Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

April 2026. Vol. 14, No. 2
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 919-937

Profiling Critical Thinking Disposition and Preliminary Psychometric Evidence among Early-Year Science Education Students

[1,2]*Ni Nyoman Sri Putu Verawati, [2]Wahyudi, [2]Nina Nisrina

[1]Science Education Department, University of Mataram, Mataram, Indonesia
[2]Physics Education Department, University of Mataram, Mataram, Indonesia

*Corresponding Author e-mail: veyra@unram.ac.id

Received: February 2026; Revised: March 2026; Published: April 2026

Abstract

This study profiled the critical thinking disposition of early-year science education students and examined preliminary psychometric evidence of an Ennis-based questionnaire. A quantitative descriptive survey with preliminary psychometric analysis was conducted with 90 second-semester science education students from three universities in West Nusa Tenggara Province, Indonesia. Data were collected using a 36-item questionnaire organized into 12 indicators of critical thinking disposition. Negatively worded items were reverse-scored before analysis. Descriptive statistics, category distribution, indicator-based profiling, Cronbach's alpha, corrected item-total correlation, and exploratory factor analysis based on 12 indicator scores were applied. The results showed that students' critical thinking disposition was generally moderate, with nearly all participants located in the moderate category. The highest indicator scores were found in precision required by situation, total situation, and employing critical thinking abilities, while the lowest scores were found in clear reasons, credible sources and observations, and being well-informed. The corrected internal consistency was modest, and several negatively worded items showed weak item-total correlations. The exploratory factor analysis suggested a tentative three-factor structure: evidence-based reasoning disposition, contextual and reflective judgment, and self-regulated critical engagement. These findings indicate that students' critical thinking disposition is still developing and that the questionnaire requires further refinement.

Keywords: Critical thinking disposition; Early-year students; Science education; Psychometric evidence; Exploratory factor analysis

How to Cite: Verawati, N. N. S. P., Wahyudi, W., & Nisrina, N. (2026). Profiling Critical Thinking Disposition and Preliminary Psychometric Evidence among Early-Year Science Education Students. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *14*(2), 919–937. https://doi.org/10.33394/j-ps.v14i2.20786

https://doi.org/10.33394/j-ps.v14i2.20786

INTRODUCTION

Critical thinking has long been recognized as one of the central aims of higher education, particularly in science education where students are expected to evaluate evidence, construct reasoned explanations, interpret data, and make justified decisions. In science learning, critical thinking is not only related to the ability to solve problems correctly, but also to the tendency to ask clear questions, seek reliable information, examine assumptions, consider alternative explanations, and revise judgments when stronger evidence is available. This makes critical thinking especially relevant for science education students, who are not only learners of scientific concepts but also prospective educators who will later be responsible for cultivating scientific reasoning in their own classrooms. Meta-analytic and assessment-oriented studies have shown that critical thinking can be developed through explicit instructional interventions and systematic assessment, although its development requires attention to both cognitive and dispositional aspects (Abrami et al., 2008; Liu et al., 2014).

However, the development of critical thinking cannot be understood only as a matter of cognitive skill. A student may know how to analyze information but may not be inclined to do so consistently. This distinction is important because critical thinking involves both ability and disposition. Ability refers to the cognitive capacity to analyze, infer, evaluate, and explain, whereas disposition refers to the willingness or tendency to use such abilities when facing problems, claims, or uncertain situations. In this sense, critical thinking disposition reflects students' intellectual habits, such as being open-minded, careful in judgment, willing to seek reasons, attentive to credible sources, and ready to change position when evidence is sufficient. Previous studies have emphasized that critical thinking disposition is conceptually distinct from critical thinking skills and therefore requires separate measurement and interpretation (Anders et al., 2019; Chen & Fan, 2022; İskifoğlu, 2013).

For early-year university students, critical thinking disposition is particularly important because the first phase of higher education is a transitional period from school-based learning to more independent academic inquiry. Students at this level are often still developing the habits required for university-level reasoning, including academic reading, evidence-based argumentation, source evaluation, and reflective judgment. In science education programs, these habits are foundational because scientific knowledge is built through questioning, observation, evidence interpretation, and reasoned explanation. If students have weak dispositions toward seeking clear reasons, using credible sources, or being well-informed, they may struggle to engage deeply with scientific inquiry even when they have access to relevant content knowledge. Therefore, profiling critical thinking disposition among early-year science education students can provide useful information for curriculum planning and instructional intervention. This is consistent with broader research indicating that critical thinking dispositions may be influenced by educational context, disciplinary experience, and learning opportunities provided in higher education programs (Abrami et al., 2008; Anders et al., 2020; Liu et al., 2014).

The conceptual foundation of this study is based on Ennis' framework of critical thinking disposition. In this framework, the ideal critical thinker is characterized by several intellectual tendencies, including seeking clear statements of questions or theses, offering clear reasons, trying to be well-informed, using credible sources, considering the total situation, keeping the basic concern in mind, being alert to alternatives, being open-minded, changing position when evidence and reasons are sufficient, seeking appropriate precision, trying to get things right, and employing critical thinking abilities (Ennis, 2015). The instrument used in this study was developed from these 12 characteristics, with each characteristic represented by three questionnaire items. The questionnaire therefore provides an indicator-based structure for examining not only the overall level of critical thinking disposition but also the relative strengths and weaknesses of specific dispositional dimensions.

Despite the growing interest in critical thinking in higher education, several issues remain underexplored. First, many studies focus more on critical thinking skills than on critical thinking disposition. Skill-based assessments are valuable, but they may not fully explain whether students are inclined to use critical thinking in academic and everyday situations. Second, when disposition is measured, the analysis often stops at total scores or general categories, while less attention is given to indicator-level profiles. Such profiles are important because students may not show the same level of disposition across all dimensions. For instance, they may be relatively open to alternatives but less consistent in using credible sources or providing clear reasons. Research on critical thinking disposition instruments, including the California Critical Thinking Disposition Inventory and its adaptations, shows that disposition is commonly treated as a multidimensional construct involving components such as truth-seeking, open-mindedness, analyticity, systematicity, self-confidence, curiosity, and maturity of judgment (Anders et al., 2019; Chen & Fan, 2022; İskifoğlu, 2013; Wang et al., 2019).

A third issue concerns the psychometric quality of the instruments used to measure critical thinking disposition. A questionnaire should not only be theoretically relevant but should also show acceptable empirical performance. Reliability, item-total correlation, and factor structure are important because they indicate whether the items function consistently and whether the proposed dimensions are supported by the data. In educational research, psychometric evaluation is especially important because instruments are often adapted to particular populations, languages, cultural contexts, or disciplinary needs. Previous studies have shown that critical thinking disposition instruments may demonstrate acceptable reliability in some populations but may not always replicate the same factorial structure across different cultural or educational contexts (Cui et al., 2020; İskifoğlu, 2013; Wang et al., 2019). This suggests that each use of a disposition instrument should be accompanied by evidence of reliability and validity that is appropriate to the target population.

