Javanese Culture-Based Numeracy Test Using the Rasch Model

Ninik Zuroidah[1*], Adisa Eka Damayanti[1], Jerhi Wahyu Fernanda[1] & Burhanudin Syaifulloh[2]

[1]Department of Mathematics Education, Institut Agama Islam Negeri Kediri, Indonesia
[2] Faculty of Education, University of Canberra, Australia

*Corresponding email: ninikzuroidah@iainkediri.ac.id

Abstract: Javanese Culture-Based Numeracy Test Using the Rasch Model. Numeracy is one of the six basic literacies students need to improve in the 21st century. **Objectives:** This research aimed to develop a numeracy test based on Javanese culture using the Rasch model. **Methods:** The ADDIE model is used in the numeracy development model. The question grid is based on modified curriculum-13 and the Minimum Competency Assessment (MCA) numeracy framework. Data was obtained from tests aimed at 300 SMP Negeri 5 Kediri City students, which were determined using simple probabilistic random sampling techniques. Quantitative descriptive analysis assisted by SPSS and Jamovi programs was used to analyze reliability, item suitability tests, and difficulty levels. **Findings:** The research results showed that 50 numeracy questions with Javanese cultural content had been developed, with 17 questions of medium validity and 33 questions of high validity. The reliability estimate is 0.833, which is very high. In the test of item suitability and item difficulty level, it was found that four items, namely items 1, 2, 6, and 9, were not suitable for use in numeracy test, while 46 other items matched the Rasch model and were suitable for use. **Conclusion:** This research provides important information for further development and analysis studies related to Javanese culture-based numeracy tests.

Keywords: test instrument, numeracy, Javanese culture, ADDIE, Rasch model.

To cite this article:

■ INTRODUCTION

Since 2006, UNESCO has emphasized that numeracy is one of the abilities that can determine the progress of a nation (Hong, Thakuriah, Mason, & Lido, 2020). Apart from that, numeracy is also one of the six basic literacies needed to face the challenges of the 21st century because, without realizing it, numeracy is embedded in daily activities, especially in the world of education. The importance of numeracy in the world of education in Indonesia can be seen in the government's policy of abolishing the National Assessment to become a Computer-Based National Assessment, where one of the test topics is numeracy, which has been implemented since 2021 (Fadhilah, Setiawan, Indriani, & Yulianti, 2022).

The numerical results from the implementation of the Computer-Based National Assessment show that student numeracy in Indonesia is still below the minimum competency level or that most students have not reached the minimum competency limit (Rohendi, 2022; Yang & Sianturi, 2021). The low numeracy of students in Indonesia has received more attention from several researchers for analysis. The results of

the analysis show that there is a lack of motivation to learn numeracy, the assumption that numeracy is a subject that is difficult to understand, a lack of reference materials, and the Computer-Based National Assessment is not used as a benchmark for graduation (Anam, Suteja, Septianto, Purnomo, & Utami, 2020; Budiman, Hufad, & Purnomo, 2022; Kustantina, Nuryadi, & Marhaeni, 2022). This problem cannot be ignored, considering that numeracy is also a form of assessment on an international scale, such as TIMSS, PISA, Lantinte, and Naplan in Australia (Hall & Zmood, 2019), etc., two of which Indonesia is a participant in implementing the assessment.

Globally, the importance of numeracy has been widely recognized, leading to various educational reforms aimed at improving students' numerical literacy (OECD, 2018). Research in countries like Norway, Bolivia, and China has demonstrated that effective numeracy education requires not only a solid curriculum but also culturally responsive teaching methods that connect learning to students' everyday experiences (Greer & Mukhopadhyay, 2015; Nortvedt & Wiese, 2020; Yu et al., 2021). For example, studies in the Bolivia have shown that incorporating local contexts and cultural elements into numeracy education can significantly enhance students' engagement and understanding (O'Shaughnessy et al., 2023). This is particularly relevant in multicultural societies, where diverse cultural backgrounds can be leveraged to make numeracy more relatable and meaningful to students (Nortvedt & Wiese, 2020; Pinxten, 2016). However, such approaches are less common in Indonesia, where numeracy education has traditionally been more standardized and less adaptive to regional cultural variations.

