Jurnal Iqra'
Kajian Ilmu Pendidikan

OPEN ACCESS

Volume 10, Number 3, December 2025

Rasch-Based Cognitive Diagnostic Assessment to Measure Critical Thinking Skills in Islamic Religious Education

Agil Bahsoan [1*], Nurhayati Tine [1], Ivan Rahmat Santoso [1], Neva Lionitha Ibrahim [1], Branislav Pupala [2]

[1] *Universitas Negeri Gorontalo, Indonesia*

[2] *Tranava University Slovakia, Slovakia*

agil@ung.ac.id *

ARTICLE INFO

Article history:

Received
July 20, 2025
Revised
October 05, 2025
Accepted
November 16, 2025

ABSTRACT

This research aimed to measure the critical thinking skills of students in the Islamic Religious Education course through the application of a cognitive diagnostic assessment based on the Rasch model. The background of this research stems from the need for an assessment instrument that not only provides summative scores but also diagnoses students' abilities in a more specific and in-depth manner. It employed a quantitative design involving 107 students, and the developed instrument was validated using Rasch analysis to ensure validity, reliability, and item fit. The results showed that the majority of students are in the moderate category (52,34%), followed by the high category (30,84%), and the low category (16,82%). These results highlight the diversity of students' critical thinking skills, which require adaptive teaching strategies. The novelty of this research lies not only in the application of the Rasch model in cognitive diagnostic assessment but also in its implementation in the context of Islamic Religious Education, which has traditionally been measured through conventional tests emphasizing memorization and conceptual understanding. This approach allows for a more detailed mapping of critical thinking sub-skills in values- and ethics-based courses, going beyond previous research practices that generally focus on science or technical skills. Thus, this research contributes to the development of more accurate, adaptive, contextual, and relevant assessment instruments to support improvements in the quality of learning in higher education.

Keywords: Cognitive Diagnostic Assessment, Rasch Model, Critical Thinking Skills, Islamic Religious Education, Higher Education.

Published by Institut Agama Islam Ma'arif NU (IAIMNU) Metro Lampung
Website https://journal.iaimnumetrolampung.ac.id/index.php/ji/index
E-ISSN 2548-7892

INTRODUCTION

Global changes marked by the Industrial Revolution 4.0 and the transition to Society 5.0 have presented new challenges for the world of education. Universities are no longer content to simply produce graduates proficient in technical skills; they are also required to prepare a generation with higher-order thinking skills, including critical thinking skills. These skills are crucial because they serve as the foundation for students to process information, analyze problems, evaluate arguments, and make decisions that are not merely rational but also based on values, ethics, and social sensitivity (Brookhart, 2010; Darminto et al., 2025; Jansen & Möller, 2022; Lewis & Smith, 1993). In the context of higher education in Indonesia, critical thinking skills are a key indicator for measuring the quality of learning, particularly in courses that integrate cognitive, affective, and spiritual dimensions.

Islamic Religious Education (IRE) in public universities plays a crucial role in shaping students' perspectives on religious, social, and national issues, not only normatively but also critically and reflectively, enabling them to adopt a moderate attitude (Une et al., 2015). At the

Faculty of Economics and Business, Universitas Negeri Gorontalo, students are required to connect Islamic principles with professional practice, even though many have limited religious backgrounds. Therefore, the learning approach must be dialogic, analytical, and contextual. However, the effectiveness of Islamic Education is difficult to measure because assessment instruments are still summative, while diagnostic assessments based on the Rasch Model are not yet available.

A limitation of summative assessment is its tendency to only display final student achievement without providing in-depth information about the cognitive processes that occur during learning (Barbeau et al., 2025; Darabi Bazvand & Rasooli, 2022). With this approach, lecturers only receive final grades but do not obtain a detailed picture of misconceptions, gaps in understanding, or individual potential for further development. However, developing critical thinking skills requires assessments capable of comprehensively diagnosing students' cognitive processes. This is where cognitive diagnostic assessments become relevant, as they can provide detailed information about students' strengths, weaknesses, and learning needs.

Previous research has shown that diagnostic assessments in higher education play a crucial role in improving learning approaches, Laaziz et al. (2024) emphasized that diagnostic assessments not only identify students' initial capacities but also help instructors design learning strategies tailored to individual needs, Zhou et al. (2024) even linked educational attainment to cognitive function and emphasized that higher education can slow cognitive decline. Thus, diagnostic assessments are not merely evaluation tools but also essential instruments for building adaptive learning experiences.

The Rasch model offers significant advantages over classical analytical approaches. According to Bond (2015) and Azis & Lubis (2023), the Rasch model can analyze item quality, difficulty level, instrument reliability, and even the distribution of student abilities. This allows assessments to not only provide final scores but also map students' critical thinking skills more objectively (Alnahdi, 2019; Chi et al., 2021; Nielsen et al., 2017; Vaughan, 2018). Supriyadi et al. (2022) emphasize the importance of validity and reliability of assessment instruments, while Lovendra & Aisiah (2023) add that clear question formulation and appropriate testing methods are prerequisites for assessments to truly measure student understanding. With the support of Rasch analysis, assessment instruments can be developed more accurately and used as a basis for learning improvement. Furthermore, following up on diagnostic assessment results is equally important. Forniawan & Wati (2024) emphasize that assessment results should be used as a reference for designing more adaptive learning activities. Students with low understanding require additional support, while students with high abilities require greater challenges. This means that Rasch Model-based diagnostic assessments serve not only as measurement but also as a driver for the realization of differentiated and inclusive learning strategies.

Several previous and recent studies related to cognitive diagnostic assessment (CDA) for critical thinking skills have developed assessments from various perspectives and approaches (Chrismawaty et al., 2022; Ebright-Jones & Cortina, 2025; Jiang et al., 2023; Kleemola et al., 2022; Lun et al., 2023; Luo et al., 2024; Miller & Long, 2022; Ng et al., 2022; Song & Cai, 2024; Susanti et al., 2024). Although recent research indicates increasing interest in assessing and developing critical thinking among college students, significant gaps remain, particularly in terms of measurable diagnostic models, long-term interventions, and contextually adaptive methods. A more robust integration of the CDA framework could address these weaknesses by providing detailed and individualized skill profiles.

Based on these issues, this research aimed to describe the application of a Rasch Model-based CDA to measure students' critical thinking skills in the Islamic Religious Education course at the Faculty of Economics and Business, Universitas Negeri Gorontalo. It addressed questions regarding the design of a valid and reliable assessment instrument, the quality of the instrument after analysis using the Rasch Model, the revealed profile of students' critical thinking skills, and the implications of the assessment results for learning strategies. This research is expected to make a significant contribution to the development of educational assessments based on modern measurement theory, expand the application of the Rasch Model in cognitive diagnostic assessment, and assist lecturers in designing more effective, adaptive

IRE learning strategies oriented toward developing students' critical thinking skills in a complex global era.

