Jurnal Elemen, 12(3), 689-709, July 2026
https:/doi.org/10.29408/jel.v12i3.35213
e- ISSN: 2442 - 4226

Can you read this chart? Assessing graph literacy competence and confidence calibration among Indonesian undergraduate health science students

Bustang Bustang [1] *, Garry Kuan [2]

[1] Department of Sport Science, Universitas Negeri Makassar, South Sulawesi, Indonesia
[2] School of Health Sciences, Universiti Sains Malaysia, Malaysia

* Correspondence: b.bustang@unm.ac.id

Abstract

Despite growing international interest in graph literacy among health science students, empirical evidence from Indonesian higher education remains scarce, particularly regarding the alignment between students' competence and their confidence in interpreting graphical health information. This study investigated graph literacy competence among undergraduate health science students at a public Indonesian university, examined the relationship between competence and confidence, and identified areas of difficulty. A cross-sectional design was employed with 78 undergraduate health science students from Years 1 to 4. The validated 13-item Graph Literacy Scale was administered alongside an item-level confidence rating task. Statistical and calibration analysis were used to examine performance, confidence, and their alignment. Results showed that students achieved a mean score of 5.31 (SD = 2.19) out of 13, substantially below international benchmarks. Calibration analysis revealed a non-significant mean calibration score, indicating adequate group-level metacognitive calibration. A weak-to-moderate positive correlation between competence and confidence was found (r = 0.34, p = .002), suggesting that more competent students tended to be more confident. These findings highlight the urgent need to integrate explicit graph literacy instruction into Indonesian health science curricula to better prepare future healthcare professionals for data-informed clinical practice.

Keywords: confidence; graph literacy; health science students; Indonesian university; metacognitive monitoring

How to cite: Bustang, B., & Kuan, G. (2026). Can you read this chart? Assessing graph literacy competence and confidence calibration among Indonesian undergraduate health science students. *Jurnal Elemen, 12*(3), 689-709. https://doi.org/10.29408/jel.v12i3.35213

Received: 4 June 2026 | Revised: 2 July 2026
Accepted: 10 July 2026 | Published: 13 July 2026

Introduction

In contemporary healthcare and scientific communication, the ability to read, interpret, and critically evaluate graphical information has become an increasingly important professional skill. Charts, diagrams, and statistical visualizations are now routinely embedded in clinical guidelines, public health campaigns, research articles, and patient education materials. This ability is commonly referred to as *graph literacy*. Roth (2002) described graph reading as a form of dynamic semiotic activity in which a reader must first identify significant visual features as signs, and then engage in a dialectic process of connecting those signs to their real-world referents. From this perspective, graph literacy is not a fixed trait. Rather, it is a developing competence shaped by experience, context, and familiarity with the conventions of visual data representation (Bodén & Stenliden, 2019; Procter-Legg et al., 2026). When a reader is deeply familiar with the subject matter, the graph becomes transparent, functioning as a window into the phenomenon it represents. In unfamiliar contexts, however, even expert readers can struggle (Roth, 2002).

The definition of graph literacy has been discussed in several ways across the literature. Durand et al. (2020) described it as the ability to understand and use information presented in graphical formats, and argued that it forms a distinct component of health literacy that is closely connected to, but not reducible to, numeracy. This view is consistent with earlier work by Galesic and Garcia-Retamero (2011), who defined graph literacy as the ability to understand and use information from graphs in everyday contexts, and who developed a validated scale to measure individual differences in this capacity across cultural settings. Friel et al. (2001) further proposed that graph comprehension operates across three hierarchical levels. At the first level, a reader should be able to *read the data*, meaning to locate specific pieces of information directly from a graph. At the second level, a reader should be able to *read between the data*, meaning to identify relationships and patterns as shown in the graph. The third and most advanced level requires the ability to *read beyond the data*, meaning to make inferences and predictions that go further than what is explicitly displayed. This three-level framework has since been widely adopted in health literacy and mathematics education research as a practical guide for both assessing and developing graph comprehension skills (Galesic & Garcia-Retamero, 2011; Soler Wenglein et al., 2025).

