ACER

ICAL 2024

Learning from evidence:
Assessing progress and impact

10-12 October 2024 | Bali, Indonesia

International Conference on Assessment and Learning

Learning From Evidence: Assessing Progress and Impact. Proceeding of the 2024 International Conference on Assessment and Learning

ACER Indonesia
Jl. Hang Tuah IX No. 9
Gunung, Kebayoran Baru
Jakarta 12120
Indonesia
Phone: +62 21 2751 6999
www.acer/org/id

ISBN 978-1-74286-792-2

© 2025 Australian Council for Educational Research

Citation APA 7
Iskandar, B. (Ed.). (2025). Learning From Evidence: Assessing Progress and Impact. Proceeding of the 2024 International Conference on Assessment and Learning. ACER Indonesia. https://doi.org/10.37517/978-1-74286-792-2

The opinions represented in these articles belong to the authors and this compilation does not warrant the accuracy of any conclusion expressed.

Table of Contents

contents

01.

Rubrics on the Radar: Contestation from Indian Higher Education | *Aarti Srivastava, Pankaj Sarkar, Joshita Negi* 11

02.

Advancing the Measurement of Fluid Reasoning: A Rasch Model Validation of The Faxtor Cognitive Ability Test | *Afiya Nur Nabila, Rafli Sodiq Bagaskara* 22

03.

The Development and Validation of the Career Interest Test for High School and First-year College Students in Indonesia | *Afiyana Eka Nurilla, Dewi Maulina, Adiyo Roebianto, Khairunissa Syafira Dumbi, Audinia Shabrina* 32

04.

Evaluation of The Standard Setting for The UKPPG Pre-Service Knowledge Test in 2024 Between General Fields of Study and Specialized Fields of Study | *Andri Sutrisno, Thomas Riyadi* 39

05.

The Influence of Mental Health and Perceived Social Support on Grit Among Young Lecturers | *Anggi Mayangsari, Irma Rosalinda Lubis, Iriani Indri Hapsari* 47

Table of Contents

contents

06.

Developing Collaborative Behavior Scale for State-Owned Enterprise Employees | *Anindita Damayanti, Isma Junida, Angela Oktavia Suryani, Laura F. N. Sudarnoto* 53

07.

Transformation of The State University Entrance Selection to Encourage Inclusive Access to Higher Education in Indonesia | *Arya Gamma Aditia, Lista Kurniawati* 60

08.

Measurement of Educators' Digital Literacy Competence | *Demitria Dini Ariyani, Maria Sumunaringtyas, Nugro Krismanto* 67

09.

Urgency of Following up on National Assessment Results | *Ema Noviah, Aip Badrujaman, Dinny Devi Triana* 78

10.

PEMANTIK Assessment as Learning Catalysator: Assisting Learning Ecosystem to Improve Children Literacy and Numeracy Skill | *Eri Yusron, Annisa Al Hadi, Angela Merici G. Adem* 84

Table of Contents

contents

11. **A Qualitative Analysis of The Effectiveness, Challenges, and Opportunities of Implementing The Assessment of Indonesian Madrasah Competencies** | *Eviana Hikamudin, Bambang Suryadi, Suprananto* 100

12. **Formulating Designer-Friendly Guidelines for Developing Game-Based Assessments to Measure Conscientiousness** | *Fadillah, R. Hidayat, A. Santoso* 109

13. **Contextualised Delivery for Early Childhood Education Personnel in Southwest Sumba** | *Flory Abiwawanti, Budiarti Rahayu, Hani Yulindrasari* 129

14. **Alignment of Teacher and Student Perceptions on Learning Methods** | *Hanifan Nur Fitrianto, Siti Raudlah* 141

15. **College-Career-Readiness Assessment Based on Inclusive and Technology Environment in Indonesia** | *Hasna Hafizhah Salma, Gary Collins Brata Winardy, Muhammad Nanang Suprayogi, Siti Elda Hiererra, Keisha Putri Theanny, Dorothy Situmorang, Cindi Nabila Maharani* 145

Table of Contents

contents

16. **Evaluation of The Combined CIPP and Goal-Oriented Model Program on Acceleration of Stunting Reduction with Specific Interventions** | *Helvy Yunida, Awaluddin Tjalla, Mahdiyah* 152

17. **Evaluation of The Situational Judgement Test (SJT) Question Format: An Effort to Measure Teachers' Competence More Equitably** | *Iin Nurul Khoiriyah, Fairuzia Rahmadiyanti Widyadhari* 164

18. **Advancing Educational Assessment in Indonesia: Developing a Problem-Solving Tool Aligned with Kurikulum Merdeka** | *Ilham Phalosa Reswara, Katie Makar* 170

19. **Beyond The Classroom: Autonomous Learning Challenges in Emancipated Curriculum** | *Indah Pratiwi, Bakti Utama, Ika Hijriani* 176