METHOD

This research was conducted at the Faculty of Economics and Business, Universitas Negeri Gorontalo, on students taking the IRE course. The research is planned to take place in the odd semester of the 2025/2026 academic year, starting from the stages of instrument development, validation, data collection, analysis, and preparation of the research report. This research employed a descriptive-quantitative design with Rasch analysis, where this design was aimed at describing the distribution and profile of students' critical thinking skills rather than to examine causal relationships (Creswell, 2017). In the meantime, the Rasch model was used as it provides several methodological advantages compared with classical test theory, including the ability to (1) evaluate item validity and reliability more objectively, (2) map item difficulty and student ability on the same logit scale, and (3) produce diagnostic information that enables classification of students into low, moderate, and high ability groups (Bond et al., 2020; Sumintono, 2014). This diagnostic capacity is particularly crucial in the context of Islamic Religious Education (IRE), where assessments have traditionally relied on summative tests emphasizing memorization. By using Rasch-based Cognitive Diagnostic Assessment, this present research offers a more detailed and objective mapping of critical thinking sub-skills, providing both methodological rigor and contextual novelty in value- and ethics-based learning environments. The population in this research was all students of the Faculty of Economics and Business, Universitas Negeri Gorontalo, who were enrolled in the IRE course. The distribution of students by study program and class is shown in Table 1 below:

Table 1. Population of Students Attending Islamic Religious Education Course, Faculty of Economics and Business, Universitas Negeri Gorontalo

Study Program	Class	Number of Students
Bachelor of Economics Education	A	29
	B	28
	C	27
	D	28
	E	28
	F	28
	G	28
Subtotal		196
Bachelor of Accounting	A	30
	B	30
	C	29
	D	30
	E	30
	F	30
	G	30
Subtotal		209
Bachelor of Management	A	28
	B	28
	C	28
	D	28
	E	28
	F	30
	G	30
	H	30
	I	30
Subtotal		260

Bachelor of Development Economics	A	18
	B	20
	C	23
	D	20
	E	8
Subtotal		89
Total		**754**

The sample in this research was determined using purposive sampling, with the primary consideration being the representativeness of classes across each study program in the Faculty of Economics and Business. This approach was chosen to ensure that the selected classes reflected similar learning contexts and comparable levels of prior knowledge. In Rasch measurement, sample adequacy is more related to item calibration stability than to population representativeness (Linacre, 1994). Linacre's guideline suggests that a minimum of 30–50 respondents is sufficient for preliminary item calibration, while 100–150 respondents provide stable and reliable estimates across most conditions. In this current research, a total of 107 students were involved, which meets the adequacy requirement for Rasch analysis.

Thus, purposive sampling was employed to ensure contextual representativeness, and the sample size fulfilled methodological requirements for Rasch measurement. The research sample is shown in Table 2 below:

Table 2. Research Sample

Study Program	Code	Number of Students
Economics Education	A	29
Accounting	B	30
Management	C	28
Development Economics	D	20
Total		107

Thus, the total sample size of the research was 107 students

Data collection was conducted through:

1. A cognitive diagnostic test, administered directly to the sample students.
2. Documentation, in the form of attendance lists, lesson plans, and student work.

The variable in this research was students' critical thinking skills, which were measured through a CDA in the IRE course. The research instrument was a multiple-choice cognitive diagnostic test with 15 questions, including:

Table 3. Test Instrument Indicators

No	Aspect/Variable	Number of Items	Critical Thinking Indicators (Ennis)	Statement/Question
1	Islamic Creed	1	Providing elementary clarification	Students can explain the meaning of faith in Allah in their own words.
2	Islamic Creed	2	Building basic support	Students can provide reasons why believing in angels is important.
3	Islamic Creed	3	Making inferences	Students can summarize the consequences of not believing in *qadha* and *qadar*.
4	Worship	4	Providing simple explanations	Students can explain the main purpose of prayer in daily life.
5	Worship	5	Providing advanced clarifications	Students can analyze the difference between obligatory and optional prayers.
6	Worship	6	Setting	Students can plan ways to

			strategies and tactics	maintain consistency in performing prayers on time.
7	Islamic Law	7	Providing simple explanations	Students can explain the difference between permissible and forbidden (halal and haram) in Islam.
8	Islamic Law	8	Making inferences	Students can draw conclusions from simple cases related to usury in everyday life.
9	Islamic Law	9	Setting strategies and tactics	Students can determine appropriate steps to avoid sinful acts.
10	Religious Tolerance	10	Providing simple explanations	Students can explain the meaning of tolerance between religious communities.
11	Religious Tolerance	11	Providing advanced explanations	Students can analyze the importance of maintaining harmony between religious communities in society.
12	Religious Tolerance	12	Setting strategies and tactics	Students can design appropriate attitudes when facing differences in beliefs with friends.
13	Love for the Homeland	13	Providing simple explanations	Students can explain the meaning of love for the homeland in daily life.
14	Love for the Homeland	14	Making inferences	Students can summarize the negative impacts if the younger generation does not love their homeland.
15	Love for the Homeland	15	Setting strategies and tactics	Students can determine concrete steps to maintain national unity.

In addition to the cognitive diagnostic test, documentation data were also collected, including attendance lists, course syllabi (RPS), and samples of student work. These documents were then integrated with the test results through data triangulation. Attendance lists provided contextual information about students' consistency and engagement during the course, whereas the RPS served as a reference to ensure alignment between the intended learning outcomes and the measured critical thinking skills. Meanwhile, student work, such as assignments and written reflections, was thematically analyzed to identify evidence of critical thinking processes, such as clarification, inference, and strategic reasoning. By triangulating test results with documentation, the validity of findings was strengthened, and the interpretation of students' critical thinking profiles became more comprehensive.

The development of the cognitive diagnostic test instrument followed several systematic steps. First, a blueprint was constructed based on Ennis' (Ennis, 1985, 1990) indicators of critical thinking and contextualized within the content of Islamic Religious Education. Second, the items were reviewed through expert judgment involving two experts in Islamic Education and one expert in educational measurement to ensure content validity, clarity, and relevance. Third, the instrument underwent a pilot test with a group of 30 students who were not part of the main study. Feedback from the pilot was used to revise ambiguous wording and adjust the difficulty levels of the items. Fourth, the finalized version of the test contained 15 multiple-choice items, representing five thematic areas (Islamic Creed, Worship, Islamic Law, Religious Tolerance, and Love for the Homeland).