The importance of graph literacy becomes especially clear in health-related settings. Healthcare professionals must interpret epidemiological trends, evaluate clinical trial outputs, and communicate statistical information about treatment risks and benefits (Reading Turchioe & Mangal, 2024). Patients, in turn, are expected to engage with increasingly visual forms of health information to make informed choices about their own care (van Strien-Knippenberg et al., 2025). Rothman et al. (2008) noted that numeracy is a distinct and vital predictor of patient safety, independent of general reading ability. Numeracy refers to the functional ability to understand and apply numerical and statistical information in everyday and professional contexts, forming an essential foundation for undergraduate health professional training (Valcheff et al., 2026). This view is supported by Bustang (2026), who conceptualized health numeracy not merely as arithmetic competence but as a form of applied mathematical reasoning deeply integrated with clinical decision-making processes. Patients with limited numeracy, including those who struggle to interpret graphs and data visualizations, are more likely to misunderstand medical instructions and miscalculate health risks, which can lead to poorer health outcomes (Danya et al., 2021; Rothman et al., 2008).

Durand et al. (2020) investigated the relationship between graph literacy, numeracy, and health literacy in a Medicaid-eligible population in the United States. Their findings showed that graph literacy was the only factor significantly associated with overall health information comprehension. Individuals with lower graph literacy struggled to extract meaning from visual formats, regardless of the format type used. Importantly, preferences for a given format did not necessarily reflect actual comprehension, which suggests that designing health communication based on patient preferences alone may be insufficient. These findings reinforce the argument that graph literacy is a skill worth measuring and developing in its own right, particularly among populations who may have had limited exposure to data visualization in their educational backgrounds. This concern extends naturally to health science students in developing countries, who represent the next generation of healthcare providers and whose quantitative skills have received relatively little empirical attention.

Gaissmaier et al. (2012) investigated how individual differences in graph literacy affect the effectiveness of different health information formats. Their study found that individuals with high graph literacy comprehended and recalled health information more effectively when it was presented graphically. Those with lower graph literacy, however, performed better with purely numerical formats. Interestingly, both groups found graphical displays more visually attractive regardless of their comprehension levels. Gaissmaier et al. (2012) concluded that there is no universal best format for communicating health risk, and that tailoring information to an individual's literacy profile is more likely to support genuine understanding. These findings highlight the need to first assess the graph literacy skills of healthcare providers and the populations they serve before designing health communication materials.

Several empirical studies have begun to document graph literacy levels among health science and medical student populations. The body of evidence, however, remains relatively limited, particularly outside Western countries. Soler Wenglein et al. (2025) conducted a cross-sectional survey of 449 undergraduate medical students in Germany using the validated 13-item Graph Literacy Scale originally developed by Galesic and Garcia-Retamero (2011). Their results showed that these students achieved a mean score substantially above the German general population average, indicating a general level of proficiency in reading and interpreting graphical data. However, a particularly concerning finding emerged at the highest level of graph comprehension. Fewer than half of the students correctly responded to items requiring recognition of the limitations of ambiguous or incomplete visual information, indicating notable weakness in their ability to read beyond the data (Soler Wenglein et al., 2025). Equally concerning, graph literacy scores did not improve significantly as students advanced through their academic years, which suggests that formal medical education alone may not be systematically building this skill over time. These findings raise important questions about whether health science curricula are adequately preparing students to critically engage with data-intensive clinical materials.

Beyond competence, an important but often underexplored dimension of health literacy research concerns the relationship between students' actual performance and their confidence in that performance. This relationship is commonly referred to as *metacognitive calibration* (Gutierrez de Blume, 2022). Kruger and Dunning (1999) demonstrated that individuals with limited competence in a given domain tend to significantly overestimate their own ability. This occurs because the cognitive skills needed to perform well are the same skills needed to recognize poor performance, a pattern they described as a double burden of incompetence. This overconfidence gap is particularly consequential in health professional education (Knof et al., 2024; Rahmani, 2020). Students who believe they can correctly interpret clinical data may not recognize their knowledge gaps or feel the need to seek additional learning support. Schraw (2009) argued that metacognitive monitoring, meaning the ability to accurately judge one's own level of understanding, is a teachable skill that can be cultivated through instructional experiences that provide frequent and accurate feedback. In a directly relevant Indonesian study, Bustang et al. (2026) found a statistically significant overconfidence gap among undergraduate nutrition students solving health numeracy tasks. These students maintained high confidence despite demonstrably inaccurate performance, and this misalignment was identified as a tangible risk to future patient safety. Bustang's (2022) earlier doctoral research on probabilistic reasoning among Indonesian students and mathematics teachers similarly documented systematic miscalibration of metacognitive judgements, suggesting this pattern may be widespread across Indonesian educational contexts and not limited to any single discipline.