20. **Evaluation of The Alignment of Mechanical Talent Subtest Assessment Results with The Selection of Engineering and Technical Study Programs in The National Selection Based on Achievement** | *Khansa Fahira Syafitra, Nova Indriani* 185

Table of Contents

contents

21. Evaluation of Community-Organized MSME Business Incubation Program Planning | *Kinkin Suartini, Ari Saptono, Riyadi* 190

22. Indonesian National Assessment for Madrasah Students: The Analysis of Different Types of Numeracy Test Items | *Kusaeri, Noor Wahyudi, Tatik Indayati, Calista Octavian Dwisanti, Elmita Irmanila* 204

23. Development of A Situational Judgement Test Blueprint to Design Five Factor Model Personality Profiles Based on Diagnostic Classification | *Medianta Tarigan, Rahmat Hidayat, Agung Santoso* 214

24. Enhancing Teacher Skills: Effectiveness of 'Using Presentation Software Adaptively as An Interactive Learning Media' Training Module | *Novita Sari, Anissa Cessarea, Billy Fariman, Kahfi Hizbullah* 232

25. Psychometric Properties of The Pauline Test: A Next-Generation Work Aptitude Test on The Talentlytica Platform | *Novita Sari, Wahyu Maulana Firdaus, Aswin Januarsjaf* 248

Table of Contents

contents

26. Iterating Professional Development Program for School Leaders & Teachers Using Kirkpatrick Framework | *Putri Aulia Silkana, Cici Tri Wanita* 259

27. Estimating The Impact of A Whole-School Improvement Intervention Using National Assessment Metrics: Lessons Learned from PINTAR Program | *Radi Negara* 265

28. Assessing Gender Bias in Vocational Interest Measurement: A Differential Item Functioning of The Eureka Interest Inventory | *Rafli Sodiq Bagaskara, Nadim Alfan Assa, Whisnu Yudiana* 275

29. Enhancing Competence in The Malaysian Construction Industry: Challenges and Solutions in The Assessment Process | *Saiful Bahri Bin Hasan Basri* 284

30. Engineering Mechanics Instruments as A Measurement Tool for Vocational School Students' Competency: A Literature Review in The Field of Technology and Engineering | *Salma Maharani, Riyan Arthur, Soeprijanto* 294

Table of Contents

contents

31.

Maritime Students' Perception of Video-Making Projects to Practice Speaking Skills | *St. Hariati Buchari, Dyah Rathaningsih, I Dewa Ayu Indira Nandhita* 304

32.

How Indonesian Students Perform in Mathematics: Learn from PISA 2022 | *Suci Paramitha Liestari, Zul Arsiah, Rina Rosdiani Patria* 314

33.

The Development of Marital Life Satisfaction Scale| *Tino Dwi Prasetiyo, Narayya Nurshadrina, Laura F. N. Sudarnoto, Angela Oktavia Suryani* 332

34.

Academic Dishonesty in EFL Students' Writings: Students' Reflections in The Era of Artificial Intelligence | *Tri Purba Dharmastuti* 341

35.

Adaptation of CRS and BESSI to Assess Pancasila Student Profile | *Tri Wahyuni, Wardani Rahayu, Deli Santun Naga* 353

Table of Contents

contents

36. Examining Item Writers' Perception on Complex Multiple Choice Questions | *Triyanto, Andrew Bony Nabasar Manurung* 360

37. The Development of As Learning and for Learning Assessment Instruments in Problem-Based Learning to Measure Elementary School Learners' Social and Spiritual Attitudes as well as Creative and Critical Thinking Skill Competencies | *Undang Rosidin, Fatkhur Rohman, Nina Kadaritna, Elza Maisinur* 368

38. Gender Differences in PISA and National Assessment Performance in Indonesia: A Biopsychosocial Model | *Wahyu Nurhayati, Haryo Susetyo* 380

39. Construction of Student Academic Stress Scale (Preliminary Study) | *Weni Endahing Warni, Dimas Riztiardhana* 390

40. Insights from "Teaching Grading": The Peer Observation of Teaching Program at The Global Islamic Boarding School | *Zulfikar Alimuddin, Sauqina* 397

Indonesian National Assessment for Madrasah Students: The Analysis of Different Types of Numeracy Test Items

Kusaeri[1]
Noor Wahyudi[1]
Tatik Indayati[1]
Calista Octavian Dwisanti[2]
Elmita Irmanila[3]

[1]Universitas Islam Negeri Sunan Ampel Surabaya
[2]SMA Progresif Bumi Shalawat Sidoarjo
[3]LKP Science Society Sidoarjo

https://doi.org/

This article is part of the research funded by UIN Sunan Ampel Surabaya through the Rector's Decree Number 180 of 2023. The researchers expressed their infinite gratitude to the Rector, his staff, and the Center for Scientific Research and Publication team at LPPM (Institute for Research and Community Service) UIN Sunan `Ampel Surabaya. The authors are also grateful to the Ministry of Religion of the Republic of Indonesia for giving access and permission to publish the AKMI data.