Prior to administration, it was expected that items requiring basic clarification (e.g., definitions and explanations of creed or worship) would be easier for students, while items requiring strategic and tactical reasoning (e.g., planning consistent prayer, designing tolerant attitudes, or maintaining national unity) would be more difficult. Although the final test consisted of 15 items, this number is sufficient in Rasch analysis because instrument reliability depends more on item quality and calibration than on the quantity of items (Bond & Fox, 2013). Previous studies have demonstrated that short tests with 10–20 items can still produce stable measures when the items are well-constructed and adequately cover the target construct (Sumintono & Widhiarso, 2014).

Data Analysis

Data were analysed using the Rasch Model with Winsteps software. Several criteria were employed to ensure instrument quality and validity:

- Item Fit: Items were considered acceptable if Outfit Mean Square (MNSQ) values were within 0.5–1.5 and standardized Z-values (ZSTD) ranged between –2 and +2 (Bond & Fox, 2013). Items outside these ranges were further examined for potential revision. Reliability: Both person and item reliability were reported. Reliability values of ≥ 0.70 were considered good, while the Cronbach's Alpha equivalent in Rasch (KR-20) above 0.60 was regarded as minimally acceptable (Sumintono & Widhiarso, 2014).
- Unidimensionality: Principal Component Analysis (PCA) of residuals was conducted to test whether the instrument measured a single construct. Following Linacre (1994), the measurement was considered unidimensional if the variance explained by Rasch measures exceeded 40% and the unexplained variance in the first contrast was less than 2 eigenvalue units. Ability Levels: Students' ability were categorized based on logit scores. Respondents with logits below -0.5 were classified as low, between -0.5 and +0.5 as moderate, and above +0.5 as high (Stone & Wright, 1999).

The analysis procedure was then carried out in six stages:

1. Testing item validity (MNSQ, ZSTD, and Point Measure Correlation).
2. Testing instrument reliability (person and item reliability).
3. Examining instrument unidimensionality using PCA of residuals.
4. Analyzing item difficulty levels (logit values).
5. Mapping students' abilities using the Wright Map (person–item map).
6. Providing descriptive interpretation of critical thinking results in low, moderate, and high categories.

The research procedure was carried out through four main stages: preparation, implementation, data analysis, and reporting. The first stage was preparation, where at this stage, the researcher compiled a diagnostic assessment question grid tailored to the topics in the IRE course and indicators of critical thinking skills. After compiling the framework, we developed an instrument consisting of multiple-choice questions covering aspects of Islamic Creed, Worship, Islamic Law, Religious Tolerance, and Love for the Homeland.

The second stage was research implementation. During this stage, we administered a cognitive diagnostic test to the students who served as the research sample. The test was administered directly in class according to the Islamic Religious Education lecture schedule. All student responses were then collected to serve as the primary data for the research. Then, the third stage was data analysis. The collected data from the test results were processed using software using the Rasch Model approach. The analysis was conducted to test the validity and reliability of the instrument, the difficulty level of the questions, and to map students' critical thinking skills using the Wright Map. The results were then interpreted based on critical thinking skill indicators to determine the students' ability profiles.

The final stage was reporting the research results. During this stage, we systematically compiled a research report based on the results obtained from the data analysis. Furthermore, we formulated recommendations regarding the application of CDA in IRE teaching, which can be used as input for lecturers in developing more effective learning strategies oriented toward students' critical thinking skills.

RESULT AND DISCUSSION

Validity Test of Question Items

The results of the Item Statistics: Misfit Order data processing showed the quality of the instrument items based on the fit parameters in the Rasch model. The information displayed includes the total score, measure value, standard error, and infit index and outfit mean square (MNSQ), along with ZSTD. The ideal MNSQ value is in the range of 0,5 – 1,5 and the ZSTD value is around -2 to +2, so the item is considered to fit the model if it is still within that range. In addition, the item correlation with the point-measure correlation measure is also displayed (Bond & Fox, 2013). A good point-measure correlation value indicates that the item has a positive contribution to the construct being measured. Alagumalai et al. (2005) classify these values into very good (>0,40), good (0,30–0,39), fair (0,20–0,29), unable to distinguish (0,00–0,19), and requires examination of the item (<0,00).

Table 4. Validity Test Based on Item Fit Order Test Analysis

Item	Outfit MNSQ	Outfit ZSTD	Pt Corr	Meaning
6	1.48	3.2	0.23	Fit
8	1.47	3.0	0.33	Fit
9	1.32	1.0	0.23	Fit
14	1.26	1.5	0.29	Fit
4	1.18	1.3	0.34	Fit
3	1.07	0.5	0.39	Fit
5	0.98	–0.1	0.41	Fit
1	0.85	–0.1	0.25	Fit
2	0.61	–0.7	0.32	Fit
11	0.61	–0.7	0.32	Fit
12	0.61	–0.7	0.32	Fit
10	0.84	–1.3	0.56	Fit
7	0.71	–2.5	0.61	Fit
15	0.63	–2.7	0.61	Fit
13	0.62	–2.6	0.61	Fit

Based on the Rasch analysis results, item quality was assessed using three main indicators: Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr. The acceptance criteria were: (1) Outfit MNSQ within the range of 0,5–1,5; (2) Outfit ZSTD within the range of -2 to +2; and (3) Pt Measure Corr being positive and not approaching zero. If at least two of the three criteria are met, the item can be declared good (Sumintono & Widhiarso, 2014). The results showed that most items met these criteria, with Pt Measure Corr values ranging from sufficient to excellent. However, several items required attention due to their lack of consistency across all indicators, particularly the ZSTD value, which tends to be influenced by sample size. Overall, this instrument can be considered reliable for use, but revisions to several items that showed discrepancies are still recommended to optimize the instrument's quality.

Reliability Test of Instrument

Based on the results of the Rasch Model analysis of 107 respondents and 15 questions, a statistical summary was obtained as presented in Table 2. This analysis provides an overview of the quality of respondents (persons) and the quality of questions (items) in the instrument used.

Table 5. Analysis of Reliability Test on Output Summary Statistics

Statistics	Person (Respondent)	Item
Total	107 respondents	15 items
Mean Score	10,7	76,6
Mean Measure	64,08	50,00
Standard Deviation Measure	12,53	11,09
Separation	1,09	3,72
Reliability	0,54	0,93
Cronbach Alpha (KR-20)	0,64	

The Rasch analysis results showed that the respondent reliability was 0,54 with a separation of 1,09, indicating that the consistency of the answers was in the moderate category and the instrument's ability to differentiate between groups of respondents was quite good, although not optimal in grouping ability levels. Conversely, in terms of item reliability, the value reached 0,93 with a separation of 3,72, which falls in the very good category. This means that the instrument is able to distinguish the level of item difficulty very well, and the overall quality of the questions is sufficient. Furthermore, the Cronbach Alpha (KR-20) value of 0,64 indicates that the overall reliability of the instrument is in the quite acceptable category.