Despite growing international interest in graph literacy, there remains a significant gap in research on this topic within Indonesian higher education, particularly in health science programs. Indonesia is a large and rapidly developing country whose health sector increasingly relies on data-informed practice, evidence-based medicine, and digital health communication. Indonesian health science students, including those in nursing, pharmacy, public health, nutrition, and other allied health disciplines, are expected throughout their careers to engage with statistical reports, graphical data in research publications, and visual health communication materials. Yet little is empirically known about how well equipped these students currently are to do so. Bustang (2026) has argued in the Indonesian context that numeracy in health domains is a foundational competency for safe and effective clinical decision-making, and that its absence at the student level represents a significant gap in health professional preparation. Furthermore, Bustang (2022) demonstrated that Indonesian students' reasoning patterns and metacognitive tendencies in mathematically oriented tasks can differ substantially from those documented in Western populations. These differences arise from culturally specific educational experiences and varying levels of prior exposure to mathematical and statistical representations. Given that most existing graph literacy studies have been conducted in high-income Western countries, their findings cannot simply be assumed to generalize to the Indonesian educational context.

The present study aimed to address this gap by investigating the graph literacy competence and calibration confidence of undergraduate health science students at a public university in Indonesia. The study used the 13-item Graph Literacy Scale developed by Galesic and Garcia-Retamero (2011), the same validated instrument used in comparable international studies. Three core research questions guided this investigation. First, what is the overall level of graph literacy competence among these students? Second, how well calibrated are their confidence levels relative to their actual performance? Third, which specific dimensions or item types within the scale present the greatest challenges? Understanding these patterns is essential for designing targeted, evidence-based curricular interventions that prepare future Indonesian healthcare professionals to read, interpret, and communicate graphical health information with both technical skill and appropriate epistemic humility (Bustang et al., 2026; Schraw, 2009). The findings are also intended to contribute to a growing body of international literature on quantitative literacy in health professional education, from a perspective that has been largely absent in existing scholarship.

Methods

Research design and participants

This study employed a cross-sectional design to assess the graph literacy competence and confidence of undergraduate health science students. A convenience sample of 78 students enrolled in Years 1 to 4 of the undergraduate health science program at the Faculty of Sport and Health Sciences, Universitas Negeri Makassar, Indonesia was recruited. Convenience sampling was used because the study required voluntary participation within an accessible institutional setting, and students were recruited during regular class schedules. The sample was predominantly female, with 72 female students (92%) and only 6 male students (8%). This gender distribution reflects the broader enrolment patterns in Indonesian health science faculties, where female students constitute the large majority of the student population, consistent with national health workforce data reported by Statistics Indonesia (BPS-Statistics Indonesia, 2025). The mean age of all participants was 20.2 years (SD = 1.3). In Indonesia, undergraduate health science programs typically span four years, with students admitted directly from senior secondary school. In terms of study year, Year 2 students formed the largest group (n = 36.46%), followed by Year 4 students (n = 26.33%), while Year 3 students were the smallest group (n = 5.6%). Regarding program of study, Health Administration was the most represented program (n = 42.54%), followed by Nutrition (n = 26.33%). The remaining participants were enrolled in Physiotherapy, Public Health, and other programs.

Participation in this study was entirely voluntary and anonymous. Students were not asked to provide any personal, biographical, or educational background information. It is worth noting, however, that statistical content is included in the Indonesian national secondary school curriculum and is also covered in undergraduate programs across all disciplines, including health sciences. This means that all participants had some prior exposure to quantitative and data-related content before taking part in the study.

Table 1. Participant characteristics

Characteristic	**N**	**Male, N = 6**[1]	**Female, N = 72**[1]	**Overall, N = 78**[1]
Age[2]	78	20.3 (2.1)	20.2 (1.3)	20.2 (1.3)
Program of study	78			
Health Administration		3 (50%)	39 (54%)	42 (54%)
Nutrition		0 (0%)	26 (36%)	26 (33%)
Physiotherapy		1 (17%)	5 (7%)	6 (8%)
Public Health		0 (0%)	2 (3%)	2 (3%)
Other		2 (33%)	0 (0%)	2 (3%)
Entry cohort	78			
2022		2 (33%)	24 (33%)	26 (33%)
2023		1 (17%)	4 (6%)	5 (6%)
2024		1 (17%)	35 (49%)	36 (46%)
2025		2 (33%)	9 (12%)	11 (14%)