Various efforts have also been made by several researchers to increase students' interest in numeracy in Indonesia, one of which is by developing numeracy tests with a variety of content and contexts, as has been done (Purnomo, Sa'dijah, Hidayanto, Permadi, & Anwar, 2022), but the results of their research reveal that the numeracy that has been developed not by the numeracy framework in ANBK. Based on the results of research and analysis of low numeracy in Indonesia, it is necessary to develop numeracy based on ANBK standardized tests to become a supporting material for increasing student numeracy in Indonesia. This effort is supported by research (Widarti, Rokhim, Septiani, & Dzikrulloh, 2022), which reports that the availability of numeracy references with context variations is less than 50%, especially in the Indonesian cultural context.

The cultural context includes the socio-cultural context, one of the contexts of numeracy (Widarti et al., 2022). Culture is the overall form of culture, special characteristics, or patterns for a certain group of individuals (Issatriadi, Rokhmad, & Timoer, 1977). As a country with thousands of islands, Indonesia also has an abundance of culture, especially the culture on the island of Java. Existing culture must be protected, developed, and preserved. One effort that can be made to preserve Javanese culture is by linking it to the educational sector, in this case, by developing numeracy tests that use the Javanese cultural context. Javanese culture, with its rich history and diverse traditions, offers a unique context for developing numeracy assessments. Incorporating cultural elements such as traditional Javanese games, architecture, and art forms into numeracy questions can make learning more relevant and engaging for students in Java (Novikasari, Muttaqin, & Elebiary, 2024; Stafford, 2008; Suryonegoro, Wuryastuti, & Dewi, 2024). Previous research has explored the integration of local culture into educational content, but there is a lack of studies focusing specifically on numeracy within the Javanese context (Kurniawan, Budiarto, & Ekawati, 2022; Primaniarta & de Mattos, 2022). This presents

an opportunity to create innovative educational tools that both support numeracy skills and promote cultural heritage.

Several studies have developed Javanese-based numeracy but still focus on one region and one topic of discussion (Kurniawan et al., 2022; Sumarni et al., 2023). However, these studies have not comprehensively explored the integration of Javanese culture into numeracy tests across a broader range of topics and regions. Additionally, there is a gap in the literature regarding the application of advanced psychometric models like the Rasch model in this context. By addressing this gap, the current study aims to contribute to both the fields of educational assessment and cultural studies by providing a model for developing culturally relevant numeracy tests that are psychometrically robust.

In test development, the question instruments developed must have good question quality in terms of validity, reliability, and question difficulty. To obtain good quality, a question instrument can be analyzed using the Rasch model as one of the Item Response Theory (TRB) models (Elvira, Retnawati, Rohaeti, & Sainuddin, 2023). Compared with other TRB models, the Rasch model has more accurate statistical analysis results and accurate calculations of measurement errors (Ardiyanti, 2016). Apart from that, the grain parameters do not depend on the sample so they have high productivity and stability in the grain parameters (Isnani, Utami, Susongko, & Lestiani, 2019). The Rasch model, developed by Georg Rasch (1960), is widely used in educational research for its ability to provide detailed item-level analysis and to ensure that test items are aligned with the underlying constructs they are intended to measure (Bond & Fox, 2015). In the context of numeracy assessment, the Rasch model offers several advantages, including the ability to assess the difficulty of items and the ability to compare students' abilities on a common scale, regardless of which items they

answered (Linacre, 2024). Previous studies have successfully applied the Rasch model in various educational settings, including in culturally diverse contexts, to ensure that assessments are fair and unbiased (Boone, Staver, & Yale, 2013). However, its application in developing numeracy assessments that are explicitly based on Javanese culture is still minimal. This research seeks to bridge this gap by applying the Rasch model to develop a numeracy test that is both culturally relevant and psychometrically sound.

By doing so, this study not only addresses the pressing issue of low numeracy skills in Indonesia but also contributes to the preservation and promotion of Javanese cultural heritage through education. The integration of cultural context into educational assessments aligns with broader educational goals of fostering inclusivity and relevance in learning (Gay, 2018). Moreover, by utilizing the Rasch model, this research ensures that the numeracy test developed is psychometrically sound, providing reliable and valid measurements of students' abilities. This approach supports the creation of culturally responsive assessment practices that are essential for meeting the diverse needs of students in a multicultural nation like Indonesia. Ultimately, this study aims to provide a model for integrating cultural contexts into educational assessments, which can be adapted and applied in other regions with rich cultural diversity. The outcomes of this research are expected to enhance students' engagement and achievement in numeracy, while also preserving cultural heritage, thus offering a significant contribution to the field of educational assessment and cultural studies.