Abstract

The types of questions can influence students' responses and the characteristics of the items, such as the index of difficulty and the index of discrimination. Hence, more studies on using various types of questions are worth doing to better understand how test takers answer and complete the questions in various forms. This research aims to reveal students' difficulties in doing various types of questions in numeracy tests and further investigate the students' difficulties with the types of questions that students mostly gave the wrong answers. The sample was 1,773 students out of 297,257 fifth graders of Islamic schools participating in the Indonesian Madrasah Competency Assessment in 2022 administered by the Ministry of Religious Affairs of the Republic of Indonesia. The instruments were question items in five types, with texts as the stimuli. The data analyses used descriptive statistics (mean and standard deviation) and t-tests. The findings are: (a) numeracy questions using matching and short answers have the lowest response, i.e., most students find them challenging, and (2) the difficulty with matching type can potentially be caused by the lack of practice during the learning process while with the short answer type may be due to the application system. The findings imply that teachers and Islamic schools need to give students more practice to help them get accustomed to these types of question items. Examples, assignments or projects for homework with various types of questions may facilitate students to get familiarized with various types of questions. Students also need more practice with short answer questions to enable them to do the questions based on the rules.

Introduction

The use of question types in assessment is a matter of debate in the education evaluation system in Indonesia

There has been a long and ongoing debate on issues related to various types of question items used in assessment in education as different types of questions may bring different responses of the test takers having diverse abilities (Bennett, 2023; Birenbaum & Tatsuoka, 1967; Hastedt & Sibberns, 2005; Liou & Bulut, 2020; Skedsmo & Huber, 2020). The types of questions commonly used to assess students' cognitive performance include multiple-choice, complex multiple-choice, true-false, matching, short-answer and essay (Tozoğlu et al., 2004). The types of questions can influence students' responses (Delaram & Sharifi, 2014) and the characteristics of the items, such as the index

of difficulty and the index of discrimination (Haladyna & Rodriguez, 2013; Kan et al., 2019). Hence, more studies on using various types of questions are worth doing to better understand how test takers answer and complete the questions in various forms (Moon et al., 2018).

Analysis of various questions is essential for teachers to get accurate information about the types of questions to construct and administer to measure students' competencies. The appropriate types of questions can serve as reliable instruments to get the accurate result of the measurement relevant to the targeted construct (Liou & Bulut, 2020). The proper selection of instruments (types of questions in particular) will provide valid information about students' competencies (Hohensinn & Kubinger, 2009; Moon et al., 2018).

The choice of questions type influences students' cognitive results

Current studies on types of questions have mainly focused on psychometric assessments (Liou & Bulut, 2020), (Brassi & Couch, 2019; Finney & McFadden, 2023; Moon et al., 2020) and comparing question types between multiple choice and short answer or essay in various disciplines (Butler, 2018; Hollingworth et al., 2007; LaFave et al., 2022; Pepple et al., 2010; Rahmawati & Retnawati, 2019). Studies focusing on psychometric assessment and question item characteristics mainly compare question items in the form of selected responses such as multiple choice and complex multiple-choice, multiple choice and true-false. The studies' focus has been the index of difficulty, index of discrimination, and guessing from the perspectives of Classical Test Theory (CTT) and Item Response Theory (IRT).

Studies comparing types of multiple-choice and essays show that students have better responses to multiple-choice questions compared to essays. Multiple choice questions need less cognitive processes, while **essays require** higher intellectual ability to answer the questions through analysis, application, and comprehension. Therefore, students potentially find problems answering essay questions compared to the multiple choice. Yet, few studies report on students' ability to respond to other types of questions, such as complex multiple-choice, matching, short answer and true-false.

One study compared Turkish students' ability to respond to multiple-choice questions, complex multiple-choice, and essay types of PISA 2003-2012 (Özkan & Özaslan, 2018). The essay type mainly had incorrect answers, followed by complex multiple-choice. Multiple-choice questions tend to be easier compared to essays and complex multiple-choice. Another study compared complex multiple-choice questions to multiple-choice essays. The findings show that complex multiple-choice is commonly used compared to essays (Albanese, 1982). The other finding highlights that multiple-choice questions are more manageable, reliable, and valid than complex ones.

Findings from previous studies indicate the limited number of studies comparing the students' responses or difficulties in answering various types of questions. Hence, studies simultaneously portray various students' responses to different questions that are deemed essential.

Type of numeracy questions in Indonesian Madrasah Competency Assessment-AKMI

Based on data from a national assessment from the Ministry of Religious Affairs of the Republic of Indonesia (termed as Indonesian Madrasah Competency Assessment-AKMI), there is a strong potential to have an in-depth and comprehensive study on students' responses and difficulties in answering various types of question. Other studies using AKMI data (Kusaeri, Lailiyah, et al., 2022; Kusaeri, Yudha, et al., 2022; Oktaviani, 2023) have presented insights and horizons on the issue. However, those studies have not yet addressed issues related to students' difficulties in responding to various types of questions used in AKMI, particularly the numeracy test items. Hence, further research on the issue is highly essential.

