluating and considering the theoretical justification and empirical evidence of the measuring tool, score, or assessment that is employed (Bergsland and Vogt, 1962). Furthermore, validity refers to the degree to which the instrument properly measures the item, is consistent with the construct being tested, and can be described and interpreted. Validity may be classified into four types: construct validity, content validity, criterion validity, and predictive validity. All of the types of validity must be based on professional assessment and statistical analysis (Ahmed & Ishtiaq, 2021). Reliability refers to the process of assessing the consistency and repeatability of a manufactured device (Bruton et al., 2000). This is due to low reliability levels, which have an impact on the instrument's consistency. Ahmad and Ishtiaq (2021) stated that reliability is classified into three categories: test-retest reliability, internal reliability, and inter-rater reliability.

Content validity entails the review of experts chosen based on their respective professions. Since there are several approaches for establishing an instrument's validity, researchers must first determine the scale utilised to acquire expert confirmation (Elangovan and Sundaravel, 2021) due to differences in computation methods and formula use. As a student's visual, auditory, and kinaesthetic abilities diagnostic test includes three measurement scales, the researcher relies on the Content Validity Ratio (CVR) score to establish the instrument's validity. The measuring scale chosen is (i) extremely important, (ii) useful but less important, and (iii) unimportant. Thus, Lawshe (1975) suggested that the three-scale instrument's content validity should be calculated using the CVR approach to determine the value of expert agreement on the developed items. Each item's content validity cannot be calculated using the Content Validity Index (CVI) value since four scales are the minimum scale for the CVI calculation technique (Muhamad Saiful Bahri Yusoff, 2019). Furthermore, the CVI computation will split the scales 1 and 2, which represent the value 0, while the scales 3 and 4 represent the value 1 in the CVI formula calculation (Muhamad Saiful Bahri Yusof, 2019).

Next, reliability highlights the need for consistency in items used to assess something to be tested (Sürücü and Maslakçi, 2020). According to the Cronbach Alpha formula, items on a Likert scale should have a reliability value greater than 0.70. As a result of the test's dichotomous scale, the Cronbach Alpha formula cannot be used to assess its reliability rating. Therefore, the Kuder-Richardson 20 (KR-20) formula is well-suited for assessing item consistency. According to Kuder and Richardson (1937), a score of 0.70 or above shows that the test item has a significant connection and may be used as a research instrument. In any event, if the value falls below 0.50, the item should be discarded or amended for improvement (Pienaar et al., 2021). The correct measurement technique of the research variables is the consequence of the instrument's strength developed by the validity and reliability processes (Mohd Hussin Musa et al., 2020).

Studies on the validity and reliability of research tools have been published in many fields. Among these are the studies conducted by Fenny Farhana et al. (2023), Subramaniam Raman et al. (2023), Vinoothini Mannogaran & Nor Azwahanum Nor Shaid (2023), and Mohd Taufek Harun and Mohd Ikhwan Hadi Yaacob (2021). Most current studies employ CVI validity and reliability tests, Cronbach's alpha coefficient values, and Rasch measurement models. Nevertheless, this study uses the CVR value for the validity test and the KR-20 value for the reliability test. These two measuring tools were selected to fulfil the calculation process's criteria based on the scale utilised, namely three expert assessment scales for the validity process and a dichotomous scale for the diagnostic test research instrument in the reliability phase. As a result, effective instruments for measuring students' skills must go through a validity and reliability procedure to ensure that they satisfy the context, criteria, and items being tested. Furthermore, this study addresses inadequacies in the assessment of student competencies using measuring instruments from various viewpoints and approaches.

In conclusion, this study examines the validity and reliability of the visual, auditory, and kinaesthetic diagnostic test instruments for the Sirah Virtual Reality (VR) Module. The development of student abilities is critical to students' survival in today's more complex and tough educational environment. As a result, there is a need for the development of instruments that can assess students' skills in order to generate various educational interventions and innovations that address challenges of learning and teaching approaches in the classroom. As a result, two objectives were set for this study: to evaluate the validity of the instrument's content for identifying students' visual, auditory, and kinaesthetic skills based on the CVR value and to assess the instrument's reliability using the KR-20 value.