■ METHOD
Participants

The participants in this research were 300 students from SMP Negeri 5 Kediri City, selected using a simple probabilistic random sampling technique. This method ensures that every student

in the population has an equal chance of being chosen, thus providing a representative sample for the study (Sugiyono, 2018). The participants were chosen based on the needs of the research, specifically for the development and validation of the Javanese culture-based numeracy test. In addition to the students, four experts served as evaluators during the content validation stage, consisting of two mathematics education experts and two field practitioners.

Research Design and Procedures

This study employs a development research design, following the ADDIE model, which includes five phases: Analysis, Design, Development, Implementation, and Evaluation.

During the Analysis phase, the researchers identified the educational needs and cultural contexts to be included in the numeracy test. The Design phase involved structuring the test items and aligning them with the 2013 teaching curriculum and the Ministry of Education and Culture's AKM framework. In the Development phase, the numeracy test was created, incorporating Javanese cultural elements such as weton, tumpengan, pitonan, wajik, jadah, traditional dance, and cow race. The Implementation phase involved administering the test to the sample population, and finally, the Evaluation phase focused on analyzing the test results using statistical methods to assess validity and reliability. The research flow is illustrated in Figure 1 below:
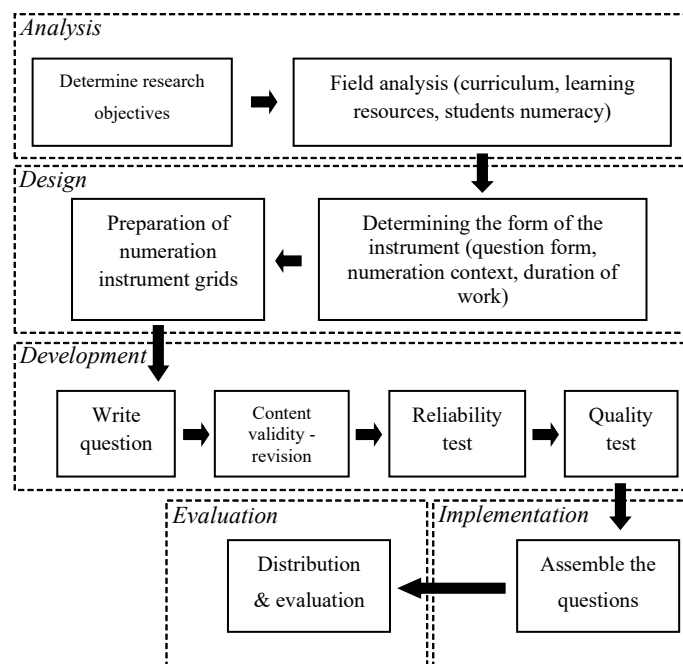

Figure 1. Research flow

The research was conducted over a period of six months, from January to June 2024, with each phase carefully planned to ensure a systematic approach to test development.

Instruments

The primary instrument developed for this research was a numeracy test consisting of 50 multiple-choice questions. The test items were

designed to assess four numeracy content areas: (1) numbers, (2) algebra, (3) geometry and measurement, and (4) data and uncertainty. These items were contextualized within Javanese cultural settings to enhance relevance and engagement for the students. The test was validated through expert judgment, involving two experts in mathematics education and two field practitioners. The validation focused on content validity, construct validity, and language clarity, assessed using a structured validation sheet with six content sub-aspects, five construct sub-aspects, and three language sub-aspects.

The content validity of the test was quantified using Aiken's V, where a V index d" 0.4 indicates low validity, 0.4 – 0.8 indicates moderate validity, and a V index e" 0.8 indicates high validity (Retnawati, 2016). Reliability was measured using Cronbach's Alpha, with categories ranging from very low to very high reliability, as shown in Table 1 below:

Table 1. Reliability coefficient category

Total score	Category
$0.80 < r_{xx'} \leq 1.00$	Very high
$0.60 < r_{xx'} \leq 0.80$	High
$0.40 < r_{xx'} \leq 0.60$	Moderate
$0.20 < r_{xx'} \leq 0.40$	Low
$-1.00 \leq r_{xx'} \leq 0.20$	Very low