The results of AKMI 2022 show that 92% of students in Islamic primary schools are at the initial competencies (Need Assistance, Capable, Basic), and 8% are at the Skilled level. None of the students are at the level of Creative (Yusqi, 2022), the highest level used in the AKMI framework. This suggests the low numeracy skills of students in Islamic primary schools in Indonesia. One of the possible causes of this lack of numeracy skills is the variety of question types used in AKMI. It is assumed that teachers mainly use multiple choice and essay types in daily practice. Hence, the affordance category

(Gibson, 2015) is appropriate to initially portray students' difficulty responding to various questions in AKMI.

From the perspective of affordance theory, students' perceptions toward the types of questions in an assessment can drive an action in the form of a response or answer (Moon et al., 2020). Various forms of answers result from a variety of question types. A specific type of question will stimulate a particular response based on the students' experience. For example, if in the teaching and learning process, students never experience doing complex-multiple choice questions, they will respond to this type of question similarly to general multiple-choice questions (Mills & Harrison, 2020). They may think there is only one correct answer to any multiple-choice question. The complex multiple-choice question needs two correct answers. This potential misperception of the students due to their unfamiliarity with the question type inspires the current research to explore further.

The purpose of this study

The objectives of the research are to (a) reveal the students' difficulty in doing various types of questions in numeracy tests (multiple choice, complex multiple-choice, true-false, matching, and short answer) in AKMI and (b) investigate any potential problems students have on the types of question in the items they mostly gave wrong responses or answers. With these two objectives, the research can identify the types of questions that can effectively portray the cognitive abilities and complexity of students' thinking skills in numeracy among students in Indonesian Islamic schools. The holistic picture of the phenomenon can provide a basis for any intervention relevant to students' weaknesses. The follow-up intervention will support students to get accustomed to various types of questions through examples and projects that students can work on at home.

Method

We conducted the study by following the International Test Commission (2017) guidelines for translating and adapting tests. We received both permission and the Indonesian translation of DCI from the original author. The Indonesian translation DCI was translated by another researcher in a non-psychometric study a few years ago (Dermawan et al., 2015). We continued working on the translated version by taking the next steps, which were backward translation until pilot study.

Participant

The participants of this research were all the fifth graders of Islamic primary schools participating in the National Madrasah Competency Assessment of the Ministry of Religious Affairs of the Republic of Indonesia in 2022. The total participants were 297,257 students from 11,061 Islamic primary schools throughout Indonesia. Out of 11.061 Islamic primary schools, 706 are public, and 10,895 are private (Ramdhani, 2022). After checking the students' answers, 1,773 students were selected as the sample of the research. The other 295,484 students did not give complete answers to all numeracy test items and were excluded from the analysis

Measures

Research instruments

The instrument used in this research is the instrument of AKMI in five question types: multiple choice (MC), complex multiple-choice (CMC), matching, multiple choice true-false (MCTF), and short answer. The variations of question types become an integrative stimulus in the form of texts presented in relevance to the content and context of numeracy. The stimulus in each numeracy question item characterizes AKMI question items. This integration is expected to trigger students to get used to reading and enhance the literacy level of students in Islamic schools. Through the tight process of instrument development, internal and external review, readability test by teachers and students and field test of the instruments. The instrument developers went through tight national selection by the Ministry of Religious Affairs team of the Republic of Indonesia.

The piloting of the instruments shows that out of 2,121 items of the numeracy test, the index of discrimination ranges from 0.05 to 8.936, and the index of difficulty is between -28.785 to 0.074. To ensure that the test items can function well as assessment instruments and provide valid information, the items used in this research are those within the discrimination index range of 0.2 to 4

and difficulty index of -4 to 4 (Umar et al., 2022). Those items are distributed proportionally in the MC, CMC, matching, MCTF and short answer types.

Data analysis

Descriptive statistical analysis (mean and standard deviation) reflecting the characteristics of each question type was measured to collect data relevant to the main focus of the research. Further analysis using a t-test was aimed at statistically measure which types of numeracy questions students mostly found problematic. The analysis used SPSS version 25. To corroborate the statistical data, there are examples of question items that students mostly give improper responses to. This corroboration further highlights the statistical findings under Gibson's affordance theory.

Findings

Display of ability-type questions of AKMI

In general, the data show that the MCTF questions perform better as an assessment instrument than other questions (Table 1). The student's ability to respond to this type of question is evenly distributed (seen from the mean and standard deviation). This differs slightly from the MC and Complex MC types, which have a lower meaning and more minor standard deviation. The matching and short answer questions have the lowest meaning and standard deviation among other types.