The use of negatively worded items also deserves attention in disposition measurement. Such items are often included to reduce acquiescence bias, but they can also create interpretation problems, especially for early-year students. If respondents misunderstand negatively worded statements or respond inconsistently, the internal consistency of the instrument may decrease. Therefore, examining item performance is not a minor technical step but an important part of interpreting the quality of a critical thinking disposition questionnaire. This concern is aligned with broader psychometric discussions showing that instrument performance can vary across populations and that translated or adapted measures may require careful evaluation before strong conclusions are drawn (Anders et al., 2019; Bakhtiari-Dovvombaygi et al., 2024; Cui et al., 2020). In this study, the questionnaire contained eight negatively worded items that required reverse scoring before total score, indicator score, reliability, and exploratory factor analysis were computed.

Another gap concerns the factor structure of critical thinking disposition instruments. Although theoretical frameworks may propose several characteristics or dimensions, empirical data do not always reproduce the same structure. Exploratory factor analysis can be useful in this situation because it provides preliminary evidence about how indicators group together in a specific sample. However, such analysis must be interpreted carefully, especially when sample size is limited. Previous research on critical thinking disposition has reported that factorial structures may vary across professions, educational levels, and cultural contexts, which makes context-specific evidence necessary before an instrument is used for broader interpretation (Anders et al., 2019; İskifoğlu, 2013; Wang et al., 2019). For the present study, EFA was conducted at the indicator level rather than the item level because the available data consisted of 90 respondents and 12 indicator scores. This approach does not provide final construct validation, but it can offer preliminary psychometric evidence regarding the possible structure of the instrument.

Based on these considerations, the present study aims to profile the critical thinking disposition of early-year science education students and examine preliminary psychometric evidence of an Ennis-based questionnaire. Specifically, the study addresses four objectives: to describe the overall level of students' critical thinking disposition, to identify the relative strengths and weaknesses across 12 disposition indicators, to examine the internal consistency and item performance of the questionnaire, and to explore the preliminary factor structure of the 12 indicators. By combining descriptive profiling and initial psychometric analysis, this study is expected to contribute both substantively and methodologically. Substantively, it provides evidence about how early-year science education students tend to engage with critical thinking dispositions. Methodologically, it offers preliminary evidence about the performance of an Ennis-based disposition instrument that may require further refinement and validation in broader samples.

METHOD

Research Design

This study employed a quantitative descriptive survey design with preliminary psychometric analysis. The descriptive component was used to profile the critical thinking disposition of early-year science education students, while the psychometric component was used to examine the initial empirical performance of the questionnaire. This design was appropriate because the study did not aim to test an intervention, examine causal relationships, or compare different groups. Instead, it focused on describing students' disposition patterns and evaluating the internal consistency, item performance, and exploratory factor structure of the instrument.

The study was conducted as a cross-sectional investigation involving second-semester science education students from three universities in West Nusa Tenggara Province, Indonesia. The analysis focused on four main aspects: students' overall critical thinking disposition level, the indicator-based profile of critical thinking disposition, the internal consistency and item performance of the questionnaire, and the preliminary factor structure of the 12 indicators. Therefore, the findings are interpreted as descriptive and exploratory rather than causal or confirmatory.

Participants

The participants were 90 second-semester students enrolled in science education programs at three universities in West Nusa Tenggara Province, Indonesia. The universities were Universitas Mataram, Universitas Islam Negeri Mataram, and Universitas Pendidikan Mandalika. These participants were selected because early-year science education students are in an important transition phase in higher education, where academic reasoning, evidence-based thinking, reflective judgment, and scientific inquiry habits are still developing. As prospective science educators, their critical thinking disposition is relevant not only to their own academic development but also to their future professional role in facilitating scientific reasoning in classrooms.

All participants completed the critical thinking disposition questionnaire. The dataset contained complete responses from all 90 students across 36 questionnaire items. No missing responses or invalid scores were identified during the data screening process. Since the available dataset did not include demographic variables such as gender, age, academic achievement, or university-specific participant distribution, the analysis was limited to students' responses to the critical thinking disposition questionnaire. The response data were organized by respondent codes from M1 to M90 and item scores from item 1 to item 36.

Instrument

The instrument used in this study was a critical thinking disposition questionnaire developed based on Ennis' framework of the ideal critical thinker. The questionnaire consisted of 36 items organized into 12 indicators, with three items representing each indicator. The indicators included clarity of question or thesis, clear reasons, being well-informed, credible sources and observations, total situation, basic concern in context, alternatives, open-mindedness, change position based on evidence, precision required by situation, get it right, and employ critical thinking abilities.

The questionnaire used a five-point response scale ranging from 1 to 5, where higher scores indicated stronger critical thinking disposition. Of the 36 items, 28 were positively worded and 8 were negatively worded. The negatively worded items were items 15, 18, 21, 24, 27, 30, 33, and 36. These items were reverse-scored before analysis so that all items had the same interpretive direction, where higher scores consistently represented stronger critical thinking disposition.

The 12 indicators and their item distribution were as follows: clarity of question or thesis was represented by items 1 to 3; clear reasons by items 4 to 6; being well-informed by items 7 to 9; credible sources and observations by items 10 to 12; total situation by items 13 to 15; basic concern in context by items 16 to 18; alternatives by items 19 to 21; open-mindedness by items 22 to 24; change position based on evidence by items 25 to 27; precision required by situation by items 28 to 30; get it right by items 31 to 33; and employ critical thinking abilities by items 34 to 36.

Data Collection Procedure

Data were collected by administering the questionnaire to second-semester science education students from the three participating universities. Students were asked to respond to each statement according to their actual condition by selecting one response option on the five-point scale. The questionnaire was designed to capture students' self-reported tendencies in engaging with critical thinking, such as seeking reasons, evaluating sources, considering alternatives, withholding judgment when evidence is insufficient, and making judgments based on evidence.

After the responses were collected, the data were tabulated into a respondent-by-item matrix. Each respondent was coded from M1 to M90, and each item was coded from item 1 to item 36. The tabulated dataset was then checked for completeness, consistency of score range, and readiness for statistical analysis. The final dataset consisted of 90 complete cases and 36 item responses for each case.

Scoring and Data Analysis

The scoring procedure began with reverse scoring of the negatively worded items. For items 15, 18, 21, 24, 27, 30, 33, and 36, the original score was transformed using the rule 1 = 5, 2 = 4, 3 = 3, 4 = 2, and 5 = 1. This step ensured that all items were scored in the same direction. After reverse scoring, the corrected total score was calculated by summing the scores of all 36 items. The theoretical minimum score was 36, and the theoretical maximum score was 180. The mean item score was also computed by dividing the total score by 36.

Indicator scores were calculated by averaging the three items belonging to each of the 12 indicators. These indicator scores were used to identify students' relative strengths and weaknesses across the dimensions of critical thinking disposition. To support descriptive interpretation, students' mean item scores were classified into three levels: low, moderate, and high. The low category covered mean scores from 1.00 to 2.33, the moderate category from 2.34 to 3.67, and the high category from 3.68 to 5.00. This classification was used only as a descriptive aid and was not treated as a diagnostic classification of individual competence.