Instrument Unidimensionality Analysis

Instrument unidimensionality analysis was conducted to ensure that each statement item truly measures critical thinking skills as one main construct. In the Rasch approach, unidimensionality testing can be seen through Principal Component Analysis (PCA) of Residuals. The results of the PCA analysis on this research instrument are displayed in the Standardized Residual Variance table.

Figure 1. Unidimensionality Test Results

```
TABLE 23.0 agil data                          ZOU812WS.TXT  Aug 27 22:44 2025
INPUT: 107 Person  15 Item  REPORTED: 107 Person  15 Item  2 CATS  WINSTEPS 3.73
-------------------------------------------------------------------------------

          Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                              -- Empirical --    Modeled
  Total raw variance in observations     =        17.4 100.0%        100.0%
    Raw variance explained by measures   =         2.4  14.0%         13.9%
      Raw variance explained by persons  =          .8   4.5%          4.5%
      Raw Variance explained by items    =         1.7   9.5%          9.4%
    Raw unexplained variance (total)     =        15.0  86.0% 100.0%   86.1%
         Unexplned variance in 1st contrast =        1.7   9.8%  11.4%
         Unexplned variance in 2nd contrast =        1.6   9.3%  10.8%
         Unexplned variance in 3rd contrast =        1.5   8.5%   9.8%
         Unexplned variance in 4th contrast =        1.4   7.9%   9.2%
         Unexplned variance in 5th contrast =        1.3   7.5%   8.7%
```

Based on the table, the total data variance is 17,4 (100%). Of this, the variance explained by the instruments (measures) is 2,4 (14,0%), with the following breakdown: variance explained by respondents (persons) at 0,8 (4,5%), and variance explained by instrument items at 1,7 (9,5%). Meanwhile, the unexplained variance remains relatively large at 15,0 (86,0%). In terms of contrast, the largest variance appears in the first contrast with a value of 1,7 (9,8%), followed by the second contrast at 1,6 (9,3%), the third contrast at 1,5 (8,5%), the fourth contrast at 1,4 (7,9%), and the fifth contrast at 1,3 (7,5%).

The variance explained by the instruments (14,0%) is not significant. However, according to the Rasch analysis criteria, if the unexplained variance in the first contrast is still around or below 2 eigenvalue units, then the instrument can be said to tend to be unidimensional. As a result, the first contrast of 1.7 is still within the tolerance limit. Thus, it can be concluded that the instrument used has a tendency to be unidimensional, even though the contribution of explained variance is relatively low.

Analysis of the difficulty level of the question items

In Rasch analysis, item difficulty is indicated by the logit value. The higher the logit value, the more difficult the item is for respondents to answer correctly and vice versa. This analysis is important for assessing the distribution of item difficulty and determining whether the instrument is balanced in measuring respondents' abilities.

Table 6. The Difficulty Level of The Question Items

Item	Variable	Statement	Measure (Logit)	Difficulty	Interpretation
1	Islamic Creed	Students can explain the meaning of faith in Allah in their own	-0,33	Easy	The majority of students are capable, as it only requires basic

		words.			explanation skills (Ennis: basic clarification).
2	Islamic Creed	Students can provide reasons why faith in angels is important.	-0,25	Easy	Relatively easy, students only need to provide simple reasons (Ennis: elementary clarification).
3	Islamic Creed	Students can summarize the consequences of not believing in *qadha* and *qadar*.	+0,12	Medium	Requires the ability to infer cause and effect (Ennis: inference).
4	Worship	Students can explain the main purpose of prayer in daily life.	-0,40	Easy	Easy because it relates to daily worship experiences (Ennis: basic clarification).
5	Worship	Students can analyze the difference between obligatory and optional prayers.	+0,28	Medium	Requires comparative analysis (Ennis: basis for decision).
6	Worship	Students can plan ways to consistently pray on time.	+0,75	Difficult	High because it involves planning and self-evaluation (Ennis: strategy & tactics).
7	Islamic Law	Students can explain the difference between halal and haram in Islam.	-0,22	Easy	Easy, simply remembering basic concepts (Ennis: basic clarification).
8	Islamic Law	Students can draw conclusions from a simple case of usury.	+0,18	Medium	Requires the ability to draw conclusions from cases (Ennis: inference).
9	Islamic Law	Students can determine steps to avoid sinful acts.	+0,60	Difficult	Difficult, because it requires planning concrete actions (Ennis: strategy & tactics).
10	Religious Tolerance	Students can explain the meaning of tolerance between religious communities.	-0,35	Easy	Easy, simply explaining general concepts (Ennis: basic clarification).
11	Religious Tolerance	Students can analyze the importance of maintaining religious harmony.	+0,22	Medium	Requires social cause and effect analysis (Ennis: inference).
12	Religious Tolerance	Students can design a stance when facing differences in belief.	+0,82	Difficult	Difficult, because it involves strategic attitude planning (Ennis: strategy & tactics).
13	Love for the Homeland	Students can explain the meaning of love for one's country in	-0,28	Easy	Easy, because it is declarative knowledge (Ennis: basic

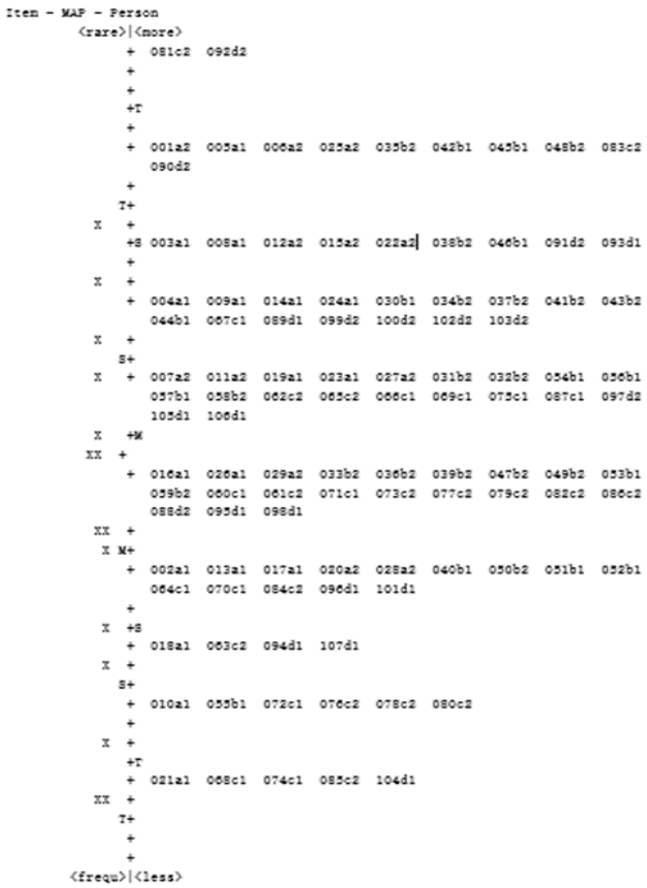
		everyday life.			clarification).
14	Love for the Homeland	Students can summarize the negative impacts if the younger generation does not love their country.	+0,30	Medium	Requires the ability to infer social consequences (Ennis: inference).
15	Love for the Homeland	Students can determine concrete steps to maintain national unity.	+0,95	Difficult	The most difficult requires strategic planning (Ennis: strategy & tactics).