Table 1. Descriptive Statistics of 5 (Five) Types of Question

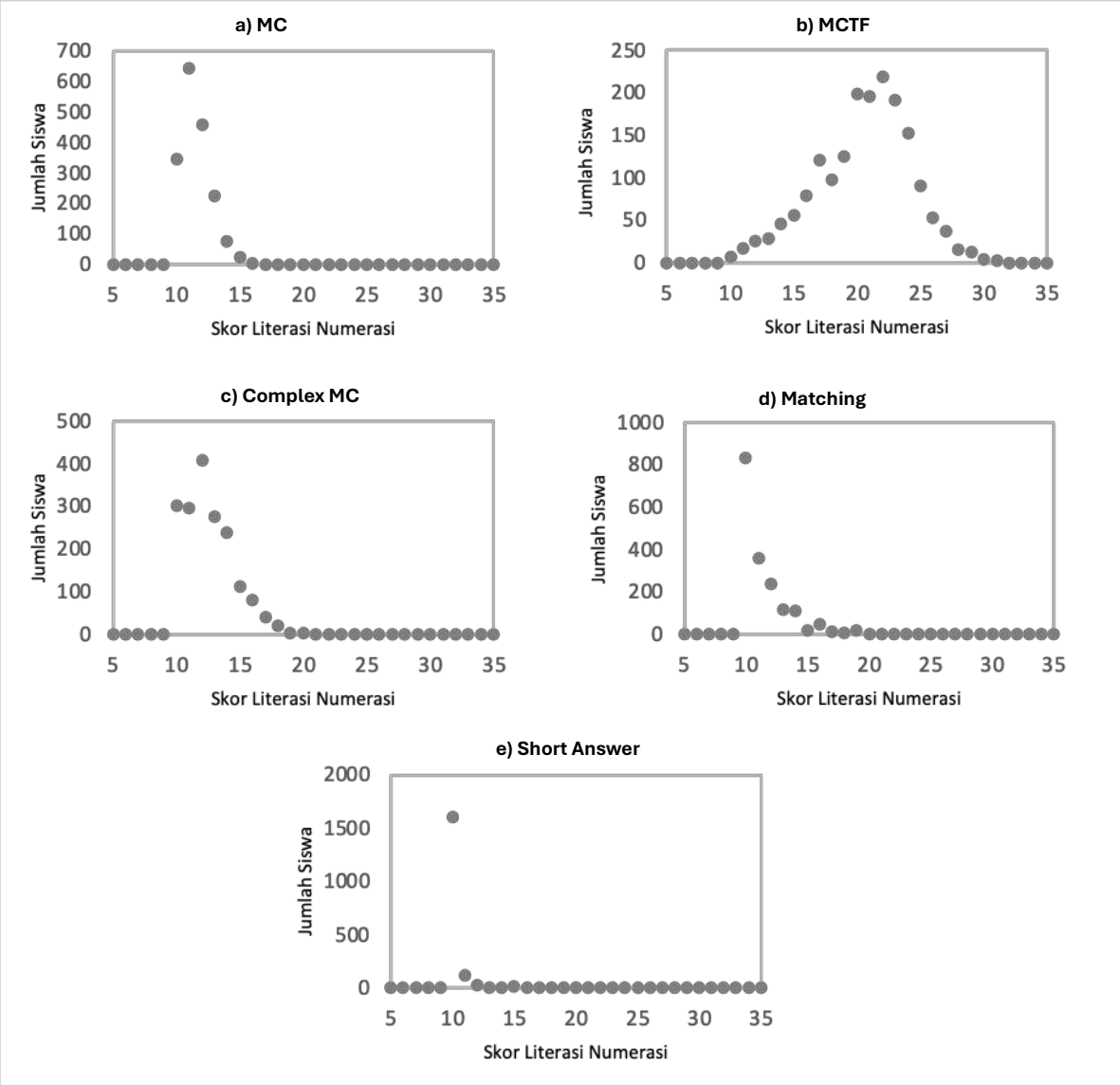
Dimension	N	Mean	Std. Deviation
MC	1773	11.50	1.154
Complex MC	1773	12.48	1.924
MCTF	1773	20.55	3.707
Matching	1773	11.34	1.860
Short Answer	1773	10.13	0.484

Numeration literacy capability is reviewed from various types of questions

Picture 1 presents the distribution of the numeracy test scores of the students. In conjunction with data presented in Table 1, both data show that all five types of questions enable students to get the maximum numeracy scores. However, students are better at doing only the MCTF type than the others.

Further exploration of the test items that have no correct answers shows that matching and short answer types contribute to students' failure to provide correct answers, each with 49.20% and 90.47%. These percentages highlight that both types of questions contribute to the low numeracy skills of the students participating in AKMI. Further analysis is needed to better understand which of the four types of questions (MC, Complex MC, matching and short answer) has the highest difficulty level.

Picture 1. The Distribution of Numeracy Scores of 5 (Five) Types of Question



Comparison results of two types of questions with the t-test

Statistic measurements using a t-test (Table 2) show that matching and short answer types are the two most difficult types of questions, with the Sig of both types being less than 0.005. This result signifies that matching and short answers are the two most difficult compared to other types of numeracy questions.