Data analysis was conducted in several stages. First, data screening was performed to identify missing values, invalid scores, and inconsistencies in the response range. Second, descriptive statistics were calculated for the corrected total score and mean item score, including minimum score, maximum score, mean, standard deviation, median, and confidence interval. Third, indicator-based descriptive analysis was conducted by calculating the mean, standard deviation, minimum, and maximum score for each of the 12 indicators. The indicators were then ranked from the highest to the lowest mean score.

Internal consistency was examined using Cronbach's alpha. Reliability was calculated for all 36 corrected items, positive items only, negatively worded items after reverse scoring, and all items without reverse scoring for diagnostic comparison. The main reliability result was based on the corrected 36-item scale because it followed the intended scoring direction of the instrument. Item performance was examined using corrected item-total correlations and alpha-if-item-deleted values. Items with weak, near-zero, or negative item-total correlations were identified as items requiring further review. However, item deletion was not conducted automatically because theoretical relevance and content validity also need to be considered before removing items from an instrument.

A preliminary exploratory factor analysis was conducted using the 12 indicator scores rather than the 36 individual item scores. This approach was selected because the available sample consisted of 90 respondents, making indicator-level EFA more defensible than item-level EFA. Before factor extraction, the suitability of the data was examined using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. Factors were extracted using Principal Axis Factoring and rotated using Varimax rotation. The interpretation of the factor structure was based on rotated factor loadings, communalities, eigenvalues, explained variance, and conceptual coherence among indicators.

The EFA results were interpreted as preliminary psychometric evidence, not as final construct validation. This caution was necessary because the sample size was limited and the KMO value indicated only marginal sampling adequacy. Therefore, the factor structure was used to provide an initial indication of how the 12 indicators might cluster in this sample, while recognizing that further testing with a larger and more diverse sample is needed.

Ethical Considerations

The study used questionnaire data collected from students for academic research purposes. The responses were analyzed in aggregate form, and individual respondents were coded anonymously using respondent labels. No personal identifying information was reported in the analysis or presentation of findings. The interpretation of results was also conducted carefully to avoid labeling individual students, since the purpose of the study was to describe group-level disposition patterns and examine preliminary instrument performance.

RESULTS AND DISCUSSION

Results

Data Screening

The dataset consisted of responses from 90 second-semester science education students who completed a 36-item critical thinking disposition questionnaire. The questionnaire was structured into 12 indicators, with each indicator represented by three items. Before conducting the statistical analysis, the data were screened for completeness, invalid scores, and scoring direction. The summary of the data screening process is presented in Table 1.

Table 1. Data screening and scoring procedure

Aspect	Result
Number of respondents	90
Study population	Second-semester science education students
Number of questionnaire items	36
Number of indicators	12
Items per indicator	3
Response scale	1 to 5
Missing data	0
Scores outside the expected range	0
Negatively worded items	15, 18, 21, 24, 27, 30, 33, 36
Scoring treatment for negative items	Reverse scoring

The results in Table 1 show that the dataset was complete and suitable for further analysis. All 90 students responded to the 36 questionnaire items, and no missing values were identified. In addition, all responses were within the expected score range of 1 to 5. This indicates that the dataset did not require case deletion, data imputation, or correction of invalid responses before analysis. The completeness of the dataset is important because the subsequent analyses, including total score calculation, indicator profiling, reliability analysis, and exploratory factor analysis, depend on consistent responses across all items.

Table 1 also shows that eight items were negatively worded and required reverse scoring before statistical analysis. These items were items 15, 18, 21, 24, 27, 30, 33, and 36. Reverse scoring was applied to ensure that all items had the same interpretive direction, where higher scores consistently reflected stronger critical thinking disposition. Without this correction, the total scores and reliability estimates would be misleading because negatively worded items would contribute scores in the opposite direction of the intended construct.

General Disposition

After reverse scoring the negatively worded items, the corrected total score was computed by summing all 36 item scores. Descriptive statistics were then calculated to describe the overall level of students' critical thinking disposition. The descriptive statistics of the corrected total scores are presented in Table 2.

Table 2. Descriptive statistics of corrected critical thinking disposition scores

Statistic	Value
N	90
Theoretical minimum score	36
Theoretical maximum score	180
Empirical minimum score	79
Empirical maximum score	114
Mean total score	98.44
Standard deviation	5.45
Median	98.00
Mean per item	2.73
95% CI for mean per item	2.70 to 2.77

The results in Table 2 indicate that the corrected total scores ranged from 79 to 114, while the theoretical score range of the instrument was 36 to 180. The mean total score was 98.44, with a standard deviation of 5.45. The median score was 98.00, which was very close to the mean score. This suggests that the distribution of total scores was relatively centered and did not show a large difference between the average and middle values. When converted into a mean item score, students obtained an average of 2.73 on a five-point scale.

Table 2 suggests that the students' critical thinking disposition was generally at a moderate level. The mean item score of 2.73 was above the low category but still far from the high category. Therefore, the result should not be interpreted as evidence that students already had a strong critical thinking disposition. A more accurate interpretation is that the students showed an emerging disposition toward critical thinking, but this tendency had not yet developed consistently. This finding is relevant because the respondents were second-semester students who were still in the early stage of developing academic reasoning habits in science education.

Disposition Category

To provide a clearer interpretation of the score distribution, students were classified into three disposition levels based on their mean item scores. The classification used three categories: low, moderate, and high. The distribution of students across these critical thinking disposition levels is presented in Table 3.

Table 3. Distribution of students across critical thinking disposition levels

Category	Score range	Frequency	Percentage
Low	1.00 to 2.33	1	1.11%
Moderate	2.34 to 3.67	89	98.89%
High	3.68 to 5.00	0	0.00%

The results in Table 3 show that almost all students were categorized as having a moderate level of critical thinking disposition. Specifically, 89 students, or 98.89%, were in the moderate category, while only one student, or 1.11%, was in the low category. No student reached the high category. This distribution confirms that the students' disposition was not concentrated at the lowest level, but it also shows that the group had not yet demonstrated a high level of critical thinking disposition.

Table 3 also indicates that the critical thinking disposition profile of the sample was relatively homogeneous. Most students were located within the same category, suggesting that the overall pattern of disposition among second-semester science education students was quite similar. This may reflect their comparable academic experience, since they were at the same semester level and likely had not yet received extensive disciplinary training in scientific argumentation, evidence evaluation, or inquiry-based reasoning. Therefore, the moderate category should be interpreted as a developmental condition rather than as a final level of competence.

Indicator Profile

The 36 questionnaire items were grouped into 12 indicators based on the critical thinking disposition framework. Each indicator consisted of three items, and the indicator score was calculated by averaging the corrected scores of the relevant items. The indicator-based profile of students' critical thinking disposition is presented in Table 4.