## 2. LITERATURE REVIEW

The Industrial Revolution 5.0 (IR 5.0) marks the advancement of technological integration in everyday life, including artificial intelligence (AI), the Internet of Things (IoT), and robots, as well as the creation of sophisticated software (Mohammad Saffuan Mat Saad, 2020). Furthermore, the technological advancement has the potential to affect human existence in terms of more efficient administration and cost savings, as well as increase overall well-being. However, IR 5.0 cannot be completely fulfilled unless the community is knowledgeable and

possesses a wide set of abilities. As a result, the skills agenda of Society 5.0 cannot be met unless application and early exposure in education are provided from the start of school. In Malaysian education, the Malaysian Ministry of Education (KPM) is the primary driving force behind training students to be highly competent and competitive on a worldwide scale. Learning materials that employ technology tools or software will almost probably include abilities that engage the primary senses, namely sight, hearing, and movement (Kayalvili Munusamy and Khairul Azhar Jamaludin, 2022). As a result, visual, auditory, and kinaesthetic abilities become top priorities for student mastery.

**2.1 Students' Visual, Auditory and Kinesthetic Skills**

Visual abilities are an individual's capacity to perceive, assess, coordinate, and comprehend what is observed through the sense of sight (Norfishah Mat Rabi, 2021). Gardner and Moran (2006) describe visual skills as a person's capacity to orient and control space and three-dimensional objects while conveying precise meaning. Fleming and Baume (2006), on the other hand, suggest that the usage of diagrams, charts, maps, graphs, symbols, and images demonstrates a desire to visualise. Visual skills are essential for students' academic development in order for the learning process to go smoothly. Several studies have shown that improving visual skills enhances students' ability to add operations in the field of mathematics (Soon Ping Sia & Kamariah Abu Bakar, 2022), visual arts education (Muhamad Firdaus Ramli & Rofidah Musa, 2020), and cognitive and affective domains (Siti Hayati binti Mohd Yusoff and Halim bin Husain, 2021). This demonstrates the favourable impact of improving students' visual abilities. Kadam et al. (2021) found that improving students' visual abilities resulted in a 10% improvement in their ICT mastery. As a result, visual skills are observed to work in parallel with attempts to increase student performance. The second focus of this study is on the students' auditory skills. Barbe and Milone (1981) defined auditory skills as a person's capacity to use their hearing to get information in the form of audio content, speech, or music. Gardner and Moran (2006) feel that persons with auditory abilities will be more aware of rhythm, verbal intonation, and even melody. Several research studies have indicated that enhancing auditory talents encourages the establishment of a community of inquiry (Setyosari et al., 2020), increases language skills (O'Brien and Yeatman, 2021), and improves students' reading capabilities (Fitri and Gani, 2020). Cohen's research (1980) discovered that, on average, each person spends 45% of his daily time listening, 30% speaking, 16% reading, and 9% writing, suggesting that auditory talents are the most important skills for all people in their daily lives. As a result, students must enhance and master their hearing talents, especially because the world of education is today dominated by digitisation and inquiry activities. Kinaesthetic abilities refer to a person's capacity to understand information based on physical activity, whether it includes gross or fine motor movements (Gardner and Moran, 2006). Kinaesthetic skills vary from visual and auditory skills in that all motions or actions may happen in any setting and involve various body parts. Furthermore, Saeed et al. (2015) claimed that students with kinaesthetic abilities prefer sensory stimulation activities and tend to lose concentration if no hands-on activities are provided during the learning session.

Several studies propose various methods for improving students' kinaesthetic skills, including outbound activities, also known as experimental activities (Hajar et al., 2020); dance art activities (Aini et al., 2023); and exercise activities (Rosmayanti and Santana, 2021). Nonetheless, Hastuti's (2020) study indicated no significant difference between reading activities in terms of students' kinaesthetic and auditory skills.

**2.2 Development Process**

The instrument development process must be based on many theories or models to ensure that the structures and items employed fulfil the testing criteria. The Coaley Model (2009), the McIntire and Miller Model (2007), the Gregory Model (2007), the Hogen Model (2007), the Murphy and Davidshofer Model (2005), the Cohen, Swerdlik, and Sturman Model (2002), the Gable and Wolf Model (1993), and the Brown Model (1983) are examples of instrument construction models that can be used as guides. All of these instrument-building models require certain stages and processes. The Gregory Model is used to create visual, auditory, and kinaesthetic skill diagnostic test instruments, and it consists of six steps: (i) defining the test, (ii) determining the scale, (iii) constructing the item, (iv) testing the item, (v) reviewing the test, and (vi) publishing the test (Gregory, 2015). The researcher chose this model because item-building procedure is straightforward and appropriate for the setting of this study. Regardless of the model employed in instrument development, the instrument's validity and reliability method must be followed to verify that the items developed are appropriate for assessing and generalising the study's findings. Furthermore, the instrument designed should be transparent, intelligible, and have an accurate measuring scale (Salwana Mahamud and Norasmah Othman, 2020). The researcher constructed a study instrument in the form of pre- and post-test instruments, which will be completed by students from the treatment and control groups. This pre- and post-test instrument is appropriate for evaluating the efficacy, effectiveness, and novel techniques for gathering empirical data before and after the intervention is implemented (Cohen and Cornwell, 1989). Campbell and Stanley (1963) suggest that pre- and post-test instruments are critical for determining the interaction, cause, and effect of a variable that is being tested. According to Creswell (2018), the pre- and post-test results can be evaluated using a scale or exam marks. Even Afriani et al. (2020) employed pre- and post-test instruments to assess the reading levels of students in the treatment and control groups. As a result, using pre- and post-test sets to measure experimental investigations is ideal. This visual, auditory, and kinaesthetic skill diagnostic test instrument was developed using many theories as primary sources, including the VARK Model (1987), Laird's Sensory Theory (1985), and the VAK Model (1979). The integration of various models and theories is designed to test hypotheses and solve the research problems.