Table 2. Result of t-test across Types of Question

Comparison of Types of Questions	F	Sig	Mean Difference	Std. Error Difference
MC and Matching	181.962	0.002	0.16469	0.05199
MC and Short Answer	1513.268	0.000	1.37563	0.02973
MC and Complex MC	390.319	0.000	-0.97800	0.05330
Complex MC and Matching	17.331	0.000	1.14270	0.06356
Complex MC and Short Answer	2168.584	0.000	2.35364	0.04713
Matching and Short Answer	1425.348	0.000	1.21094	0.04564

Field review

Results from field exploration highlight the following important information. Ms Winahyu, S. Pd. (MI Negeri Pangkajene Kepulauan Sulawesi Selatan) stated that: "As for the short answer, most students were confused with the instruction. For example, in one item, the correct answer is 1,000. Some students answered 1000, "one thousand" or having the one thousand with an apostrophe ".", that is "1,000". It is possible that the application cannot read these varieties of answers and conclude the answers as incorrect." The same concern is reflected by Mr. Nanang Rianto (MIN 2 Batam Kepulauan Riau), "students have problems with short-term questions. They are used to this type of question but in paper-based form. They can answer more than one word, which does not follow the conditions of answers accepted by the system used for AKMI." Both teachers have the same concerns with the application used for AKMI. The possible cause of many incorrect answers can be the desynchronisation between the various forms of students' answers with the system used in AKMI.

Discussion

The findings confirm the affordance theory in that student's perceptions of the types of questions and applications used in AKMI stimulate various answers (Moon et al., 2020). Data on matching and short answer types confirm the theory. The application system requires students to give short answers in one word or phrase for short answer questions. The students often missed this instruction and gave answers longer than the requirement. Moreover, most of the students used mobile phones to do the assessment. Using mobile phones may create another challenge as the students need to type the answer letter by letter (Mohammadi et al., 2020). This step creates problems for the students and influences the end results of their assessment.

The different mode of doing the assessment, from paper-based to computer–or mobile-based, seems to cause the assessment to be daunting. Research Lloyd et al., (2012) and Andriyani et al., (2021) found that students find it challenging to use technology such as mobile phones in this situation. Students unfamiliar with computers or mobile phones have more challenges than those who have had the experience. In this context, student's ability to use computers and mobile phones becomes the condition for them to be able to do the test in this context. Hence, teachers should provide test simulations to the students before the actual administration of AKMI. The simulation can potentially reduce students' anxiety and the daunting mode of the assessment. Hence, students will be familiar with the application's features and get used to the mode and webpage of the assessment.

The findings also highlight that data on matching and short answer questions also confirm the affordance theory in that student's ability to complete these two types of questions is influenced by their perception (Moon et al., 2018, 2020). Students are unfamiliar with the matching type, in which they must match the premise with the response. In their learning processes or previous assessments, it seems they do not interact with this type of question. Hence, Islamic primary school teachers must give students more exercises on this question during teaching and learning. This will not only familiarize students with this matching type of question but will also make them accustomed to various types of questions that will increase their confidence when doing national and international assessments (such as PISA). So, more regular practice in doing matching and short answer questions is needed for teaching and learning in class.

Another contributing factor to the low response to the two types of questions is students' level of seriousness in doing the AKMI. They think that AKMI will not significantly influence their pursuit of further education. They do not think AKMI is essential and do the assessment in a shorter time than the allotted time. Hence, students need to be given orientation about the impact of AKMI on their future education and consider it essential. So, they will take the assessment seriously. The schools need to give socialization about the benefits and impact of AKMI on the learning environment at schools. With this socialization, students can be more serious about doing AKMI and get optimal results.

To better understand students' difficulty completing the question item with matching type, below is the example of the item with the lowest number of correct responses. This example provides information on the item's characteristics that students find difficult (in terms of the construction and the form of the question).

The Makmur Regency

One of the government programs is to build subsidized housing to provide low-income people with affordable and livable houses. The Makmur Regency developer will build a subsidised housing complex, as presented in the picture. The land for the complex is 48 meters and 21 meters wide. The house to build will be three types, each with 36m2, 45 m2, and 72 m2 and a 3-meter-wide road. Please select the pair that best suits the following statement from this information.

The land area for each complex is the same (…)
The land area for the road is less than half of the total area of the whole complex (…)
[-] the land area is 108 m2. (2)
[-] the land area is 144 m2. (3)
[-] the land area is 63 m2. (4)
Key: 2|3

Komplek Perumahan Makmur

Salah satu program pemerintah adalah pembangunan rumah bersubsidi dengan tujuan membantu masyarakat berpenghasilan rendah memiliki rumah layak huni. Pengembang Perumahan Makmur akan membangun komplek perumahan bersubsidi seperti tampak pada gambar. Panjang tanah komplek perumahan seluruhnya adalah 48 m dan lebar 21 m. Rumah yang akan dibangun direncanakan dalam tiga tipe yaitu, rumah dengan luas 36 m^2, 45 m^2 dan 72 m^2, serta dilengkapi dengan jalan selebar 3 m.