The analysis of the instrument, which consisted of 15 items covering five main variables (Islamic Creed, Worshipping, Islamic Law, Religious Tolerance, and Love for the Homeland), showed that each item had a different level of difficulty based on the logit value. A lower logit value indicates that the item is relatively easier for students to answer or understand, while a higher logit value indicates that the item is relatively more difficult and requires higher critical thinking skills.

Mapping student abilities in using a Wright Map (person-item map)

An item-person mapping analysis (Item-Person Map/Wright Map) was conducted to determine the correspondence between respondents' ability levels and the difficulty levels of the instrument items. This mapping depicts the relative position of the respondent (person) on the left side and the items on the right side based on the logit values generated from the Rasch analysis. This allows us to determine whether the item difficulty level aligns with the respondents' abilities and how well the instrument is able to differentiate between different abilities.

Figure 2. Person Map



Based on the results of the Item-Person (Wright Map) mapping analysis, the distribution of respondents' ability levels compared to the difficulty of the items is visible. The left side of the map displays respondents' ability using an identity code, while the right side displays the item difficulty level. The logit scale used illustrates the relative position between respondents' ability and item difficulty, with higher positions on the logit scale indicating higher ability or difficulty.

Most respondents were distributed in the medium to high logit range, indicating a fairly even variation in ability with a tendency toward the upper levels. This indicates that the instrument is capable of measuring respondents with medium to high ability levels. The items analyzed mostly fell within the medium logit range, with some items occupying lower and higher positions, indicating variation in difficulty levels. The distribution of items tended to be concentrated in the middle, indicating that the instrument primarily measured ability at a medium level. The mapping results also showed that some respondents occupied higher logit positions than the highest difficulty items. This indicates that some respondents had a higher ability than the available items, making the instrument not entirely challenging for them. Conversely, some respondents fell below the easiest item position, indicating a low-ability group who experienced difficulty answering even relatively easy items. Overall, the match between the distribution of respondents' abilities and the item difficulty levels can be said to be quite good.

Descriptive interpretation of critical thinking skills results (low, medium, high categories).

Based on the results of the Rasch analysis of 15 items measuring students' critical thinking skills, a descriptive interpretation can be made by dividing the measurement results into three categories: low, medium, and high. This categorization refers to the logit distribution of respondents' abilities compared to the distribution of item difficulty. Students in the low category generally had difficulty answering relatively easy items (negative logit), such as explaining the meaning of faith in Allah, distinguishing between halal and haram, or explaining the meaning of love for one's country. This indicates that they are still limited to mastering simple declarative knowledge and are not yet fully capable of basic clarification (Ennis, 1985, 1990). In the medium category, students were able to answer items with a medium level of difficulty (logit close to zero to moderately positive), such as deducing the consequences of not believing in *qadha* and *qadar*, analyzing the differences between obligatory and sunnah prayers, or analyzing the importance of interfaith harmony. This indicates that their critical thinking skills already encompass inference and analysis, although they are not yet fully consistent when faced with more complex strategic demands.

Meanwhile, students in the high category were able to complete items with large logits (high positive), such as planning how to maintain consistency in prayer, determining steps to avoid sin, designing attitudes when facing differences in belief, and determining concrete steps to maintain national unity. These items require advanced critical thinking skills (Thi Nhat et al., 2018), which involve the ability to formulate strategies, make decisions, and connect abstract concepts with concrete actions. Students in this category demonstrate a more mature capacity for reflective thinking and planning. In general, the distribution of respondents shows that the majority of students fall into the medium to high category, with only a small minority falling into the low category. This indicates that the instrument successfully distinguishes variations in students' critical thinking abilities and demonstrates that the majority of respondents have mastered critical thinking skills at the analytical and strategic levels, although a small minority still requires attention in developing basic skills.

DISSCUSSION

The results provide a comprehensive overview of the instrument's quality, item difficulty, reliability, and the distribution of student abilities. Overall, it indicated that the developed instrument is sufficient for measuring critical thinking skills, although several aspects still require improvement. The item validity analysis revealed that all items fit the Rasch model, although there was variation in the Outfit ZSTD scores, some of which fell outside the ideal range. This is acceptable considering the influence of sample size on ZSTD scores, so the

instrument remains valid overall. The instrument's reliability also qualifies as good, particularly for the items, which have very high reliability (0,93) with a separation of 3,72. This demonstrates that the questions clearly differentiate levels of difficulty, although respondent reliability remains moderate (0,54), indicating a tendency toward homogeneity in student abilities. These findings suggest that while the instrument provides robust evidence of item quality, the moderate person reliability reflects a persistent gap in students' critical thinking distribution, which requires further pedagogical attention.

The results of the unidimensionality analysis indicate that this instrument tends to be unidimensional, with a first contrast value of 1,7, which is still within the tolerable limits. This means that the developed items truly measure one main construct, namely critical thinking skills, although the percentage of explained variance is still relatively low. It also suggests that the instrument requires further refinement to optimally capture students' critical thinking constructs. Based on item difficulty, the logit distribution shows a variation in the questions, ranging from easy to medium to difficult. Items related to simple explanations (e.g., items 1, 4, 7, 10, and 13) have low logit values and are therefore considered easy. This is understandable because these types of questions only require students to reproduce conceptual knowledge that is relatively close to everyday experience. Conversely, items that require strategic and tactical skills (e.g., items 6, 9, 12, and 15) have the highest logit values and are considered difficult. This means that students still face challenges in developing plans, strategies, or applicable actions based on conceptual understanding. These results support the argument that students' critical thinking skills are still stronger in basic clarification than in advanced analysis and strategic planning.

The Rasch analysis showed that items requiring basic clarification and inference were relatively easier for students, whereas those related to strategy and tactics were the most difficult. This pattern is consistent with Ennis (1985) critical thinking framework, where basic clarification and inference are fundamental components of critical thinking, while advanced clarification, which involves planning, strategy, and evaluation, requires higher-order reasoning. Similarly, Facione (1990), in accordance with the Delphi report, emphasizes that evaluation and self-regulation are among the most complex cognitive skills in critical thinking. Students' difficulties with items involving strategy and evaluation suggest that, although they can master the basic aspects of critical thinking, they still face challenges in the advanced dimensions that require reflective judgment and self-regulation. These findings suggest that instructional practices in Islamic Religious Education should place a greater emphasis on developing higher-order reasoning skills, not only to strengthen basic comprehension and inference, but also to foster evaluation and self-regulation through reflective tasks, case analysis, and problem-based learning.