Data Analysis

The data analysis in this research involved several statistical techniques. First, content validity was assessed using Aiken's V, providing a quantitative measure of the relevance and clarity of the test items based on expert judgment. Reliability was evaluated using Cronbach's Alpha, where the reliability coefficient categories determined the consistency of the test scores (Guilford, 1950). Additionally, the Rasch model was used to assess the fit and difficulty level of the test items. Items were considered to fit the

model if their infit and outfit values fell within the range of 0.50 – 1.50, ensuring that the items were both challenging and appropriate for the students' abilities (Sumintono & Widhiarso, 2014). The difficulty level of each item was categorized using established thresholds, as outlined in Table 2 below:

Table 2. Difficulty level category

Difficulty index	Category
<-2.00	Easy
-2.00 s.d 2.00	Moderate
>2.00	Difficult

■ RESULT AND DISCUSSION

This research, which aims to develop a numeracy test based on Javanese culture, was developed based on the ADDIE model with the following description of the results (Meihami & Malmir, 2024).

Analyze

At this stage, the researcher did several things which became the initial foundation for this research (Mai, Sun, Tan, & Zhao, 2024), namely (1) determining the research objectives, namely producing a test instrument that can measure and explore numeracy abilities, (2) curriculum analysis where developing this numeracy test used the 2013 curriculum which synchronized with the AKM framework for numeracy at junior high school/equivalent level, (3) analysis of learning resources was extracted from the location of the research respondent's school, namely SMP Negeri 5 Kediri City, where numeracy learning resources were still limited. To train students' numeracy, the teacher only facilitates five questions per meeting, which are shared via the WhatsApp application, and (4) an analysis of students' numeracy abilities, which is the foundation or main background of this research, where students' numeracy abilities in Indonesia are still low when

seen from the AKM numeracy results since 2021, as well as students at SMP Negeri 5 Kediri City. One of the causes of students' low numeracy skills is a lack of interest in reading, so when faced with numeracy questions standardized by the AKM framework, students are unable to solve them or, in other words, are not yet familiar with numeracy questions.

Design

At this stage, the researcher (1) selected the test instrument in the form of 50 numeracy multiple-choice questions. The multiple choice question format is the choice because the scoring can be done easily, quickly, and objectively (Romm, Kika, & Schoer, 2019; Tangianu et al., 2018) often used in measuring national to international standards, the most effective form of assessment, can last a long time and is economical (Gierl, Bulut, Guo, & Zhang, 2017), and can test measurements quickly in large groups. (2) determine the numeracy context in the form of Javanese culture because it is based on research by Mawaddah et la. (2022), where questions related to culturally based mathematical skills have good potential effects. Several researchers have also developed numeracy test instruments with their respective characteristic contexts, both in the form of instruments and learning modules (Mawaddah et al., 2022). (3) determine the test time of 150 minutes to complete 50 numeracy questions. (4) compiling a grid is a crucial stage in constructing the numeracy questions being developed. The grid compiled in this research uses two references, namely indicators of numeracy ability and KD by Ministry of Education and Culture regulations and the implications of the 2013 Curriculum. The instrument grid contains indicators of numeracy ability, KD, class, semester, material, question form, question indicators, and question number.

Development

At this stage, the researcher carried out (1) writing questions in a question card format containing the question number, question form, class, level, KD, material, numeracy ability indicators, question indicators, cognitive processes, cognitive level, and answers (Meihami & Malmir, 2024). The question cards were constructed to make it easier for researchers to validate the content. Apart from that, the thing that is taken into account in compiling these numeracy questions is the Javanese cultural context, which is the identity of the numeracy test that has been developed. The Javanese cultural context used is related to *weton*, *tumpengan*, *pitonan*, *wajik, jadah*, traditional dance, cow race, etc., which are adapted to class VII middle school material. Examples of questions that have been developed can be seen in Figure 2.

(2) Content validation was carried out on 2 experts in the field of mathematics and 2 field practitioners. The aspects assessed at this validation stage consist of six sub-aspects, constructs with five sub-aspects, and language with three sub-aspects. The assessment aspects in question are multiple choice questions, Javanese cultural context, and numeracy. The validation results were then analyzed using Aiken's V formula, which produced calculations, as seen in Table 3.