Table 4. Indicator-based profile of students' critical thinking disposition

Rank	Indicator	Items	Mean	SD	Min	Max
1	Precision required by situation	28 to 30	2.92	0.44	2.33	4.00
2	Total situation	13 to 15	2.89	0.34	2.33	3.67
3	Employ critical thinking abilities	34 to 36	2.89	0.39	2.00	4.00
4	Get it right	31 to 33	2.86	0.37	2.00	3.67
5	Alternatives	19 to 21	2.85	0.35	1.67	3.67
6	Open-mindedness	22 to 24	2.84	0.30	2.00	3.33
7	Basic concern in context	16 to 18	2.83	0.37	2.00	3.67
8	Change position based on evidence	25 to 27	2.75	0.37	1.67	3.67
9	Clarity of question/thesis	1 to 3	2.64	0.42	1.67	3.67
10	Being well-informed	7 to 9	2.46	0.39	1.67	3.67
11	Credible sources and observations	10 to 12	2.45	0.37	1.00	3.00
12	Clear reasons	4 to 6	2.44	0.45	1.67	3.67

The results in Table 4 show that the highest indicator score was found in precision required by situation, with a mean of 2.92. This was followed by total situation and employ critical thinking abilities, each with a mean of 2.89. These indicators represent students' tendency to adjust the level of precision according to situational demands, consider broader contextual factors, and use critical thinking abilities when analyzing problems. Although these were the highest indicators, their mean scores still remained within the moderate range. This means that these indicators were relatively stronger than the others, but they were not yet strong in an absolute sense.

Table 4 also shows that the lowest mean scores were found in clear reasons, credible sources and observations, and being well-informed. These three indicators are closely related to evidence-based reasoning. The low scores suggest that students were less consistent in providing clear reasons, using credible sources, and seeking sufficient information before drawing conclusions. This pattern is substantively important for science education because scientific thinking requires the ability to justify claims, evaluate evidence, and refer to reliable sources. The finding suggests that students may already show general awareness of critical thinking, but their academic reasoning habits still need to be strengthened.

Internal Consistency

The internal consistency of the questionnaire was examined using Cronbach's alpha. Reliability was calculated for all corrected items, positive items only, negatively worded items after reverse scoring, and all items without reverse scoring for comparison. The results of the reliability analysis are presented in Table 5.

Table 5. Internal consistency of the critical thinking disposition questionnaire

Reliability analysis	Cronbach's Alpha
All 36 items after reverse scoring	0.522
Positive items only	0.651
Negatively worded items after reverse scoring	0.558
All 36 items without reverse scoring	0.682

The results in Table 5 show that the Cronbach's alpha coefficient for all 36 items after reverse scoring was 0.522. This value indicates modest internal consistency and suggests that the questionnaire did not function as a highly consistent scale in this sample. The result is important because the corrected alpha is the most appropriate reliability coefficient to report, since the instrument includes negatively worded items that must be scored in the opposite direction. Therefore, the instrument should not be described as having strong reliability based on the present dataset.

Table 5 also shows that the alpha coefficient for positive items only was 0.651, which was higher than the coefficient for the full corrected scale. Meanwhile, the alpha coefficient for all items without reverse scoring was 0.682. However, this latter value should not be used as the main reliability result because it does not follow the scoring logic of the instrument. The comparison suggests that the negatively worded items may have contributed to the lower reliability after reverse scoring. This finding indicates the need for further item review, especially to determine whether the wording of negative items was clear enough for early-semester students.

Item Performance

To identify items that may weaken the internal consistency of the instrument, corrected item-total correlations and alpha-if-item-deleted values were examined. Table 6 presents the items with the weakest corrected item-total correlations.

Table 6. Items with the weakest corrected item-total correlations

Item	Direction	Corrected mean	SD	Corrected item-total r	Alpha if item deleted
15	Negative	3.71	0.59	-0.156	0.546
35	Positive	2.78	0.73	-0.122	0.549
18	Negative	3.53	0.71	-0.052	0.539
24	Negative	3.51	0.60	-0.031	0.533
1	Positive	2.83	0.64	-0.019	0.532
32	Positive	2.54	0.66	-0.003	0.531
27	Negative	3.53	0.67	0.013	0.529
36	Negative	3.28	0.86	0.024	0.533
30	Negative	3.60	0.73	0.035	0.528
34	Positive	2.60	0.73	0.038	0.527
26	Positive	2.43	0.64	0.044	0.525
22	Positive	2.52	0.55	0.068	0.521

The results in Table 6 show that several items had weak corrected item-total correlations. Item 15 showed the weakest value, with a corrected item-total correlation of -0.156. Other items with negative item-total correlations included item 35, item 18, item 24, item 1, and item 32. Negative or near-zero item-total correlations indicate that these items did not move consistently with the total corrected score. In other words, students' responses to these items were not strongly aligned with their general critical thinking disposition score.

Table 6 also shows that several weak items were negatively worded, including items 15, 18, 24, 27, 30, and 36. This pattern suggests that negative wording may have caused response inconsistency or interpretation difficulty among students. However, item 35 was also problematic even though it was positively worded, which means the issue was not limited only to reverse-scored items. These findings should not be used as an immediate basis for deleting items without further review. Instead, they indicate that the items require qualitative evaluation, expert judgment, and further testing with a larger sample before final decisions are made.

EFA Suitability

A preliminary exploratory factor analysis was conducted using the 12 indicator scores rather than the 36 individual items. This decision was made because the sample size of 90 respondents was more defensible for an indicator-level exploratory analysis than for an item-level factor analysis. The suitability tests and initial eigenvalues are presented in Table 7.

Table 7. EFA suitability tests and initial eigenvalues

Test or component	Result
KMO measure of sampling adequacy	0.547
Bartlett's test of sphericity	χ^2 = 105.20
Degrees of freedom	66
p-value	0.0015
Component 1 eigenvalue	2.098
Component 2 eigenvalue	1.485
Component 3 eigenvalue	1.457
Component 4 eigenvalue	1.149
Component 5 eigenvalue	1.051
Component 6 eigenvalue	0.920

The results in Table 7 show that Bartlett's test of sphericity was significant, $\chi^2(66)$ = 105.20, p = 0.0015. This indicates that the correlation matrix among the 12 indicators was not an identity matrix and could be explored further using factor analysis. However, the KMO value was 0.547, which reflects only marginal sampling adequacy. This means that the EFA result should be interpreted carefully. It can provide an initial indication of the possible structure among indicators, but it is not strong enough to serve as final evidence of construct validity.

Table 7 also shows that five components had eigenvalues greater than 1.00. Based only on the Kaiser criterion, this could suggest a five-component solution. However, such a solution would be too fragmented because the analysis involved only 12 indicators and 90 respondents. A five-factor solution would also be difficult to interpret theoretically. Therefore, a three-factor solution was considered more reasonable because it provided a more coherent and interpretable structure. This decision was based not only on eigenvalues, but also on conceptual interpretability and the exploratory purpose of the analysis.

Explained Variance

To complement the eigenvalue information, the explained variance of each extracted component and the cumulative explained variance were visualized. Figure 1 presents two types of information simultaneously: the bars indicate the individual explained variance contributed by each component, while the line chart shows the cumulative explained variance across the 12 components. This visualization helps clarify not only how much variance was explained by each component, but also how the contribution of additional components changed across the factor solution.