**2.3 Pre- and Post Test**

Experimental studies often employ pre- and post-tests as research instruments. This is because experimental investigations seek to understand the source and impact of a phenomenon, as well as to assess the efficacy of treatments,

new methodologies, or adopted programs (González-Alonso et al., 2020). Other researchers, particularly in the field of education, commonly employ pre- and post-test instruments to stimulate the development of innovative learning techniques. Among them are studies on the effectiveness of e-learning implementation (Siti Azura Abu Hassan et al., 2021), the effectiveness of multimedia methods (Che Suriani Kiflee et al., 2020), the effectiveness of role-playing activities (Kanniamah Arumainathan et al., 2023), and the effectiveness of Malay language learning videos (Zazila Shafinaz Zaharon Lai and Ahmad Nurzid Rosli, 2021). The pre- and post-tests in this study are referred to as visual, auditory, and kinaesthetic skill diagnostic tests. It assesses the students' visual, auditory, and kinaesthetic abilities before and after the intervention. In reality, this test set goes through a validity and reliability procedure to ensure that the instrument is correctly delivered and measures the students' competencies.

## 3. METHODOLOGY

This study employs a quantitative design with a pre- and post-test as data collection tools. The pre- and post-test instruments are known as visual, auditory, and kinaesthetic diagnostic tests for students. It has four sections: the preliminary and parts A, B, and C. The preliminary section of the instrument comprises questionnaire items pertaining to student demographics such as age, gender, position at school, and socioeconomic status. Part A discusses the students' kinaesthetic abilities test, which consists of 15 items. Part B consists of ten items for auditory abilities, while part C is a five-item test for kinaesthetic skills. Overall, it consists of 30 items, which all research participants will answer. This research included 38 fifth-year students chosen based on the class's original physical state. As this is quasi-experimental research, the participant population cannot be drawn at random and must consist of current students. The validity procedure includes five experts who were selected to assess the validity of the diagnostic tests. The CVR value was utilised by the researcher to assess the built instrument's content validity. The experts are provided with a three-measuring-scales instrument: extremely important, useful but less important, and unimportant. Lawshe (1975) believes that the content validity of the three-scale instrument should be evaluated using the CVR method to determine expert agreement. The corresponding values will vary from -1 to +1. $ne$ is the number of experts who select the item as significant, whereas N is the total number of experts. The formula for calculating the CVR value is given below:

$$\text{CVR Formula} = \frac{ne - (N/2)}{N/2}$$

Lawshe (1975) also indicates that if the number of experts who agree to accept an item exceeds half of the total number, the item can be accepted since the CVR will be positive, between 0 and 0.99. However, if the CVR value is negative, the item must be modified or removed from the list of items (Romero Jeldres et al., 2023). Table 1 displays the CVR values based on the number of experts involved with the instrument's content validation procedure.

Table 1: Minimum value of CVR based on the number of experts (Lawshe 1975)

| Number of Experts | Minimum Value |
|---|---|
| 5 | .99 |
| 6 | .99 |
| 7 | .99 |
| 8 | .75 |
| 9 | .78 |
| 10 | .62 |
| 11 | .59 |
| 12 | .56 |
| 14 | .54 |
| 14 | .51 |
| 15 | .49 |
| 20 | .42 |
| 25 | .37 |
| 30 | .33 |
| 35 | .31 |
| 40 | .29 |

Creswell (2018) asserts that reliability is critical to establishing the consistency of the developed items after they are answered by the research population. The measuring scale chosen to create the instrument will influence the type of testing utilised to get the reliability results. Therefore, the KR-20 value is employed in the reliability procedure of this research instrument, given that the scale used is dichotomous. The symbol k represents the number of items, p represents the number of correct responses provided by the participant, and q represents the number of incorrect answers. $\sigma^2$ indicates the variance of the entire population score. Furthermore, Kuder and Richardson (1937) indicated that the appropriate value of KR-20 is between 0.7 and 0.9, using the following formula:

$$\text{KR-20 Formula} = \frac{k}{k-1}\left(1-\frac{\Sigma pq}{\sigma^2}\right)$$