Table 3 shows that overall, the questions that have been developed have an Aiken's V index above 0.711. A total of 17 questions have moderate validity because the resulting index is between 0.400 – 0.800. Meanwhile, the other 33 questions have an index greater than 0.800, which means high validity. Valid questions show their suitability to the indicators, content, and context to be measured. This was then reinforced by research (Degroote, DeSmet, De Bourdeaudhuij, Van Dyck, & Crombez, 2020),

Figure 2. Numeracy questions based on javanese culture

which stated that the question instrument is content valid if the questions reflect what is to be measured, with a clear construct. (3) a reliability test, estimated after the instrument has been tested for content validity, is useful for knowing the instrument's stability, in this case, the Javanese culture-based numeracy test. The results of reliability estimation assisted by the SPSS program can be seen in Table 4.

Table 3. Aiken's v results

Item	V	Item	V	Item	V	Item	V	Item	V
1	0.785	11	0.825	21	0.713	31	0.832	41	0.817
2	0.794	12	0.821	22	0.764	32	0.798	42	0.784
3	0.799	13	0.820	23	0.816	33	0.795	43	0.839
4	0.820	14	0.799	24	0.789	34	0.832	44	0.813
5	0.823	15	0.801	25	0.810	35	0.831	45	0.835
6	0.763	16	0.781	26	0.711	36	0.835	46	0.839
7	0.821	17	0.772	27	0.752	37	0.839	47	0.831
8	0.820	18	0.816	28	0.820	38	0.824	48	0.828
9	0.824	19	0.819	29	0.806	39	0.807	49	0.828
10	0.817	20	0.810	30	0.749	40	0.785	50	0.817

Table 4. Reliability estimation

Estimate	McDonald's ω	Cronbach's α
Point estimate	0.817	0.833
95% CI lower bound	0.787	0.804
95% CI upper bound	0.846	0.858

Table 4 above shows that the reliability estimate using Cronbach Alpha produces a value of 0.833, which means the numeracy instrument questions are reliable/reliable in the very high category. So, it can be concluded that the numeracy test that has been developed is reliable

(4) the quality test is carried out by testing the difficulty level with the Rasch model and testing item suitability. The Jamovi program estimated item difficulty using the Rasch model for dichotomous data. However, what is done first is to estimate the suitability of the items to the Rasch model using infit and outfit values. Results will be obtained from this item suitability test by estimating student response patterns to instrument items, such as the numeracy test (Bond, Yan, & Heene, 2020).
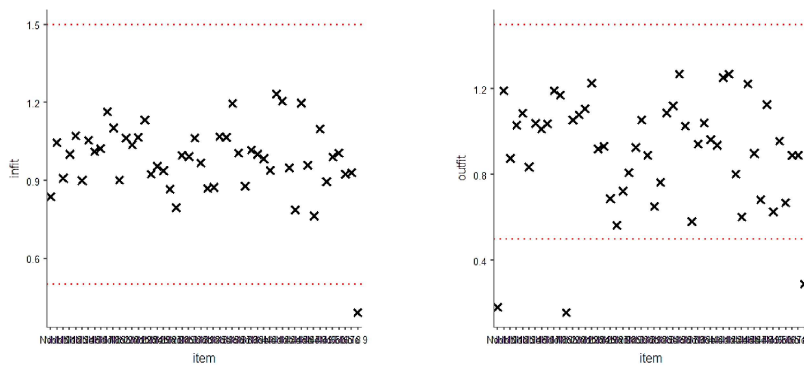

Figure 3. Item suitability test

Figure 3 shows that three questions are misfits because they do not meet the outfit (item 1, item 2, item 9), and one question item is a misfit because it does not meet the outfit (item 9). In other words, the three items are outside the red line, which indicates that the items do not match the Rasch model because the estimated value is below 0.50. So, 47 questions match the items and are acceptable. Furthermore, the estimated level of difficulty can be seen in Figure 4.