The Wright Map results showed a distribution of respondents with the majority at medium to high ability levels, while the test items tend to be concentrated at medium difficulty levels. The Rasch-based Cognitive Diagnostic Assessment produced valid and reliable measures of students' critical thinking skills in Islamic Religious Education. The Wright Map results showed that most students clustered at the moderate ability level, with fewer students at the higher end of the scale. This finding suggests that while students can demonstrate basic clarification and inference skills, they encounter greater challenges in performing strategic and evaluative reasoning. These results suggest that instructional practices in Islamic Religious Education should extend beyond reinforcing fundamental knowledge to foster higher-order reasoning through problem-based tasks, case analysis, and reflective discussions. In addition, the results highlight a gap between the intended learning outcomes stated in the course syllabus (RPS) and the actual distribution of students' abilities, suggesting that further refinement of both curriculum design and assessment strategies is required. Thus, the study not only validates the use of Rasch-based CDA as an effective diagnostic tool but also underscores the importance of aligning assessment practices with pedagogical approaches to enhance critical thinking at an advanced level.

The distribution of students' critical thinking skills shows that 16,82% were in the low category, 52.34% were in the medium category, and 30.84% were in the high category. The

majority of students are in the medium category, meaning they have basic critical thinking skills, but have not yet fully developed inference and thinking strategies. This aligns with the trend in the item analysis results, where students find it relatively easier to answer questions that require simple explanations than questions that require logical reasoning and strategic planning. Furthermore, these results demonstrate the relevance of implementing CDA in local contexts, particularly in IRE learning. This contrasts with previous studies that relied primarily on global summative assessments such as the California Critical Thinking Skills Test (Ng et al., 2022). Compared with studies using a sectoral approach, such as Chrismawaty et al. (2022) in the context of dental education or Miller & Long (2022) in aviation education, the results of this research demonstrate that CDA can be integrated into the humanities context, particularly religious education. This broadens the scope of CDA, previously more commonly used in science or technology fields, to include religious studies, which also require critical, reflective, and analytical thinking skills. In other words, this research demonstrates that CDA is relevant not only for fields oriented toward practical problem-solving but also for courses emphasizing values, ethics, and in-depth conceptual understanding.

The results on the distribution of student abilities also enrich the literature on variations in critical thinking achievement. A study Luo et al. (2024) highlighted the growth of critical thinking in Chinese students through a longitudinal model, while this study presents evidence that the distribution of students' critical thinking abilities within a single class is highly heterogeneous. This situation emphasizes the importance of diagnostic assessments to capture the diversity of abilities, which cannot be sufficiently captured by conventional summative tests. Furthermore, this research makes a tangible, practical contribution. By utilizing the Rasch model, the instrument not only measures critical thinking skills but also provides a detailed distribution map of student abilities. This aligns with recommendations Jiang et al. (2023) that emphasize the importance of formative assessment in supporting the development of critical thinking skills through peer assessment.

These results also address the shortcomings of previous research, which tends to lack direct instructional implications. For example, Ebright-Jones & Cortina (2025) focused on the debate about the effectiveness of lecture versus seminar formats in fostering critical thinking, without producing a skills map that can be directly used by instructors. Instead, this study demonstrates how assessment results can be transformed into a basis for instructional decision-making. Thus, the contribution of this research lies not only in measurement but also in applying the measurement results to improve IRE learning strategies. By identifying that the majority of students fall into the middle category, this study also suggests the need to strengthen the IRE curriculum to further encourage the development of critical thinking. This aligns with the results of Susanti et al. (2024), who found variations in critical thinking achievement based on students' cognitive styles. These results suggest that uniform learning may not be effective for all students. This study confirms this by demonstrating that there are groups of students who require different teaching approaches based on their diagnostic results.

Several studies have employed Rasch modeling to assess critical thinking skills in various domains. In Indonesia, Prasetya & Pratama (2023) analyzed item quality in biology and found that higher-order reasoning items often showed misfit. In contrast, Karoror & Jalmo (2022) reported that many junior high school students still performed at low levels in ecosystem materials. Ridwan et al. (2023) also demonstrated that evaluation-related items in climate change contexts were among the most challenging, reinforcing the notion that strategic and evaluative reasoning persist as persistent challenges across subjects. At the international level, Nielsen et al. (2021) conducted a cross-cultural study of psychology students in Denmark and Spain, demonstrating that while the instruments achieved good overall fit, several items functioned differently across groups, highlighting the importance of testing item invariance. Likewise, Khamman & Meejang (2024) applied cognitive diagnostic assessment to reading comprehension and found that inference and evaluation subskills consistently posed the greatest difficulties. Taken together, both national and international evidence converge with the findings of this study, which identified strategy and evaluation items as the most challenging in IRE. This reinforces the contribution of Rasch-based CDA not only as a diagnostic tool in a

specific domain but also as part of the broader scholarly discussion on assessing critical thinking across contexts.

Theoretically, these results reinforce the importance of implementing cognitive diagnostic assessments in IRE learning in higher education. Rasch Model-based assessments not only demonstrate the validity and reliability of the instrument but also provide a detailed map of student abilities. This information is highly useful for lecturers in developing more adaptive learning strategies. For example, low-achieving students can be reinforced with basic concepts, medium-achieving students can be guided to improve their analytical skills, and high-achieving students can be facilitated with project-based learning or problem-solving to hone advanced critical thinking skills. The theoretical implication of this research is that Rasch Model-based CDA can bridge the gap between theory and practice in measuring critical thinking. Methodologically, this research offers an instrument that not only produces a total score but also a more meaningful map of the distribution of abilities for learning purposes. Practically, the results of this research support the design of differentiated learning strategies in Islamic Religious Education courses, which have tended to be more oriented toward understanding the material than developing higher-order thinking skills.

Thus, the implementation of Rasch Model-based cognitive diagnostic assessments has proven to be an effective approach for measuring and developing students' critical thinking skills in Islamic Religious Education courses. Furthermore, this research confirms that Rasch Model-based diagnostic assessments have significant potential for improving the effectiveness of learning in higher education. The success in differentiating students based on critical thinking ability categories demonstrates that this instrument is not only theoretically valid but also practically relevant. This positions this research as an important contribution to expanding the application of CDA to more varied contexts, while also providing an empirical basis for developing IRE curricula that are more responsive to student needs.

The Rasch-based CDA results also provide meaningful guidance for instructional practices in Islamic Religious Education. Given the variation in student ability levels, differentiated strategies are necessary to ensure that all learners are both supported and challenged. For students at the lower level, scaffolding through guided questioning and step-by-step reasoning exercises can help build a solid foundation in basic clarification and inference. Students at the moderate level can be engaged through structured debates and collaborative problem-solving tasks, which encourage them to evaluate arguments and strengthen intermediate reasoning skills. Meanwhile, students at the higher level should be challenged with problem-based projects or small-scale research tasks that demand integration of evaluative judgment and self-regulation. By aligning instructional strategies with students' diagnostic profiles, educators can more effectively foster the development of critical thinking across all levels.