The results in Figure 1 show that the first component explained the largest proportion of variance, accounting for 17.48% of the total variance. The second and third components explained 12.38% and 12.14%, respectively. Together, the first three components explained 42.00% of the total variance. The fourth and fifth components added 9.58% and 8.76%, increasing the cumulative explained variance to 51.58% and 60.33%, respectively. This pattern is consistent with the eigenvalue results in Table 7, where the first five components had eigenvalues greater than 1.00.

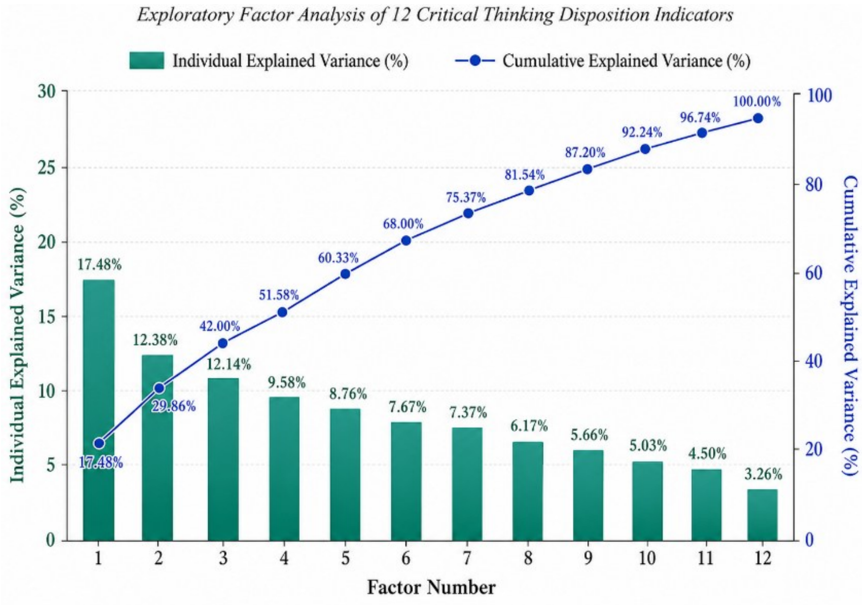

Figure 1. Explained variance by factor

However, Figure 1 also shows that the variance contribution gradually decreased after the first component. Although the Kaiser criterion could suggest retaining five components, the additional components after the third factor contributed increasingly smaller portions of variance and would have produced a more fragmented solution. Retaining five factors from only 12 indicators would also be less parsimonious and more difficult to interpret conceptually. For this reason, the three-factor solution was retained because it provided a more coherent and interpretable structure for preliminary analysis.

The first component did not explain a dominant proportion of variance, and the first three components together explained only 42.00% of the total variance. This indicates that the extracted three-factor structure captured an important part of the indicator variation but did not account for all meaningful variance in the data. Therefore, Figure 1 supports the use of a three-factor solution as a preliminary exploratory model, not as definitive evidence of a stable construct structure. Further analysis with a larger sample is still needed to test whether this factor structure can be replicated.

Factor Structure

The three-factor solution was extracted using Principal Axis Factoring and rotated using Varimax rotation. The rotated factor loadings and communalities are presented in Table 8. To make the pattern of factor loadings easier to interpret, the preliminary three-factor EFA structure is also visualized in Figure 2.

Table 8. Rotated factor loadings of the 12 indicators

Indicator	Factor 1	Factor 2	Factor 3	Communality
Clarity of question/thesis	0.226	0.023	0.052	0.054
Clear reasons	0.968	0.013	0.060	0.941
Being well-informed	0.462	0.165	-0.204	0.282
Credible sources and observations	0.283	0.340	-0.126	0.212
Total situation	0.036	0.402	-0.321	0.267
Basic concern in context	-0.007	0.525	0.027	0.276
Alternatives	0.171	0.480	-0.086	0.267
Open-mindedness	0.023	0.346	0.152	0.143
Change position based on evidence	0.146	0.087	0.098	0.039
Precision required by situation	-0.060	0.348	0.220	0.173
Get it right	0.138	-0.021	0.661	0.457
Employ critical thinking abilities	-0.010	0.060	0.325	0.109

The results in Table 8 suggest a tentative three-factor structure. Factor 1 was mainly represented by clear reasons, being well-informed, and credible sources and observations. This factor can be interpreted as evidence-based reasoning disposition because its dominant indicators relate to the tendency to justify claims, seek sufficient information, and use credible sources. Factor 2 was mainly represented by basic concern in context, alternatives, total situation, open-mindedness, and precision required by situation. This factor can be interpreted as contextual and reflective judgment because it reflects the tendency to consider context, alternatives, and situational demands before making judgments.

Table 8 also shows that Factor 3 was mainly represented by get it right and employ critical thinking abilities. This factor can be interpreted as self-regulated critical engagement because it reflects students' tendency to use their thinking abilities and strive for appropriate judgment. However, several indicators had low communalities, especially change position based on evidence, clarity of question/thesis, and employ critical thinking abilities. These low communalities indicate that some indicators were not well explained by the extracted factor structure. Therefore, the three-factor structure should be treated as a preliminary result that needs to be tested again using a larger and more diverse sample.

Preliminary EFA Model

The three-factor structure was visualized to clarify the relationship between the 12 indicators and their dominant factors. In this figure, solid lines represent dominant loadings of 0.30 or higher, while dashed lines represent weak loadings below 0.30. The preliminary three-factor EFA structure is presented in Figure 2.