Figure 4 shows that overall, students' abilities are in the interval -4.0 to 5.0. For item 9, it is indicated by the presence of the rightmost
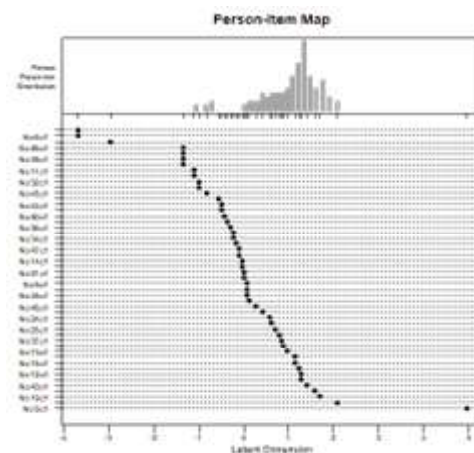

Figure 4. Person item map rasch model

point, which indicates that the item has a difficult category (ability of 5.0). Meanwhile, there are three points on the left, which represent item 1, item 2, and item 6 in the easy category (ability interval -3.0 to -4.0). Apart from that, the remaining questions are in the ability interval -1.50 – 2.0. (Fernanda & Hidayah, 2020; Sainuddin, 2018) states that good test items are in the ability interval -2.0 – 2.0. Based on the analysis above, the four question items (items 1, 2, 6, and 9) must be revised if they are still used as a question bank.

The topic of Javanese culture, which is raised in points 1 and 2, is the addition of weton. This topic in culture is used in traditional processions leading to marriage. The *weton* of men and women who are about to get married are added up according to the *neptu* obtained. The results of the Rasch model analysis show that item 1 has an outfit value of 0.182 and item 2 is 0.156 (less than 0.50), where Sumintono & Widhiarso (2014) explain that this estimate is less productive in measurement but does not reduce quality. This low value causes errors with high reliability. So this question item is not good (recommended to be rejected). Besides that, item 1 and item 2 are classified as easy questions because they have a logit value of -3.7004.

The Javanese cultural topic raised in point 6 is the cost of making tumpeng, illustrated with pictures and a price list. The results of the Rasch model analysis show that item 6 is classified as an easy question because it has a logit value of -3.7004. The topic of Javanese culture in point 9 is the cutting of *jadah*. The results of the Rasch model analysis show that item 9 has an infit value of 0.393 and an outfit of 0.290 (less than 0.5), where Sumintono & Widhiarso (2014) explain that this estimate is less productive in measurement but does not reduce quality. This low value causes errors with high reliability. So this question item is not good (recommended to be rejected). Besides that, item 9 is classified as

an easy question because it has a logit value of -2.9868.

Implementation and Evaluation

At the Implementation stage, the questions tested in the previous stage were reassembled and then distributed in research schools for mass use. At the Evaluation stage, researchers carry out a comprehensive evaluation of all the stages that have been carried out. There is a need for further evaluation regarding the test instruments developed after the implementation stage.

■ **CONCLUSION**

This research uses the Rasch model to develop a numeracy test based on Javanese culture. The development results show that a numeracy question instrument has been obtained that is valid, reliable and can be implemented in numeracy tests. The research results also confirm that Javanese culture is capable and worthy of being a content topic about numeracy. For further research, it is recommended to carry out item suitability tests from all measurement approaches, namely classical test theory and item response theory (2 PL, 3 PL, 4 PL), until a model is synthesized that best fits the numeracy test that has been developed. Apart from that, the use of Javanese cultural topics can be developed in numeracy tests at elementary and high school levels.

■ **REFERENCES**

Anam, F., Suteja, J. R., Septianto, A., Purnomo, A., & Utami, Y. P. (2020). Improving the numeracy mathematics ability: The role of abacus learning model. *Journal of Physics: Conference Series*, *1594*(1), 12041. IOP Publishing.

Ardiyanti, D. (2016). *Aplikasi model rasch pada pengembangan skala efikasi diri dalam pengambilan keputusan karir siswa* [Application of the rasch model in

developing a self-efficacy scale for students' career decision-making]. *Jurnal Psikologi*, *43*(3), 248–263.

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3th ed.). Psychology Press.

Bond, T. G., Yan, Z., & Heene, M. (2020). Applying the rasch model: Fundamental measurement in the human sciences. In *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (4th ed.). New York: Routledge. https://doi.org/10.4324/9780429030499

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). Rasch analysis in the human sciences. Springer.

Budiman, A., Hufad, A., & Purnomo, P. (2022). Development of campus potential-based integration education tourism model. *4th International Conference on Arts and Design Education (ICADE 2021)*, 21–24. Atlantis Press.

Degroote, L., DeSmet, A., De Bourdeaudhuij, I., Van Dyck, D., & Crombez, G. (2020). Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *17*, 1–13.

Elvira, M., Retnawati, H., Rohaeti, E., & Sainuddin, S. (2023). Measurement of students' chemistry practicum skills using many facets rash model. *European Journal of Educational Research*, *12*(3).