While these findings provide meaningful implications for instructional practice, it is also important to acknowledge several limitations of the present study, which at the same time open avenues for future research. This study has several limitations that should be acknowledged. First, the sample size was relatively small (n = 107), which limits the generalizability of the findings and may reduce the stability of the parameter estimates. Second, the study's context was limited to Islamic Religious Education, making it difficult to extend the results to other disciplines without further validation. Third, although the Rasch-based CDA provided strong evidence for item validity and reliability, the PCA of residuals indicated a relatively high proportion of unexplained variance, suggesting the possibility of additional latent dimensions that were not captured by the current instrument.

Thus, future research should address these limitations by applying the instrument to larger and more diverse samples, as well as across different subject areas, to test the robustness and generalizability of the findings. Longitudinal studies would also be valuable for examining how students' critical thinking skills develop over time. Additionally, integrating digital assessment platforms, such as computerized adaptive testing (CAT), can enhance both efficiency and diagnostic precision, providing real-time feedback to support informed instructional decision-making.

CONCLUSION

Based on the analysis, it can be concluded that the majority of students, 56 (52,34%), fall into the moderate category. This indicates that most students possess fairly good abilities but still require reinforcement to reach the high category. Thirty-three students (30,84%) have been in the high category, indicating their good mastery of the material. Meanwhile, 18 students (16,82%) fall into the low category, requiring special attention in the learning process. Therefore, the instrument used is able to clearly differentiate student ability levels and can serve as a basis for improving learning strategies.

Suggestions for the education sector include: First, students in the low category should be provided with additional guidance, either through remedial classes, study mentoring, or the use of technology-based learning media to enhance understanding. Second, students in the moderate category should be provided with HOTS (Higher Order Thinking Skills)-based practice questions and collaborative assignments to help them develop their skills to a higher level. Third, students in the high category should be given challenges in the form of projects, small research projects, or case studies that can hone their analytical and problem-solving skills in greater depth. Furthermore, lecturers are advised to continuously evaluate learning instruments and methods to ensure they meet the diverse needs of students. Future research is recommended to expand the sample size to include various study programs and universities to ensure more representative results and allow for cross-disciplinary comparisons of students' critical thinking skills. Furthermore, integrating non-cognitive factors such as motivation, critical thinking dispositions, and sociocultural backgrounds can enrich understanding of the factors influencing student achievement. A longitudinal approach is also recommended so that the development of critical thinking skills can be dynamically monitored throughout the course, rather than just at a single point in time. Furthermore, the development of digital technology-based instruments or computerized adaptive testing will open up opportunities for more efficient, interactive, and personalized assessments. Future research could also examine the relationship between diagnostic assessment results and the effectiveness of specific learning strategies, thus going beyond measurement to assessing the most appropriate interventions to improve students' critical thinking skills.

REFERENCES

Alagumalai, S., Curtis, D., & Hungi, N. (2005). Applied Rasch Measurement: A Book of Exemplars. In *Applied Rasch Measurement: A Book of Exemplars*. https://doi.org/10.1007/1-4020-3076-2

Alnahdi, G. H. (2019). Rasch validation of the Arabic version of the Teacher Efficacy for Inclusive Practices (TEIP)scale. *Studies in Educational Evaluation*, *62*, 104–110. https://doi.org/10.1016/j.stueduc.2019.05.004

Azis, A. C. K., & Lubis, S. K. (2023). Asesmen Diagnostik Sebagai Penilaian Pembelajaran Dalam Kurikulum Merdeka Di Sekolah Dasar. *J. Pena Anda*, *1*(2), 20–29. https://doi.org/10.33830/penaanda.v1i2.6202

Barbeau, N., Frenette, E., & Hébert, M. H. (2025). Effects of perceptions of assessment on the relationship between learning strategies and academic adjustment in a summative exam preparation context. *Social Sciences and Humanities Open*, *11*, 101448. https://doi.org/10.1016/j.ssaho.2025.101448

Bond, T. (2015). Applying the Rasch Model. In *Applying the Rasch Model*. Routledge. https://doi.org/10.4324/9781315814698

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition*. Taylor & Francis. https://books.google.co.id/books?id=MRr_AQAAQBAJ

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*. Taylor & Francis. https://doi.org/10.4324/9780429030499

Brookhart, S. M. (2010). How to Assess Higher-Order Thinking Skills in Your Classroom. In *Ascd*. Ascd. http://www.ascd.org/Publications/Books/Overview/How-to-Assess-