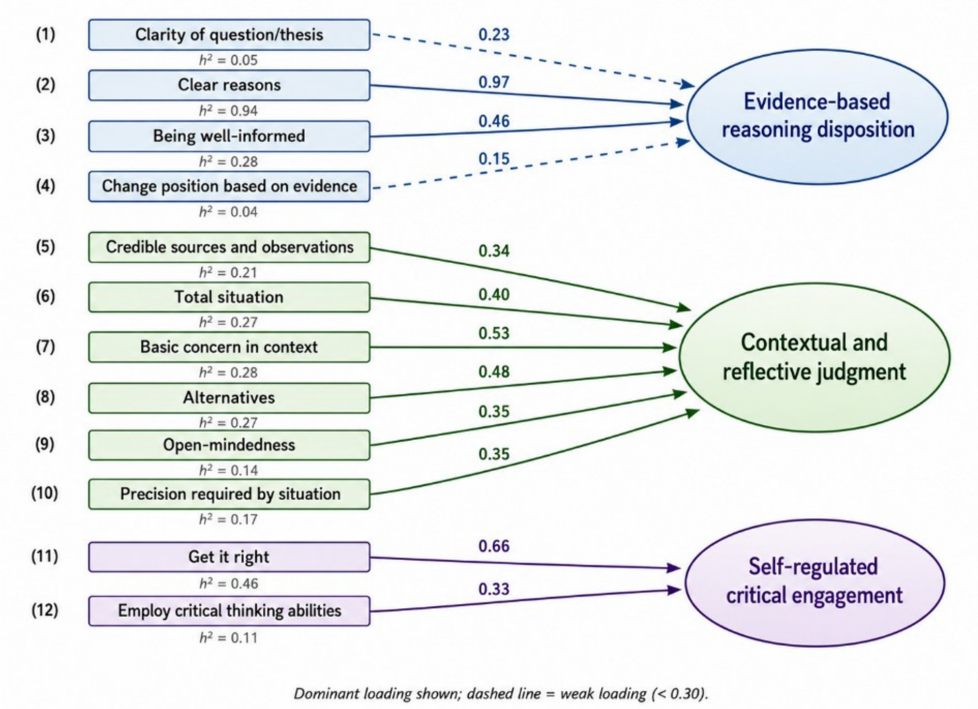

Figure 2. Preliminary three-factor EFA structure

The results in Figure 2 show that the first factor was represented most strongly by clear reasons, with a loading of 0.97. Being well-informed also loaded on this factor with a loading of 0.46. However, clarity of question/thesis and change position based on evidence showed weak loadings of 0.23 and 0.15, respectively. This means that although these indicators were visually associated with the first factor for interpretive purposes, their statistical contribution was weak. Therefore, Factor 1 should be interpreted primarily through clear reasons and being well-informed, with additional caution regarding the weaker indicators.

Figure 2 also shows that the second factor had a more distributed loading pattern, with basic concern in context, alternatives, total situation, open-mindedness, precision required by situation, and credible sources and observations contributing to the factor. The third factor was represented by get it right and employ critical thinking abilities. The diagram makes the preliminary structure easier to interpret, but it also highlights the uneven strength of the indicators. Some indicators showed strong loadings, while others had low communalities or weak loading values. Thus, the figure supports the interpretation that the EFA structure is useful as an exploratory model but should not be treated as a confirmed measurement model.

Discussion

This study aimed to profile the critical thinking disposition of early-year science education students and examine preliminary psychometric evidence of an Ennis-based critical thinking disposition questionnaire. The findings show that the students' critical thinking disposition was generally at a moderate level. This pattern was evident from the corrected mean item score and the distribution of students across categories, where almost all participants were located in the moderate category and none reached the high category. This finding suggests that second-semester science education students had begun to demonstrate an orientation toward critical thinking, but this tendency had not yet developed into a strong and stable intellectual disposition. Therefore, the result should not be interpreted as evidence of a high level of critical thinking disposition. A more defensible interpretation is that students were in a developmental phase in which critical thinking dispositions were emerging but still required systematic academic support.

The moderate level of critical thinking disposition is understandable considering the characteristics of the participants. The respondents were early-year university students who were still transitioning from school-based learning to higher education learning cultures. At this stage, students are generally still adapting to academic reading, evidence-based discussion, independent inquiry, scientific explanation, and reflective judgment. This developmental condition is important because critical thinking disposition does not emerge automatically from exposure to science content. It needs to be cultivated through learning environments that require students to ask questions, justify claims, examine sources, evaluate evidence, and revise conclusions when stronger reasons are available. This interpretation is consistent with the broader view that critical thinking development involves both cognitive skills and dispositional tendencies, and that educational interventions need to address both aspects rather than treating critical thinking only as a cognitive performance (Abrami et al., 2008; Liu et al., 2014).

The indicator-level findings provide a more specific explanation of the students' disposition profile. The relatively higher indicators were precision required by situation, total situation, and employ critical thinking abilities. These indicators suggest that students tended to show some awareness of the need to adjust the precision of their thinking, consider broader situational factors, and use critical thinking abilities when analyzing problems. However, these indicators still remained within the moderate range. This means that they were relatively stronger than other indicators, but they cannot be interpreted as strong dispositions in an absolute sense. This pattern suggests that students may already possess a general awareness of careful thinking, but this awareness has not yet become a consistently enacted habit across different academic situations.

The most important substantive finding is that the lowest indicators were clear reasons, credible sources and observations, and being well-informed. These three indicators are closely related to evidence-based reasoning, scientific argumentation, and source evaluation. In science education, this is a critical issue because scientific thinking requires learners to connect claims with evidence, evaluate the quality of data, recognize the limits of information sources, and justify conclusions through transparent reasoning. Evidence-based reasoning is not merely the ability to state an opinion; it involves making claims grounded in appropriate data, considering counter-evidence, and explaining why particular conclusions are more justified than alternatives (Greene & Yu, 2015; Khan & Krell, 2019; Oxman & García, 2020). Therefore, the relatively low scores on these indicators suggest that students may still need stronger support in the epistemic practices that are central to science learning.

This finding has important implications for science teacher education. The participants in this study were science education students who may later become science teachers. If prospective teachers are not strongly disposed to seek clear reasons, use credible sources, and become well-informed before making judgments, they may face difficulties in facilitating scientific inquiry in their future classrooms. Science teachers are expected to help students distinguish claims from evidence, evaluate the credibility of information, and construct explanations based on reliable data. In the current digital information environment, this competence is increasingly important because students frequently encounter scientific claims through online media, where the quality and credibility of information vary widely. Studies on scientific information engagement and science misinformation have emphasized that learners need explicit support in evaluating sources, weighing evidence, and participating responsibly in scientific discourse (Allchin, 2022; Hendriks et al., 2020; Sinatra & Hofer, 2016). Thus, the weak disposition toward credible sourcing and clear reasoning found in this study should be treated as an important curricular signal.

The findings also suggest that science education programs should integrate scientific argumentation more explicitly into early coursework. Scientific argumentation requires students to construct, evaluate, and revise claims using evidence and reasoning. It is not only a communication activity but also a disciplinary practice that reflects how scientific knowledge is built and justified. Students should be given repeated opportunities to formulate claims, identify relevant evidence, compare competing explanations, and defend conclusions with logical reasoning. Inquiry-based tasks, argumentation rubrics, reflective writing, source evaluation assignments, and evidence-based classroom discussions may help strengthen these dispositions. Such strategies are important because telling students to "think critically" is not sufficient. Critical thinking disposition develops when students are repeatedly placed in learning situations where careful reasoning, source evaluation, and evidence-based judgment are expected, practiced, and assessed (Khan & Krell, 2019; Mauriz et al., 2021).

The psychometric findings of this study also require careful interpretation. The internal consistency of the full 36-item questionnaire after reverse scoring was modest, with Cronbach's alpha of 0.522. This value indicates that the instrument did not function as a highly consistent scale in the present sample. Therefore, the questionnaire should not yet be described as having strong reliability. However, this does not mean that the instrument has no value. Rather, it indicates that the instrument is still at a preliminary stage of empirical refinement. This interpretation is consistent with previous studies showing that critical thinking disposition instruments may perform differently across populations, languages, cultures, and educational contexts (Anders et al., 2019; Cui et al., 2020; İskifoğlu, 2013; Wang et al., 2019). Consequently, the use of any critical thinking disposition instrument should be accompanied by population-specific reliability and validity evidence.