Fadhilah, V., Setiawan, F., Indriani, T., & Yulianti, S. (2022). *Analisis kebijakan asasemen nasional pengganti ujian nasional terhadap evaluasi pendidikan di indonesia. Alsys*, *2*(4), 514–523.

Fernanda, J. W., & Hidayah, N. (2020). *Analisis kualitas soal ujian statistika menggunakan classical test theory dan rasch model* [Analysis of the Quality of Statistics Exam Questions Using Classical Test Theory and the Rasch Model]. *Square: Journal of Mathematics and Mathematics Education*, *2*(1), 49–60.

Gay, G. (2018). Culturally responsive teaching: Theory, research, and practice. teachers college press.

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082–1116.

Greer, B., & Mukhopadhyay, S. (2015). Honoring diversity in intercultural mathematics education. *Intercultural Education*, Vol. 26, pp. 261–265. Taylor & Francis.

Guilford, J. P. (1950). Fundamental statistics in psychology and education.

Hall, J., & Zmood, S. (2019). Australia's literacy and numeracy test for initial teacher education students: Trends in numeracy for low-and high-achieving students. *Australian Journal of Teacher Education (Online)*, *44*(10), 1–17.

Hong, J., Thakuriah, P. (Vonu), Mason, P., & Lido, C. (2020). The role of numeracy and financial literacy skills in the relationship between information and communication technology use and travel behaviour. *Travel Behaviour and Society*, *21*, 257–264. https://doi.org/10.1016/j.tbs.2020.07.007

Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *REiD (Research and Evaluation in Education)*, *5*(2), 95–102.

Issatriadi, I., Rokhmad, U., & Timoer, S. (1977). *Sejarah seni budaya daerah jawa timur [History of Art and Culture in East Java]*. Direktorat Jenderal Kebudayaan.

Kurniawan, A. P., Budiarto, M. T., & Ekawati, R. (2022). *Pengembangan soal numerasi berbasis konteks nilai budaya primbon Jawa. JRPM (Jurnal Review Pembelajaran Matematika)*, 7(1), 20–34.

Kustantina, V. A., Nuryadi, N., & Marhaeni, N. H. (2022). *Meningkatkan kemampuan literasi numerasi siswa menggunakan komik matematika interaktif pada materi teorema phytagoras* [Improving students' numeracy literacy skills through the use of interactive mathematics comics on pythagorean theorem]. *Jurnal Ilmu Pendidikan Muhammadiyah Kramat Jati*, 3(1), 10–16.

Linacre, J. M. (2024). *A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*. winsteps.com.

Mai, F., Sun, J., Tan, Q., & Zhao, B. (2024). A preliminary investigation of niche clothing brand design optimization based on ADDIE model. *Applied Mathematics and Nonlinear Sciences*, 9(1). https://doi.org/10.2478/amns.2023.2.00894

Mawaddah, S., Noorbaiti, R., Aulia, M., Eryanto, A. N. E., & Mahlina, O. (2022). *Instrumen asesmen kompetensi minimum numerasi konteks lingkungan lahan basah khas Kalimantan Selatan* [Minimum competency assessment instrument for numeracy in the context of wetland environments in south kalimantan]. *EDU-MAT: Jurnal Pendidikan Matematika*, 10(1), 24–32.

Meihami, H., & Malmir, A. (2024). Student-teachers' professional agency development through ADDIE model of CALL teacher preparation. *Language Teaching Research*, 13621688241254908. https://doi.org/10.1177/13621688241254907

Nortvedt, G. A., & Wiese, E. (2020). Numeracy and migrant students: A case study of secondary level mathematics education in Norway. *ZDM*, 52(3), 527–539.

Novikasari, I., Muttaqin, A., & Elebiary, N. (2024). Teaching math and preserving culture: the intersection of values in indonesian pedagogy. In *Values and Valuing in Mathematics Education: Moving Forward into Practice* (pp. 361–379). Springer.

O'Shaughnessy, D. M., Cruz Cordero, T., Mollica, F., Boni, I., Jara-Ettinger, J., Gibson, E., & Piantadosi, S. T. (2023). Diverse mathematical knowledge among indigenous Amazonians. *Proceedings of the National Academy of Sciences*, 120(35), e2215999120.