Gambar Denah Komplek Perumahan Makmur

Setelah mencermati informasi tersebut, pilihlah pasangan yang tepat untuk setiap pernyataan berikut.
Luas tanah untuk setiap komplek perumahan adalah sama.(.....)
Luas tanah untuk pembuatan jalan kurang dari setengah luas komplek perumahan seluruhnya. (.....)

[-] Luas tanah adalah 108 m^2. (2)
[-] Luas tanah adalah 144 m^2. (3)
[-] Luas tanah adalah 63 m^2. (4)
Kunci : 2|3

The above item assesses students' ability to "do mathematical operation, present data, make models or interpret problems related to facts, relations, processes, concepts, or procedure." To complete this type of question, students need logical reasoning to interpret the problems related to mathematical operations. This still becomes the main challenge for the students(Kusaeri, Lailiyah, et al., 2022). Students are also unfamiliar with matching questions requiring them to match the premise and response. This type of question has not been widely used for the level of Islamic primary education. Hence, students find it challenging to understand the instructions for responding to the question. Therefore, in the teaching and learning processes, teachers need to provide students with more types of questions as their exercises, such as matching type. Students are expected to be familiar with various types of questions and know how to respond to each type. Therefore, more maximal results of AKMI can be expected.

Conclusion

Some essential points of this research are highlighted: (1) the numeracy questions having low responses and bringing more difficulties to the students are those in the types of matching and short answer and (2) lack of practice during the teaching and learning process, the requirement for the students to put premises and responses become the main difficulty in matching type. As with the short answer type, the problem is with the application system that requires students to have short and concise answers. In fact, students often give elaborate answers that exceed the required length of answer.

The findings are relevant to the affordance theory. The student's perception of the unfamiliar types of questions and the application system used in AKMI seem to cause incorrect responses from the students. Hence, the teacher and Islamic schools need to facilitate students with more practice during teaching and learning through various types of examples, assignments and projects they can do at home. Hence, the students can answer correctly based on the test conditions.

References

Albanese, M. A. (1982). Multiple-choice items with combinations of correct responses. Evaluation & the Health Professions, 5(2), 218–228. https://doi.org/10.1177/016327878200500205

Andriyani, S., Sa'idah, N., Yulistianti, H. D., & Hamid, A. (2021). Students' Perspectives In Online Learning; Descriptive Study on Students of The Faculty of Tarbiyah and Teacher Science at Unisnu Jepara. Psychology and Education Journal, 58(2), 7358–7365.

Bennett, R. E. (2023). Toward a Theory of Socioculturally Responsive Assessment. Educational Assessment, 28(2), 83–104. https://doi.org/10.1080/10627197.2023.2202312

Birenbaum, M., & Tatsuoka, K. K. (1967). On scoring models for tests with multiple response analyses.

Brassi, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education*, 6(1). https://doi.org/10.1186/s40594-019-0169-0

Butler, A. C. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition*, 7(3), 323–331. https://doi.org/10.1016/j.jarmac.2018.07.002

Delaram, M., & Sharifi, A. (2014). A comparison of students' scores in multiple-choice and essay questions in mother and child health examination. *Scientific Information Database*.

Dermawan, A., Goei, R., & Kirana, D. (2015). Validitas alat ukur Dyadic Coping Inventory versi Bahasa Indonesia: sebuah penelitian non-psikometrik [Validity of the Indonesian version of the Dyadic Coping Inventory: A non-psychometric study]. *Jurnal Psikologi*, *12*(1), 1–15. *(In Indonesian)*

Finney, S. J., & McFadden, M. E. (2023). Examining the Question-Behavior Effect in Low-Stakes Testing Contexts: A Cheap Strategy to Increase Examinee Effort. *Educational Assessment*, 1(1), 1–8. https://doi.org/10.1080/10627197.2023.2222588

Gibson, J. J. (2015). *The ecological approach to visual perception* (Reprint ed.). Psychology Press.

Haladyna, T., & Rodriguez, M. C. (2013). *Developing and Validating Test Items* (1st ed). Routledge.

Hastedt, D., & Sibberns, H. (2005). Differences between multiple-choice items and constructed response items in the IEA times surveys. *Studies in Educational Evaluation*, *31*(2–3), 145–161.

Hohensinn, C., & Kubinger, K. D. (2009). On varying item difficulty by changing response format for a mathematical competence test. *Psychology Science Quarterly, 51*(4), 345–362.