Higher-Order-Thinking-Skills-in-YourClassroom.aspx
Chi, S., Liu, X., & Wang, Z. (2021). Comparing student science performance between hands-on and traditional item types: A many-facet Rasch analysis. *Studies in Educational Evaluation*, *70*, 100998. https://doi.org/10.1016/j.stueduc.2021.100998
Chrismawaty, B. E., Emilia, O., Rahayu, G. R., & Ana, I. D. (2022). Critical Thinking, Clinical Reasoning Skills And Cognitive Abilities Of Dental Students. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, *11*(3), 221. https://doi.org/10.22146/jpki.72915
Creswell, J. W. . (2017). *Research design : qualitative, quantitative, and mixed methods approaches*. SAGE Publications. https://books.google.co.id/books?id=4uB76IC_pOQC
Darabi Bazvand, A., & Rasooli, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation*, *72*, 101118. https://doi.org/10.1016/j.stueduc.2021.101118
Darminto, D., Handoyo, E., Wardani, S., Subali, B., & Widiarti, N. (2025). Effective Teaching Strategies to Enhance Critical Thinking Skills in Education: A Literature Review. *Edunesia : Jurnal Ilmiah Pendidikan*, *6*(2), 948–964. https://doi.org/10.51276/edu.v6i2.1227
Ebright-Jones, B. D., & Cortina, K. S. (2025). Critical Thinking in College – Can We Teach It? *Zeitschrift Für Pädagogik*, *2*, 145–164. https://doi.org/10.3262/ZP2502145
Ennis, R. H. (1985). A Logical Basis for Measuring Critical Thinking Skills. *Educational Leadership*, *43*, 44–48. https://api.semanticscholar.org/CorpusID:17938065
Ennis, R. H. (1990). The Extent to Which Critical Thinking Is Subject-Specific: Further Clarification. *Educational Researcher*, *19*(4), 13–16. https://doi.org/10.3102/0013189X019004013
Facione, P. A. (1990). Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction Executive Summary " The Delphi Report. *The California Academic Press*, *423*(c), 1–19. http://www.insightassessment.com/pdf_files/DEXadobe.PDF
Forniawan, A., & Wati, D. R. (2024). Analisis Dan Tindak Lanjut Hasil Asesmen Diagnostik Kognitif Mata Pelajaran Ilmu Pengetahuan Alam Dan Sosial Sekolah Dasar. *Al Jahiz Journal of Biology Education Research*, *4*(2). https://doi.org/10.32332/al-jahiz.v4i2.7962
Jansen, T., & Möller, J. (2022). Teacher judgments in school exams: Influences of students' lower-order-thinking skills on the assessment of students' higher-order-thinking skills. *Teaching and Teacher Education*, *111*, 103616. https://doi.org/10.1016/j.tate.2021.103616
Jiang, J., Hu, J.-Y., Zhang, Y.-B., & Yin, X.-C. (2023). Fostering college students' critical thinking skills through peer assessment in the knowledge building community. *Interactive Learning Environments*, *31*(10), 6480–6496. https://doi.org/10.1080/10494820.2022.2039949
Karoror, I., & Jalmo, T. (2022). Profile of Critical Thinking Ability in Ecosystem Materials using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1599–1604. https://doi.org/10.29303/jppipa.v8i3.1394
Khamman, W., & Meejang, S. (2024). Cognitive diagnostic assessment: Concepts for application in diagnosing reading comprehension ability. *Journal of Education and Innovation*, *26*(4), 506–522. https://so06.tci-thaijo.org/index.php/edujournal_nu/article/view/269527
Kleemola, K., Hyytinen, H., & Toom, A. (2022). Exploring internal structure of a performance-based critical thinking assessment for new students in higher education. *Assessment & Evaluation in Higher Education*, *47*(4), 556–569. https://doi.org/10.1080/02602938.2021.1946482
Laaziz, Y., Chemsi, G., & Radid, M. (2024). The Influence of E-Assessment on Students' Cognitive Engagement in Higher Education. *International Journal of Engineering Pedagogy (Ijep)*, *14*(4), 54–67. https://doi.org/10.3991/ijep.v14i4.48293
Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588
Linacre, J. M. (1994). Sample Size and Item Calibration or Person Measure Stability. *Rasch Measurement Transactions*, *7*(4), 328. http://www.rasch.org/rmt/rmt74m.htm
Lovendra, C., & Aisiah, A. (2023). Pengembangan Instrumen Asesmen Diagnostik Kognitif

Pembelajaran Sejarah. *Jurnal Kronologi*, *5*(4), 44–55. https://doi.org/10.24036/jk.v5i4.764
Lun, V. M.-C., Yeung, J. C., & Ku, K. Y. L. (2023). Effects of mood on critical thinking. *Thinking Skills and Creativity*, *47*, 101247. https://doi.org/10.1016/j.tsc.2023.101247
Luo, J., Joen, M., & Shen, H. (2024). Chinese University Students' Growth in Critical Thinking: Accounting for School Transition and Selection Effects. *Chinese/English Journal of Educational Measurement and Evaluation*, *5*(1), 1–17. https://doi.org/10.59863/JCUB3602
Miller, I., & Long, T. (2022). Development of Critical Thinking Skills in Collegiate Aviation Programs. *College Aviation Review International*, *40*(2), 165–183. https://doi.org/10.22488/okstate.22.100262
Ng, S.-Y., Cheung, K., & Cheng, H.-L. (2022). Critical Thinking Cognitive Skills and Their Associated Factors in Chinese Community College Students in Hong Kong. *Sustainability*, *14*(3), 1127. https://doi.org/10.3390/su14031127
Nielsen, T., Makransky, G., Vang, M. L., & Dammeyer, J. (2017). How specific is specific self-efficacy? A construct validity study using Rasch measurement models. *Studies in Educational Evaluation*, *53*, 87–97. https://doi.org/10.1016/j.stueduc.2017.04.003
Nielsen, T., Martínez-García, I., & Alastor, E. (2021). Critical thinking of psychology students: A within- and cross-cultural study using Rasch models. *Scandinavian Journal of Psychology*, *62*(3), 426–435. https://doi.org/10.1111/sjop.12714
Prasetya, W. A., & Pratama, A. T. (2023). Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(1), 76–91. https://doi.org/10.21831/pep.v27i1.58873
Ridwan, I. M., Kaniawati, I., Suhandi, A., Ramalis, T. R., Mahmudah, I. R., & Vilmala, B. K. (2023). Rasch modeling: Developing a critical thinking ability test for students on climate change material. *Momentum: Physics Education Journal*, *7*(2), 247–258. https://doi.org/10.21067/mpej.v7i2.8310
Song, H., & Cai, L. (2024). Interactive learning environment as a source of critical thinking skills for college students. *BMC Medical Education*, *24*(1), 270. https://doi.org/10.1186/s12909-024-05247-y
Stone, M. H., & Wright, B. D. (1999). Measurement Essentials. In *Measurement*. Wilmington, DE: Wide Range.
Sumintono, B., & Widhiarso, W. (2014). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi). In *Trim Komunikata Publishing House.*
Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan Instrumen Asesmen Diagnostik Untuk Persiapan Kurikulum Merdeka. *Journal of Community Empowerment*, *2*(2), 67–73. https://doi.org/10.15294/jce.v2i2.61886
Susanti, W., Nasution, T., Tendra, G., Simeru, A., & Yuliendi, R. R. (2024). Assessing Students' Critical Thinking Skills in Terms of Cognitive Style: A Study of the Application of the Inquiry Collaborative Based V-Lab Model in Programming Courses. *Journal of Applied Business and Technology*, *5*(2), 67–71. https://doi.org/10.35145/jabt.v5i2.165
Thi Nhat, H., Thi Lien, N., Thi Tinh, N., Vu Thu Hang, N., & Thu Trang, N. (2018). The Development of Critical Thinking for Students in Vietnamese Schools: From Policies to Practices. *American Journal of Educational Research*, *6*(5), 431–435. https://doi.org/10.12691/education-6-5-10
Une, D., Bahsoan, A., Katili, L. D., Kasim, N. M., Tamu, Y., Dude, S., & Pomalingo, S. (2015). *Pendidikan Agama Islam di Perguruan Tinggi*. Ideas Publishing.
Vaughan, B. (2018). A Rasch analysis of the Revised Study Process Questionnaire in an Australian osteopathy student cohort. *Studies in Educational Evaluation*, *56*, 144–153. https://doi.org/10.1016/j.stueduc.2017.12.003
Zhou, W., Feng, H., Tao, H. J., Sun, H., Zhang, T., Wang, Q., & Zhang, L. (2024). Factors Influencing Poststroke Cognitive Dysfunction: Cross-Sectional Analysis. *Jmir Formative Research*, *8*, e59572. https://doi.org/10.2196/59572