The item performance analysis helps explain why the internal consistency was modest. Several items showed weak or negative corrected item-total correlations, and many of these items were negatively worded. This finding is consistent with broader psychometric literature showing that reverse-worded items can create method effects, reduce item-total correlations, lower internal consistency, and distort factor structures (Chae et al., 2017; Kortum et al., 2020; Venta et al., 2022). Reverse-worded items are often included to reduce acquiescence bias, but they may also increase the cognitive burden on respondents. Students must process not only the content of the statement but also the reversed direction of the meaning. For early-year students, especially when the construct is abstract, this can lead to inconsistent responses. Therefore, the weak performance of several negatively worded items in this study suggests that item wording should be reviewed carefully before the instrument is used in larger studies.

The issue of reverse-worded items should not be treated as a minor technical problem. It has direct implications for validity. If negatively worded items introduce wording effects rather than measuring the intended construct, the resulting scores may partly reflect response confusion instead of critical thinking disposition. Previous research has shown that reverse-coded items may function differently across language groups, educational levels, and cultural contexts, and that they can undermine measurement coherence when not carefully designed or modeled (Rammstedt et al., 2010; Song et al., 2020; Venta et al., 2022). For this reason, future refinement of the questionnaire should consider whether the negatively worded items should be rephrased, simplified, replaced with positively worded indicators, or tested using models that account for method effects. Item deletion should not be automatic, because content validity also matters, but weak items should not be retained without further examination.

The exploratory factor analysis provided additional preliminary evidence about the structure of the 12 indicators. Bartlett's test was significant, suggesting that the correlation matrix was factorable. However, the KMO value of 0.547 indicated only marginal sampling adequacy. This means that the EFA results should be interpreted cautiously. A three-factor solution was retained because it was more conceptually coherent than a five-factor solution suggested by the Kaiser criterion. The three factors were interpreted as evidence-based reasoning disposition, contextual and reflective judgment, and self-regulated critical engagement. Conceptually, this structure is reasonable because critical thinking disposition may involve a tendency to use evidence, attend to context, and regulate one's own thinking processes.

Nevertheless, the EFA results should not be interpreted as final construct validation. Several indicators showed low communalities, which means that they were weakly explained by the extracted factors. In addition, the cumulative variance explained by the first three factors was 42.00%, leaving a substantial proportion of variance unexplained. The factor structure is therefore useful as an exploratory indication, but it is not yet sufficient to support a stable measurement model. This conclusion is consistent with the view that critical thinking disposition instruments require repeated psychometric testing across samples and contexts before their factor structures can be considered stable (Anders et al., 2019; Bakhtiari-Dovvombaygi et al., 2024; Cui et al., 2020; Wang et al., 2019). Future studies should use larger samples and confirmatory factor analysis to test whether the three-factor structure can be replicated.

The interpretation of the factor structure also needs to follow the empirical loading pattern rather than only theoretical expectations. Factor 1 was represented most strongly by clear reasons and being well-informed, making the label evidence-based reasoning disposition appropriate but still somewhat narrow. Credible sources and observations, although conceptually related to evidence-based reasoning, loaded more strongly on Factor 2 in the present data. This means that the use of credible sources may have been associated by students with broader contextual and reflective judgment rather than with clear reasoning alone. Factor 2 was more distributed and included basic concern in context, alternatives, total situation, open-mindedness, precision required by situation, and credible sources and observations. Factor 3 was represented mainly by get it right and employ critical thinking abilities, supporting its interpretation as self-regulated critical engagement. These patterns are meaningful, but they remain preliminary because several loadings were weak and some communalities were low.

Taken together, the findings suggest two interconnected implications. Pedagogically, science education programs should strengthen students' evidence-based reasoning, source evaluation, and scientific argumentation from the early semesters. This can be done through learning tasks that require students to justify claims with evidence, compare sources, identify limitations of data, and reflect on how conclusions are formed. Psychometrically, the questionnaire requires further refinement before it can be used as a stable measure of critical thinking disposition. The modest reliability, weak item-total correlations, and marginal EFA suitability indicate that the instrument should be revised and retested. These two implications should be considered together because high-quality instruction requires high-quality assessment. If the measurement tool is not sufficiently reliable, its use for diagnosing students' dispositional strengths and weaknesses will be limited.

This study contributes to the literature by combining an indicator-based profile of critical thinking disposition with preliminary psychometric evidence. Many studies on critical thinking disposition focus on total scores or broad categories, but the present study shows that indicator-level analysis can reveal more specific patterns. The students in this study were not uniformly weak or strong across all dimensions. They showed relatively higher tendencies in situational and reflective aspects, but lower tendencies in clear reasoning, credible source use, and information seeking. This distinction is important because it suggests that instructional interventions should not be generic. Science education programs should target the specific dispositions that are less developed, especially those related to evidence-based reasoning and epistemic judgment.

The study also has limitations. The sample consisted of 90 second-semester science education students from three universities in West Nusa Tenggara Province, Indonesia. Although this provides useful regional evidence, the sample size remains limited for broader generalization and stronger psychometric testing. The dataset also did not include demographic or academic variables, so the study could not examine differences by gender, university, achievement, or prior learning experience. In addition, the study relied on self-report data, which may be influenced by response styles and students' self-perceptions. The EFA was conducted at the indicator level rather than the item level because of sample size considerations. This decision was appropriate for preliminary analysis, but it limits the depth of construct validation.

Future research should address these limitations by involving larger and more diverse samples, including students from different regions, universities, and academic levels. Future studies should also include additional validity evidence, such as correlations with academic achievement, performance-based critical thinking tasks, scientific argumentation assessments, or reflective writing products. Psychometric analysis should be extended through confirmatory factor analysis, measurement invariance testing, and possible method-factor modeling for reverse-worded items. In addition, qualitative methods such as cognitive interviews could be used to examine how students interpret the questionnaire items, especially negatively worded statements. Such evidence would help determine whether weak items reflect poor wording, construct mismatch, or genuine variation in students' dispositions.

In summary, this study indicates that early-year science education students demonstrated a moderate and still-developing critical thinking disposition. Their main challenge was not the complete absence of critical thinking awareness, but the limited strength of more demanding epistemic dispositions, particularly giving clear reasons, using credible sources, and becoming well-informed. The questionnaire provides a useful starting point for profiling these tendencies, but its psychometric performance shows that refinement is still needed. Therefore, the study should be understood both as a substantive profile of students' critical thinking disposition and as a preliminary methodological step toward improving the measurement of critical thinking disposition in science education contexts.

CONCLUSION

This study profiled the critical thinking disposition of early-year science education students and examined preliminary psychometric evidence of an Ennis-based questionnaire. The findings indicate that second-semester science education students from three universities in West Nusa Tenggara Province demonstrated a moderate level of critical thinking disposition. Nearly all students were located in the moderate category, while none reached the high category. At the indicator level, students showed relatively stronger tendencies in precision required by situation, consideration of the total situation, and use of critical thinking abilities.

However, the lowest scores were found in clear reasons, credible sources and observations, and being well-informed. These findings suggest that students had begun to develop general awareness of critical thinking, but their evidence-based reasoning, information-seeking, and source-evaluation dispositions were still limited.