[1]*Mean (SD); n (%)*
[2]*Age in years*

Instrument: Graph literacy scale

The primary instrument used in this study was the Graph Literacy Scale developed by Galesic and Garcia-Retamero (2011). This is a validated 13-item scale designed to measure an individual's ability to read and interpret graphical information. The scale comprises nine items requiring numeric responses and four multiple-choice items. The items are distributed across three hierarchical levels of graph comprehension as described by Friel et al. (2001). The first level, "*reading the data*", consists of four numeric response items that require students to directly extract specific values from a graph. The second level, "*reading between the data*", includes three numeric response items and one multiple-choice item, requiring students to identify relationships and patterns within graphical displays. The third and most cognitively demanding level, "*reading beyond the data*", consists of two numeric response items and three multiple-choice items, requiring students to make inferences and draw conclusions that extend beyond what is explicitly shown in the graph. Example tasks included reading a specific point from a line chart (reading the data) and comparing the effectiveness of two fictitious medications presented in a visual format (reading between the data). Calculators were not permitted during administration. Only fully completed questionnaires were included in the analysis, and incomplete responses were excluded. The complete Bahasa Indonesia version of the instrument is accessible at https://bit.ly/LGK2026.

Prior to administration, the Graph Literacy Scale was translated into Indonesian through a rigorous forward and backward translation procedure following established guidelines for cross-cultural adaptation of self-report instruments (Beaton et al., 2000; Degroot et al., 1994). The forward translation was carried out by an independent bilingual translator, followed by a synthesis and reconciliation step. The translated version was then back-translated into English by a separate translator who had no access to the original scale. Discrepancies between the back-translated and original versions were reviewed and resolved by an expert panel to ensure semantic, idiomatic, and conceptual equivalence. The instrument was subsequently administered in Indonesian (Bahasa Indonesia) to all participants. Internal consistency of the scale in the present sample was assessed using Cronbach's alpha, which yielded a value of .54. Although this value is lower than the .74 and .79 reported by Galesic and Garcia-Retamero (2011) for German and United States samples, respectively, it is consistent with the scale's design as a heterogeneous instrument covering multiple distinct graph types and comprehension levels, which inherently limits internal consistency as a reliability indicator. The relatively low alpha in this study may also reflect the homogeneity of the sample and the increased difficulty of the items in the Indonesian context.

Table 2. Indicators of graph literacy scale (Galesic & Garcia-Retamero, 2011)

Indicators	Items	Contents
Reading the data	Question 1 (Q1)	Reading off a point on a bar chart
	Question 3 (Q3)	Knowing what a quarter of a pie is in %
	Question 5 (Q5)	Reading off a point on a line chart
	Question 8 (Q8)	Reading off number of icons
Reading between the data	Question 2 (Q2)	Determining difference between 2 bars
	Question 4 (Q4)	Summing slices within a quarter of a pie
	Question 6 (Q6)	Comparing slopes of a line at 2 intervals
	Question 9 (Q9)	Determining difference between 2 groups of icons
Reading beyond the data	Question 7 (Q1)	Projecting future trend from a line chart
	Question 10 (Q1)	Comparing 2 bar charts: Attending to scale range
	Question 11 (Q1)	Comparing 2 line charts: Attending to scale labels
	Question 12 (Q1)	Differentiating slope and height of a line
	Question 13 (Q1)	Reading off a point on a bar chart when bar falls between 2 labels

Confidence assessment

In addition to the graph literacy items, students were asked to rate their confidence for each item immediately after providing their answer. This confidence assessment approach is consistent with the metacognitive monitoring methodology used in prior research on calibration (Bustang et al., 2026; Foster, 2016; Schraw, 2009). For each of the 13 items, students indicated how confident they were that their answer was correct by selecting a whole number on a scale from 0 (completely unsure) to 10 (completely certain). This procedure allowed for the collection of paired data, that is, an accuracy score and a confidence rating, for each item and each participant. The use of item-level confidence ratings is a well-established method for examining the alignment between students' subjective certainty and their objective performance in educational assessment contexts (Kruger & Dunning, 1999; Schraw, 2009).

Scoring

Each correctly answered item was awarded one point, giving a possible total score ranging from 0 to 13. For numeric response items, only exactly correct answers were accepted, with no margin of error permitted. For multiple-choice items, only the designated correct option was awarded the point. Partially completed questionnaires were not included in the analysis. The confidence ratings for each participant were averaged across all 13 items to produce a mean confidence score per participant. Both accuracy scores and confidence ratings were then converted to percentages to allow for direct comparison. The calibration score for each student was calculated by subtracting the mean accuracy percentage from the mean confidence percentage. A calibration score of zero indicated perfect calibration, where a student's confidence perfectly matched their actual performance. Positive values indicated *overconfidence*, meaning a student's confidence exceeded their actual accuracy, while negative values indicated *underconfidence*, meaning a student's confidence was lower than their actual performance. This approach to computing calibration is consistent with the definition of metacognitive monitoring accuracy used in the broader literature (Schraw, 2009; Stankov & Lee, 2014).