Ntumi et al. (2023) agreed that if the KR-20 reliability coefficient is 0.7 or above, the instrument may be considered acceptable. The KR-20 consistency value of 0.5 to 0.69 is acceptable for use as a research instrument; however, certain items must be modified and updated to achieve the acceptable values (Athiworakun et al., 2018). If the KR-20 value is less than 0.5, the instrument should be discarded because it contains elements that are unrelated to one another (Ekolu and Quainoo, 2019). Thus, the values acquired based on the calculating procedure are shown in Table 2 below:

Table 2: KR-20 scale values (source Kuder and Richardson, 1937)

| KR-20 Scale | Indicator |
|---|---|
| > 0.90 | Excellent |
| 0.80 – 0.89 | Good |
| 0.70 – 0.79 | Acceptable |
| 0.60 – 0.69 | Questionable |
| 0.50 – 0.59 | Weak |
| < 0.50 | Discarded |

## 4. FINDINGS

The visual, auditory, and kinaesthetic abilities diagnostic test instruments have been validated using CVR values. Five experts had assessed the diagnostic test's content validity by evaluating 30 items. Table 3 shows the values derived from the experts' agreement:

Table 3: Content validity value of the diagnostic test

| Item Number | Expert 1 | Expert 2 | Expert 3 | Expert 4 | Expert 5 | CVR Value |
|---|---|---|---|---|---|---|
| A1 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| A2 | | 1 | 1 | | 1 | 0.20 |
| A3 | | 1 | | 1 | 1 | 0.20 |
| A4 | | 1 | | 1 | | -0.20 |
| A5 | | 1 | | 1 | | -0.20 |
| A6 | | 1 | | | 1 | -0.20 |
| A7 | | | 1 | 1 | | -0.20 |
| A8 | | 1 | | 1 | 1 | 0.20 |
| A9 | 1 | 1 | | 1 | 1 | 0.60 |
| A10 | | 1 | | | | -0.60 |
| A11 | | 1 | 1 | 1 | | 0.20 |
| A12 | | 1 | | 1 | 1 | 0.20 |
| A13 | | 1 | | 1 | 1 | 0.20 |
| A14 | 1 | 1 | | 1 | 1 | 0.60 |
| A15 | | 1 | 1 | 1 | | 0.20 |
| B16 | | 1 | 1 | 1 | | 0.20 |
| B17 | | 1 | | | 1 | -0.20 |
| B18 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| B19 | | 1 | 1 | | 1 | 0.20 |
| B20 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| B21 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| B22 | 1 | 1 | | 1 | 1 | 0.60 |
| B23 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| B24 | | 1 | 1 | 1 | | 0.20 |
| B25 | | 1 | 1 | 1 | | 0.20 |
| C26 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| C27 | | 1 | 1 | 1 | | 0.20 |
| C28 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| C29 | 1 | 1 | 1 | 1 | 1 | 1.00 |
| C30 | 1 | 1 | 1 | 1 | 1 | 1.00 |

According to Table 3, 24 items have positive values, and at least three out of five experts agree to accept the developed items. The KR-20 value is used to calculate the reliability of the diagnostic test instrument, although six items

exhibited negative values that need to be investigated further to see if they should be adjusted or discarded. Table 4 shows the dependability values.

Table 4: Reliability values of the diagnostic tests

| Value | Total | KR-20 Value |
|---|---|---|
| K | 30 | |
| Σpq | 5.14 | 0.71 |
| $\sigma^2$ | 16.29 | |

The instrument's reliability for the developed diagnostic test obtained a KR-20 rating of 0.71. The researcher achieved the reliability value as the result of the pilot study which assessed the instrument before deploying it in a study with a real population. The test contains 30 items measuring visual, auditory, and kinaesthetic ability. The average number of correct responses (p) and wrong responses (q) in the research sample is 5.14, with a variance of 16.29.

## 5. DISCUSSION

The validity and reliability of the instrument ensure that the items and constructs developed match the study's objectives and effectively measure a phenomenon. It is related to the instrument's limitations during measurement, which may affect the data's accuracy. Table 3 shows that nine items received 1.00 and three items received 0.60 when experts reviewed them throughout the validity procedure. According to Lawshe (1975), a score between 0.60 and 1.00 indicates that an item is likely to be accepted. It demonstrates that all experts accepted the developed items. The 12 items had a positive score of 0.20, indicating that more than half of the experts agreed with the items. According to Jeldres et al. (2023) and Rodríguez et al. (2017), a positive CVR score of less than 0.60 is considered acceptable if the majority of experts agree. As a result, the researcher maintained the items on the diagnostic test instrument.