The preliminary psychometric findings suggest that the questionnaire requires further refinement before it can be used as a stable measurement instrument. The internal consistency of the full corrected scale was modest, and several items, particularly negatively worded items, showed weak or negative corrected item-total correlations. The exploratory factor analysis based on 12 indicators produced a tentative three-factor structure consisting of evidence-based reasoning disposition, contextual and reflective judgment, and self-regulated critical engagement. However, the marginal KMO value, low communalities for several indicators, and limited cumulative explained variance indicate that this structure should be interpreted as preliminary rather than confirmatory. Therefore, this study contributes both a substantive profile of students' critical thinking disposition and an initial methodological basis for improving the measurement of critical thinking disposition in science education contexts.

RECOMMENDATION

Science education programs should strengthen early-semester learning experiences that explicitly develop evidence-based reasoning, scientific argumentation, credible source evaluation, and information-seeking habits. Lecturers can integrate inquiry-based tasks, argumentation rubrics, source comparison activities, reflective writing, and evidence-based discussions into science education courses so that students repeatedly practice justifying claims with reliable evidence. For future research, the questionnaire should be revised by reviewing weak items, especially negatively worded statements, and then retested with a larger and more diverse sample. Further validation should include confirmatory factor analysis, measurement invariance testing, method-effect analysis for reverse-worded items, and external validity evidence from academic performance, performance-based critical thinking tasks, or scientific argumentation assessments.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the second-semester science education students from Universitas Mataram, Universitas Islam Negeri Mataram, and Universitas Pendidikan Mandalika who participated in this study. Appreciation is also extended to the respective study programs and lecturers who supported the administration of the questionnaire and facilitated the data collection process. The authors acknowledge that the participants' contributions were essential for profiling critical thinking disposition and generating preliminary psychometric evidence for the instrument used in this study.

FUNDING INFORMATION

This research did not receive external funding.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ni Nyoman Sri Putu Verawati	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Wahyudi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Nina Nisrina	✓	✓		✓			✓	✓		✓		✓	✓	✓

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

All participants were informed about the purpose of the study, the voluntary nature of their participation, the anonymity and confidentiality of their responses, and their right to decline or withdraw without consequence before they provided consent to complete the questionnaire.

ETHICAL APPROVAL

The researchers meticulously followed ethical protocols throughout the research process, adhering to the principles outlined in the Declaration of Helsinki.

REFERENCES

Abrami, P. C., Bernard, R., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. https://doi.org/10.3102/0034654308326084

Allchin, D. (2022). Ten competencies for the science misinformation crisis. *Science Education, 107*(2), 261–274. https://doi.org/10.1002/sce.21746

Anders, P. L., Davis, E., & McCall, W. D. (2020). Psychometric properties of an instrument to assess critical thinking disposition and metacognition in dental students. *Journal of Dental Education, 84*(5), 559–565. https://doi.org/10.1002/jdd.12038

Anders, P. L., Stellrecht, E., Davis, E., & McCall, W. D. (2019). A systematic review of critical thinking instruments for use in dental education. *Journal of Dental Education, 83*(4), 381–397. https://doi.org/10.21815/jde.019.043

Bakhtiari-Dovvombaygi, H., Pourhasan, K., Rahmaty, Z., Zare-Kaseb, A., Abbaszadeh, A., Rashtbarzadeh, A., & Borhani, F. (2024). Evaluation of cross-cultural adaptation and validation of the Persian version of the critical thinking disposition scale: Methodological study. *BMC Nursing, 23*(1). https://doi.org/10.1186/s12912-024-02129-y

Chae, D., Kang, K. H., Benkert, R., & Doorenbos, A. Z. (2017). Evaluation of the psychometric properties of the Korean version of the Cultural Competence Assessment. *Japan Journal of Nursing Science, 15*(1), 56–66. https://doi.org/10.1111/jjns.12169

Chen, S., & Fan, J. (2022). Validation of the psychometric properties of the Self-Directed Learning Readiness Scale. *Nursing Open, 10*(3), 1639–1646. https://doi.org/10.1002/nop2.1418

Cui, L., Zhu, Y., Qu, J., Tie, L., Wang, Z., & Qu, B. (2020). Psychometric properties of the Critical Thinking Disposition Assessment Test amongst medical students in China: A cross-sectional study. *Research Square*. https://doi.org/10.21203/rs.3.rs-16612/v1

Ennis, R. H. (2015). Critical thinking: A streamlined conception. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 31–47). Palgrave Macmillan US. https://doi.org/10.1057/9781137378057_2

Greene, J. A., & Yu, S. B. (2015). Educating critical thinkers. *Policy Insights From the Behavioral and Brain Sciences, 3*(1), 45–53. https://doi.org/10.1177/2372732215622223

Hendriks, F., Mayweg-Paus, E., Felton, M., Iordanou, K., Jucks, R., & Zimmermann, M. (2020). Constraints and affordances of online engagement with scientific information: A literature review. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.572744

İskifoğlu, G. (2013). Cross-cultural equivalency of the California Critical Thinking Disposition Inventory. *Educational Sciences: Theory & Practice, 14*(1), 159–178. https://doi.org/10.12738/estp.2014.1.1840

Khan, S., & Krell, M. (2019). Scientific reasoning competencies: A case of preservice teacher education. *Canadian Journal of Science, Mathematics and Technology Education, 19*(4), 446–464. https://doi.org/10.1007/s42330-019-00063-9

Kortum, P., Acemyan, C. Z., & Oswald, F. L. (2020). Is it time to go positive? Assessing the positively worded System Usability Scale (SUS). *Human Factors, 63*(6), 987–998. https://doi.org/10.1177/0018720819881556

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series, 2014*(1), 1–23. https://doi.org/10.1002/ets2.12009

Mauriz, B. P., Blanco-Anaya, P., & Maceira, J. J. P. (2021). “Fake news” or real science? Critical thinking to assess information on COVID-19. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.646909

Oxman, A. D., & García, L. M. (2020). Comparison of the Informed Health Choices Key Concepts Framework to other frameworks relevant to teaching and learning how to think critically about health claims and choices: A systematic review. *F1000Research, 9*, 164. https://doi.org/10.12688/f1000research.21858.1

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*(1), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

Sinatra, G. M., & Hofer, B. K. (2016). Public understanding of science: Policy and educational implications. *Policy Insights From the Behavioral and Brain Sciences, 3*(2), 245–253. https://doi.org/10.1177/2372732216656870

Song, Y., Luzzi, L., Chrisopoulos, S., & Brennan, D. S. (2020). Are trust and satisfaction similar in dental care settings? *Community Dentistry and Oral Epidemiology, 48*(6), 480–486. https://doi.org/10.1111/cdoe.12559

Venta, A., Bailey, C., Walker, J., Mercado, A., Colunga-Rodríguez, C., Ángel-González, M., & Dávalos-Picazo, G. (2022). Reverse-coded items do not work in Spanish: Data from four samples using established measures. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.828037

Wang, X., Sun, X., Huang, T., He, R., Hao, W., & Zhang, L. (2019). Development and validation of the critical thinking disposition inventory for Chinese medical college students (CTDI-M). *BMC Medical Education, 19*(1). https://doi.org/10.1186/s12909-019-1593-z