OECD. (2018). *Education at a Glance 2018/ : OECD Indicators*. Paris: OECD Publishing.

Pinxten, R. (2016). Plea in favor of 'trivial mathematics' in a multimathemacy educational perspective. *Intercultural Education*, 27(2), 188–200.

Primaniarta, M. G., & de Mattos, J. R. L. (2022). Javanese Ethnomathematics: Exploration of the" Tedhak Siten" Tradition for Class Learning Practices. *Journal on Mathematics Education*, 13(4), 661–680.

Purnomo, H., Sa'dijah, C., Hidayanto, E., Permadi, H., & Anwar, L. (2022). Development of instrument numeracy skills test of minimum competency assessment (MCA) in Indonesia. *International Journal of Instruction*, 15(3), 635–648.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian [Quantitative*

Analysis of Research Instruments]. Yogyakarta: Parama Publishing.

Rohendi, R. (2022). *Analisis kemampuan literasi matematis siswa smp kelas viii pada materi statistika* [analysis of mathematical literacy skills of eighth grade junior high school students on statistics]. *SCIENCE/ : Jurnal Inovasi Pendidikan Matematika Dan IPA*, *2*(3), 398–404. https://doi.org/10.51878/science.v2i3. 1591

Romm, A. T., Kika, J. C., & Schoer, V. (2019). A test taker's gamble: The effect of average grade to date on guessing behaviour in a multiple choice test with a negative marking rule. *South African Journal of Economic and Management Sciences*, *22*(1), 1–12.

Sainuddin, S. (2018). *Analisis karakteristik butir tes matematika berdasarkan teori modern (teori respon butir)* [Analysis of characteristics of mathematics test items based on modern test theory (item response theory)]. *Jurnal Penelitian Matematika Dan Pendidikan Matematika*, *1*(1), 1–12. Retrieved from http://www.journal.uncp.ac.id/index.php/ proximal/article/view/842/727

Stafford, C. (2008). Linguistic and cultural variables in the psychology of numeracy. *Journal of the Royal Anthropological Institute*, *14*, S128–S141.

Sugiyono, S. (2018). *Metode penelitian pendidikan pendekatan kualitatif, kuantitatif dan R & D* [Educational research methods: qualitative, quantitative, and r&d approaches]. *Alfabeta, Bandung*.

Sumarni, S., Prayitno, A. T., Syafari, R., Basir, M. A., Febrianti, D., & Putri, A. (2023). *Instrumen tes kemampuan literasi numerasi berbasis budaya lokal kabupaten kuningan jawa barat* [Numeracy literacy test instrument based on local culture in kuningan regency, west

java]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(1), 789. https://doi.org/10.24127/ajpm.v12i1. 6439

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch/ : untuk penelitian ilmu-ilmu sosial (revisi) [application of the rasch model: for social sciences research (revised edition)]*. Trim Komunikata Publishing House.

Suryonegoro, B. M., Wuryastuti, M. L., & Dewi, N. R. (2024). Literature review: inquiry social complexity-steam model based on math trail-virtual reality activity nuanced with javanese culture in improving critical thinking ability. *Journal Evaluation in Education (JEE)*, *5*(2), 89–99.

Tangianu, F., Mazzone, A., Berti, F., Pinna, G., Bortolotti, I., Colombo, F., … Filannino, C. (2018). Are multiple-choice questions a good tool for the assessment of clinical competence in Internal Medicine? *Italian Journal of Medicine*, *12*(2), 88–96.

Widarti, H. R., Rokhim, D. A., Septiani, M. O., & Dzikrulloh, M. H. A. (2022). Identification of science teacher practices and barriers in preparation of minimum competency assessment in the covid-19 pandemic era. *Orbital: The Electronic Journal of Chemistry*, 63–67.

Yang, D.-C., & Sianturi, I. A. J. (2021). Sixth grade students' performance, misconception, and confidence on a three-tier number sense test. *International Journal of Science and Mathematics Education*, *19*(2), 355–375. https:// doi.org/10.1007/s10763-020-10051-3

Yu, M. V. B., Liu, Y., Soto Lara, S., Puente, K., Carranza, P., Pantano, A., & Simpkins, S. D. (2021). Culturally responsive practices: Insights from a high quality math afterschool program serving underprivileged Latinx youth. *American Journal of Community Psychology*, *68*(3–4), 323–339.