Six items (A4, A5, A6, A7, A10, and B17) had negative CVR values, which indicated expert evaluation and agreement. The poor score is due to the fact that the majority of experts disagree with the items developed in response to some of the identified challenges. Among them are the item's technical difficulties, which include spelling errors, incorrect language changes, and sentence structure. In addition, experts recommend that items be updated and summarised so that the study population may easily grasp them. In practice, the selected experts provided guidance on how to adapt the items to the focus of the study. According to Napitupulu (2020) and Nur Farhana binti Ramli et al. (2018), items having negative values should be discussed and modified to fit the study's purpose. As a result, the researcher and supervisor had extra discussions concerning the items with negative values. As a result, the items were allowed for inclusion in the diagnostic test instrument; however, the revised items were also reviewed by the experts. To make the items more relevant, modifications were made to the item structure as well as to the proper use of language and term accuracy.

Regarding reliability, the researcher is guided by the KR-20 value, obtained through analysis and calculation using a particular formula. The researcher used Microsoft Excel to analyse the pilot study data. Kuder and Richardson (1937) concluded that the optimal degree of reliability is 0.50 to 1. However, the lowest score implies poor reliability, whereas a high value indicates outstanding reliability. The diagnostic test instrument had a reliability grade of 0.71. It shows that the instrument is acceptable (Kuder Richardson, 1937). Acceptable reliability values for KR-20 are 0.60 (Rönning et al., 2013) and 0.80 to 0.90 (Mueller, 1974; Raju, 1982). For this reason, the researcher assigned a reliability score of 0.70 to these diagnostic test instruments.

Since the scale indication verifies that the instrument is acceptable and the reliability value exceeds 0.70, this diagnostic test may be used in research (Kuder Richardson, 1937). The validity and reliability procedure provides the best indicator of the instrument's accuracy and consistency prior to use, as well as its ability to measure accurately. Accurate measurement based on the construct being studied provides both relevant information and accurate data for generalisation purposes. As a result, each of these procedures needs a justification for how they are used in the study and why this measurement is used. Furthermore, several validity and reliability tests can be used; however, these tests are based on research and the use of appropriate scales.

## 6. CONCLUSION

This study examines the validity and reliability of a diagnostic test designed to assess visual, auditory, and kinaesthetic learning skills. The instrument was specifically developed for this purpose and evaluated using Content Validity Ratio (CVR) and Kuder–Richardson Formula 20 (KR-20) values. The item development process involved an extensive review of relevant literature to ensure the generalisability of the findings.

In addition, the results of this study are consistent with the findings of Lechien et al. (2024), who employed Cronbach's alpha to establish reliability and achieved a coefficient of 0.75, indicating acceptable internal consistency. Similarly, Ghafouri et al. (2023) reported comparable reliability outcomes, further supporting the reliability results of the present study.

In the current digital era, researchers and educators—particularly within educational institutions—are committed to equipping students with a diverse range of skills that enable them to adapt to and thrive in modern learning environments. This diagnostic instrument provides a valuable foundation for future research on student competency assessment, teaching strategies, and learning challenges. Moreover, analysing students' skill profiles can help address existing gaps in prior studies. As education continues to evolve within the framework of Industry Revolution (IR) 5.0, strengthening research on student learning skills remains essential for achieving educational excellence.

**REFERENCE**

Afriani, Z. L., Anggraini, M., & Riswanto, R. (2020). The effect of Question Answer Relationship (QAR) Strategy in enhancing students' reading comprehension. *Journal of English Education and Teaching*, *4*(4), 548-558.

Ahmed, I., & Ishtiaq, S. (2021). Reliability and Validity: Importance in medical research. *Methods*, *12*(1), 2401-2406.

Aini, H., Wahyuningsih, R., & Bahari, J. I. (2023). Implementasi Pembelajaran Seni Tari Tradisional Gandrung Dalam Upaya Meningkatkan Kecerdasan Kinestetik Pada Anak Usia 5-6 Tahun Di Ra Perwanida Ii Purwoasri Tegaldlimo Banyuwangi. *JIEEC (Journal of Islamic Education for Early Childhood)*, *3*(2), 11-22.

Athiworakun, C., Vathanalaoha, K., Thongprayoon, T., Rajprasit, K., & Yaemtui, W. (2018). SWU-SET as a CEFR standardized English test. *Journal of Language Teaching and Research*, *9*(2), 261-267.

Barbe, W.B. & Milone, Jr., M.N. (1981). What We Know About Modality Strengths. *Educational Leadership* (Association for Supervision and Curriculum Development): 378–380.