Hollingworth, L., Beard, J. J., & Proctor, T. P. (2007, December). *An investigation of item type in a standards-based assessment*. *Practical Assessment, Research & Evaluation, 12*(18). Retrieved from https://pareonline.net/pdf/v12n18.pdf

Kan, A., Bulut, O., & Cormier, D. C. (2019). The Impact of Item Stem Format on the Dimensional Structure of Mathematics Assessments. *Educational Assessment, 24*(1), 13–32. https://doi.org/10.1080/10627197.2018.1545569

Kusaeri, K., Lailiyah, S., Arrifadah, Y., & Asmiyah, S. (2022). Enhancing creative reasoning through mathematical task: The quest for an ideal design. *International Journal of Evaluation and Research in Education, 11*(2), 482–490. https://doi.org/10.11591/ijere.v11i2.22125

Kusaeri, K., Yudha, Y. H., Kadarisman, Y. P., & Hidayatullah, A. (2022). Do Instructional Practices by Madrasah Teachers Promote Numeracy? *Jurnal Pendidikan Matematika IPA*, 13(1), 29–38.

LaFave, A. J., Taylor, J. A., Barter, A. M., & Jacobs, A. S. (2022). Student Engagement on the National Assessment of Educational Progress (NAEP): A Systematic Review and Meta-Analysis of Extant Research. *Educational Assessment, 27*(3), 205–228. https://doi.org/10.1080/10627197.2022.2043151

Liou, P. Y., & Bulut, O. (2020). The effects of item format and cognitive domain on students' science performance in TIMSS 2011. *Research in Science Education, 50*(1), 99–121.

Lloyd, S., Chalder, T., & Rimes, K. A. (2012). Family-focused cognitive behavioural therapy versus psycho-education for adolescents with chronic fatigue syndrome: Follow-up of an RCT. *Behaviour Research & Therapy, 50*(11), 719–729. https://doi.org/10.1016/j.brat.2012.08.011

Mills, V. L., & Harrison, C. (2020). Intentional professional learning design: Supporting teacher growth. *Educational Assessment, 25*(4), 331–354. https://doi.org/10.1080/10627197.2020.1766961

Mohammadi, M. S., Sarvestani, M. S., & Nouroozi, S. (2020). Mobile phone use in education by technical-engineering faculty members: Concurrent mixed-methods. *Frontiers in Education, 5*, Article 16. https://doi.org/10.3389/feduc.2020.00016

Moon, J. A., Keehner, M., & Katz, I. R. (2018). Affordances of item formats and their effects on test-taker cognition under uncertainty. *Educational Assessment Quarterly, 30*(4), 345–360.

Moon, J. A., Keehner, M., & Katz, I. R. (2020). Test-takers' response tendencies in alternative item formats: A cognitive science approach. *Educational Assessment, 25*(3), 236–250. https://doi.org/10.1080/10627197.2020.1804350

Oktaviani, E. C. (2023). Analyzing Madrasah readiness for the Indonesian Madrasah Competency Assessment: Study at MTs Negeri 2 Bekasi City. *Journal of Education, Administration, Training, and Religion, 4*(1), 49–58.

Özkan, Ö., & Özaslan, N. (2018). Student achievement in Turkey by question type in PISA math tests. *International Journal of Evaluation and Research in Education, 7*(1), 57–64.

Pepple, D. J., Young, L. E., Carroll, R. G., & Pepple, D. J. (2010). Essay vs. multiple-choice: MBBS Stage I physiology exam scores at UWI. *Advances in Physiology Education, 34*, 86–89. https://doi.org/10.1152/advan.00087.2009

Rahmawati, F., & Retnawati, H. (2019). Student difficulty in solving PISA-like math problems. *IOP Conference Series: Journal of Physics: Conference Series, 1200*(1), 012015. https://doi.org/10.1088/1742-6596/1200/1/012015

Ramdhani, M. A. (2022). Reformasi Pendidikan Madrasah Indonesia: Sebuah keharusan. *Prosiding Seminar Nasional AKMI*.

Skedsmo, G., & Huber, S. G. (2020). Unintended effects of accountability policies on assessment quality. *Educational Assessment, Evaluation and Accountability, 32*(4), 427–429. https://doi.org/10.1007/s11092-020-09338-8

Tozoğlu, D., Tozoğlu, M. D., Gürses, A., & Doğar, Ç. (2004). The students' perceptions: Essay versus multiple-choice exams. *Journal of Baltic Science Education, 2*(6), 52–59.

Umar, A., Kusaeri, A., Ridho, A., Yusuf, A., & Asyhar, A. H. (2022). Does opportunity to learn explain the math score gap between Madrasah and non-Madrasah students in Indonesia? *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan, 41*(3), 792–805. https://doi.org/10.21831/cp.v41i3.40169

Yusqi, M. I. (2022). Reformasi mutupembelajaran madrasah: Peningkatan kompetensi literasi Net Generation di madrasah. Prosiding Seminar Nasional Hasil AsesmenKompetensi Madrasah Indonesia (AKMI) 2022.

ACER
TUT WURI HANDAYANI
IKHLAS BERAMAL
HIMPUNAN EVALUASI
HEPI
PENDIDIKAN INDONESIA
APSIMETRI
HIMPSI