Bergsland, K., & Vogt, H. (1962). On the validity of glottochronology. *Current anthropology*, *3*(2), 115-153.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured?. *Physiotherapy*, *86*(2), 94-99.

Campbell. D. T.& Stanley. J. C. (1963). Experimental And Quasi-Experiment Designs For Research. London: Handbook Research on Teaching.

Che Suriani Kiflee, Siti Aishah Hassan, Fariza Khalid, Noorlin Maaulot & Rorlinda Yusof. (2020). Analisis Keberkesanan Kaedah Multimedia dalam Pengajaran dan pembelajaran terhadap pelajar pintar dan berbakat. *Jurnal Personalia Pelajar*, *23*(2).

Chua, W. L., Smith, D., Wee, L. P. C., Ting, K. C., Yeo, M. L. K., Mordiffi, S. Z., & Liaw, S. Y. (2023). Development and psychometric evaluation of the Attitudes Towards Recognising Early and Noticeable Deterioration (ATREND) scale. *Journal of Clinical Nursing*, *32*(11-12), 2684-2699.

Cohen, E., & Cornwell, L. (1989). A question of ethics: Developing information system ethics. *Journal of Business Ethics*, *8*, 431-437.

Cohen, R. L. (1980, May). Auditory skills and the communicative process. In *Seminars in Hearing* (Vol. 1, No. 02, pp. 107-114). Copyright© 1980 by Thieme Medical Publishers, Inc.

Creswell J. W. & Creswell J. D. (2018). Research Design Qualitative, Quantitative, and Mixed Methods Approaches Fifth Edition. California: SAGE Publications.

Ekolu, S. O., & Quainoo, H. (2019). Reliability of assessments in engineering education using Cronbach's alpha, KR and split-half methods. *Global journal of engineering education*, *21*(1), 24-29.

Elangovan, N., & Sundaravel, E. (2021). Method of preparing a document for survey instrument validation by experts. *MethodsX*, *8*, 101326.

Fenny Farhana binti Mohd. Sajari**,** Hafizhah binti Zulkifli & Shahlan bin Surat . (2023). Kesahan dan Kebolehpercayaan Instruman Kesediaan GPI Melaksanakan Model Tasmik j-QAF Secara Dalam Talian. *Journal of Quran Sunnah Education & Special Needs*, *7*(1), 16-29.

Fitri, D., & Gani, E. (2020). Differences in Reading Comprehension Text Reading Skills by Using Auditory, Intellectually, Repetition (AIR) and Brainstorming Learning Models in Terms of Students Reading Interest. In *1st Progress in Social Science, Humanities and Education Research Symposium (PSSHERS 2019)* (pp. 43-47). Atlantis Press.

Fleming, N., & Baume, D. (2006). Learning Styles Again: VARKing up the right tree!. *Educational developments*, *7*(4), 4.

Gardner, H., & Moran, S. (2006). The science of multiple intelligences theory: A response to Lynn Waterhouse. *Educational psychologist*, *41*(4), 227-232.

González-Alonso, F., Guillén-Gámez, F. D., & de Castro-Hernández, R. M. (2020). Methodological analysis of the effect of an anti-bullying programme in secondary education through communicative competence: A pre-test–post-test study with a control-experimental group. *International journal of environmental research and public health*, *17*(9), 3047.

Ghafouri, S.,Jahani, S., Maraghi, E., Sayadi, N. 2023. Psychometrics and Validation of the Intensive Care Unit Memory Assessment Tool in the Iranian Population. *Caspian Journal of Neurological Sciences*. DOI: 10.32598/CJNS.9.32.391.1

Gregory, R. J. 2015. Psychological Testing History, Principles, and Applications seventh edition. England: Pearson Education, Inc

Hajar, W. S., Fajrin, L. P., Setiawan, E., & Setiyatna, H. (2020). Upaya Meningkatkan Kecerdasan Kinestetik Anak Melalui Outbound. *ABNA: Journal of Islamic Early Childhood Education*, *1*(2).

Hastuti, P. E. (2020). A Comparative Study On Preferred Learning Style Between Auditory And Kinesthetic Learners On Reading Achievement. *LangEdu Journal.* Pg 1-13

Lechien, J.R., Maniaci, Vaira, L.A. 2024. Validity and reliability of an instrument evaluating the performance of intelligent chatbot: the Artificial Intelligence Performance Instrument (AIPI). European Archives of Oto-Rhino-Laryngology. DOI: 10.1007/s00405-023-08219-y

Jamal Nordin Yunus (2022). Opportunities and Challenges in Tertiary Education in the Era of Society 5.0 in the Malaysian Higher Education Context: Peluang Dan Cabaran Pendidikan Tinggi Di Era Society 5.0 Dalam Konteks Pendidikan Tinggi Di Malaysia. *Management Research Journal*, *11*(2), 74-79.

Jeldres, M. R., Costa, E. D., & Faouzi, T. (2023). A review of Lawshe's method for calculating content validity in the social sciences. *Front. Educ*. Vol. 8. https://doi.org/10.3389/feduc.2023.1271335

Kadam, P. R., Gaikwad, U. K., & Bhamre, P. D. (2021). Student centred learning using student's learning style. *Journal of Engineering Education Transformations*, *34*, 391-395.

Kanniamah Arumainathan, Kway Eng Hock &Kama Shaffeei. (2023). Kajian keberkesanan aktiviti main peranan “I Can Communicate Too” terhadap murid masalah pembelajaran berkefungsian rendah. *Jurnal Pendidikan Bitara UPSI*, *16*, 55-76.

Kayalvili Munusamy & Khairul Azhar Jamaludin (2022). Cabaran Guru Untuk Mengintegrasikan Teknologi Maklumat Dan Komunikasi (TMK) Bagi Meningkatkan Kemahiran Membaca Dalam Kalangan Murid. *Jurnal Dunia Pendidikan*, *4*(3), 265-273.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Miller B., Camarda A., Mercier. M, Burkhardt J. M, Morisseau T,, Bougrine S. B., Vinchon F., Hayek S, Landais M. A., Mourey F., 2023. Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. Journal of Intelligence 11: 54. https://doi.org/10.3390/ jintelligence11030054

Mohammad Saffuan Mat Saad. (2020). Revolusi Industri dan Masyarakat 5.0. *E-Jurnal Penyelidikan Dan Inovasi*, *7*, 29–44. Retrieved from https://ejpi.kuis.edu.my/index.php/ejpi/article/view/41

Mohd Hussin Musa ,Christina Andin & Syahrul Nizam Salam. (2020). Kesahan Dan Kebolehpercayaan Instrumen Pentaksiran Kemahiran Insaniah Murid (IPKIM) Mata Pelajaran Rekacipta. *International Journal of Education, Psychology and Counseling*, *5*(37), 96-107.

Mohd Taufek Harun & Mohd Ikhwan Hadi Yaacob. (2021). Penilaian Kualiti Modul FZ-STEM: Aspek Kesahan dan Kebolehpercayaan: Assessing the Quality of FZ-STEM Module: Validity and Reliability Aspects. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, *11*, 107-117.

Mueller, D. J. (1974). A test of the validity of two scales on Rokeach's Value Survey. *The Journal of Social Psychology*, *94*(2), 289-290.

Muhamad Firdaus Ramli & Rofidah Musa (2020). Ekplorasi seni visual melalui aktiviti lakaran asas terhadap kanak-kanak prasekolah: An exploration of visual arts through fundamental sketch activities to the preschool children. *Jurnal Pendidikan Awal Kanak-Kanak Kebangsaan*, *9*(1), 35-47.

Muhamad Saiful Bahri Yusoff. (2019). ABC of content validation and content validity index calculation. Education in Medicine Journal. 2019;11(2):49–54. https://doi.org/10.21315/eimj2019.11.2.6

Napitupulu, D. (2020). User-Acceptance instrument development: a content validity study in the e-participation context. *Journal of applied research and technology*, *18*(1), 34-43.

Norfishah Mat Rabi. (2021). *Rawatan alternatif untuk kanak-kanak autistik*. Penerbit USM.

Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the Psychometric Properties (" Item Difficulty, Discrimination and Reliability Indices") of Test Items Using Kuder-Richardson Approach (KR-20). *Shanlax International Journal of Education*, *11*(3), 18-28.

Nur Farhana binti Ramli, Othman bin Talib, Umi Kalthom binti Abdul Manaf, Siti Aishah binti Hassan. (2018). Content Validity of STEMTIP Using CVR Method. International Journal of Academic Research in Business and Social Sciences, 8(7), 1118–1125.

O'Brien, G., & Yeatman, J. D. (2021). Bridging sensory and language theories of dyslexia: Toward a multifactorial model. *Developmental Science*, *24*(3), e13039.

Pienaar, L., Prince, R., & Abrahams, A. (2021). Development of a baseline assessment tool to establish students' foundational knowledge of life sciences at entry to university. *African Journal of Health Professions Education*, *13*(1), 77-82.

Raju, N. S. (1982). On test homogeneity and maximum KR-20. *Educational and Psychological Measurement*, *42*(1), 145-152.

Rodrigues, I. B., Adachi, J. D., Beattie, K. A., & MacDermid, J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal disorders*, *18*, 1-9.. https://doi.org/10.1186/s12891-017-1914-5

Romero Jeldres, M., Díaz Costa, E., & Faouzi Nadim, T. (2023). A Review of Lawshe's Method for Calculating Content Validity in the Social Sciences. In *Frontiers in Education* (Vol. 8, p. 1271335). Frontiers. https://doi.org/10.3389/feduc.2023.1271335

Rönning, H., Årestedt, K., Nielsen, N. E., Swahn, E., & Strömberg, A. (2013). Development and psychometric evaluation of the knowledge scale for adults with congenitally malformed hearts. *Journal of Cardiovascular Nursing*, *28*(3), 228-237.

Rosmayanti, E., & Santana, F. D. T. (2021). Pembelajaran Senam Fantasi Untuk Meningkatkan Kecerdasan Kinestetik Anak Usia 5-6 Tahun. *Ceria (Cerdas Energik Responsif Inovatif Adaptif)*, *4*(3), 242-250.

Saeed, M. O., Ondracek, J., Bertsch, A., Shefstad, B., Procyshen, M., & Hernandez, M. (2015). An investigation of learning style theory and GPA at a US University. *Inspira-Journal of Modern Management & Entrepreneurship*, *5*(03), 1-14.

Salwana Mahamud & Norasmah Bt Othman. (2020). Kesahan dan Kebolehpercayaan Instrumen Kecerdasan Menghadapi Cabaran Teknologi Menggunakan Model Rasch (Validity and Reliability the Instruments Adversity Quotient for Technological Challenges Using the Rasch Model). *Akademika*, *90*(S3).

Setyosari, P., Kuswandi, D., & Sartika, D. (2020). Effect of the Community of Inquiry (CoI) Learning Model and Learning Style towards Social Skills. *European Journal of Educational Research*, *9*(2), 56-578.

Siti Azura Abu Hassan, Suzana Binti Zainol Abidin & Zulkurnain Bin Hassan. (2021). Keberkesanan Pembelajaran Dan Pengajaran Dalam Talian (Epembelajaran) Terhadap Pembelajaran Pelajar Di Kolej Komuniti Hulu Langat. *International Journal of Humanities Technology and Civilization*, 1-8.

Siti Hayati binti Mohd Yusoff & Halim bin Husain. (2021). Penggunaan perisian aplikasi teknologi maklumat dan komunikasi meningkatkan integrasi domain pembelajaran dalam karya murid belajar gaya visual. *Jurnal Penyelidikan Dedikasi*, *18*(1), 140-160.

Soon Ping Sia & Kamariah Abu Bakar. (2022). Meningkatkan Kemahiran Operasi Tambah Kanak-kanak Prasekolah melalui Penggunaan Bahan Konkrit dan Visual. *Jurnal Pendidikan Awal Kanak-kanak Kebangsaan*, *11*, 103-113.

Subramaniam a/l Raman, Muhammad Zuhair bin Zainal & Fadzilah binti Amzah. (2023). Pembinaan, Kesahan dan Kebolehpercayaan Modul Penulisan Karangan Bahasa Melayu Tingkatan Dua. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *8*(1), e002031. https://doi.org/10.47405/mjssh.v8i1.2031

Sürücü, L., & Maslakçi, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, *8*(3), 2694-2726.

Tobón, S., & Luna-Nemecio, J. (2021). Complex thinking and sustainable social development: Validity and reliability of the complex-21 scale. *Sustainability*, *13*(12), 6591.

Verawati, N. N. S. P., Prayogi, S., Yusup, M. Y., & Taha, H. (2020). Development of the Test Instrument for Measuring Students' Critical Thinking Abilities on Fluid Material. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *8*(1), 46-56.

Vinoothini Mannogaran & Nor Azwahanum Nor Shaid. (2023). Kesahan dan Kebolehpercayaan Instrumen Mengenai Pengetahuan, Sikap dan Cabaran Guru dalam Pelaksanaan Pentaksiran Bilik Darjah. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *8*(1), e001958. https://doi.org/10.47405/mjssh.v8i1.1958

Zazila Shafinaz Zaharon Lai & Ahmad Nurzid Rosli. (2021). Keberkesanan penggunaan video pembelajaran Bahasa Melayu dalam kalangan murid tahun 6 semasa pengajaran dan pembelajaran di rumah (PdPR). Journal of ICT in Education, 8(3), 76-99. https://doi.org/10.37134/jictie.vol8.sp.1.7.2021

Ghafouri, S.,Jahani, S., Maraghi, E., Sayadi, N. 2023. Psychometrics and Validation of the Intensive Care Unit Memory Assessment Tool in the Iranian Population. ***Caspian Journal of Neurological Sciences***. DOI: 10.32598/CJNS.9.32.391.1