Data analysis

Descriptive statistics, including means and standard deviations, were calculated to summarize the overall graph literacy performance of the participants as well as their confidence and calibration scores. Students were further categorized into distinct calibration profiles based on the relationship between their accuracy and their confidence. These profiles included students who were *well-calibrated* (accurate and appropriately confident), *overconfident* (high confidence but relatively low accuracy), and *underconfident* (low confidence but relatively high accuracy). To test whether the calibration scores of the group as a whole differed significantly from zero, a one-sample t-test was conducted using zero as the reference value representing perfect calibration. This approach allowed for an evaluation of whether students, on average, systematically overestimated or underestimated their graph literacy ability. All statistical analyses were performed at a significance level of .05.

Results

Overall graph literacy performance

The 78 Indonesian undergraduate health science students who participated in this study achieved a mean graph literacy score of 5.31 points (SD = 2.19) out of a maximum possible score of 13. This result was considerably lower than scores reported in comparable studies conducted in Western countries. Soler Wenglein et al. (2025) reported a mean score of 11.42 (SD = 1.42) among undergraduate medical students in Germany, while Galesic and Garcia-Retamero (2011) reported mean scores of 9.4 (SD = 2.6) and 9.3 (SD = 2.9) for German and United States general population samples, respectively. The present findings therefore suggest that the Indonesian health science students in this study performed substantially below international benchmarks, achieving roughly half the score reported among German medical students. This gap is notable and points to a potential concern regarding the quantitative readiness of health science students in the Indonesian context, particularly given the growing reliance on data-driven practice in contemporary healthcare settings.

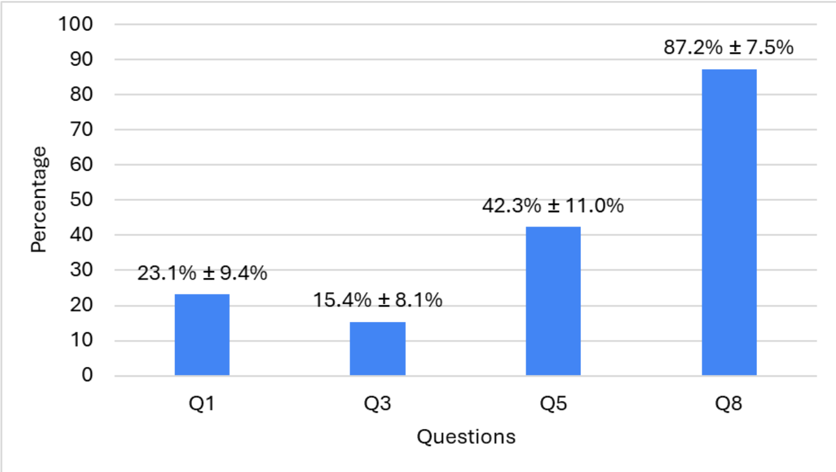

Figure 1. Percentage of correct responses for "reading the data" items

Performance on reading the data items

Figure 1 presents the percentage of students who answered each item correctly within the "reading the data" category, which comprised four questions (Q1, Q3, Q5, and Q8). This category represents the most foundational level of graph comprehension, requiring students to directly extract specific information from a graph without the need for inference or interpretation (Galesic & Garcia-Retamero, 2011). Despite the relatively straightforward cognitive demands of these items, the results revealed considerable variation in student performance. Item Q8, which asked students to count the number of women among 100 patients with a particular disease using an icon array, was the most successfully answered item in this category, with approximately 87% of students providing a correct response. This result is somewhat expected, as icon arrays present information in a highly concrete and perceptually direct format that minimizes the need for mathematical reasoning. In contrast, item Q1, which required students to read the height of a specific bar in a fully labelled bar chart showing the percentage of cancer patients who recovered after chemotherapy, was answered correctly by only approximately 23% of students. Similarly, Q3, which asked students to identify the approximate percentage of cancer-related deaths attributable to lung cancer from a pie chart, yielded a correct response rate of only around 15%. These results are particularly striking given that in the original Galesic and Garcia-Retamero (2011) study, Q1 and Q3 were answered correctly by approximately 84% and 86% of the German and US general population samples, respectively. The sharp contrast between these international benchmarks and the performance of Indonesian health science students on even the most basic items suggests a fundamental gap in foundational graph reading skills that warrants serious attention.

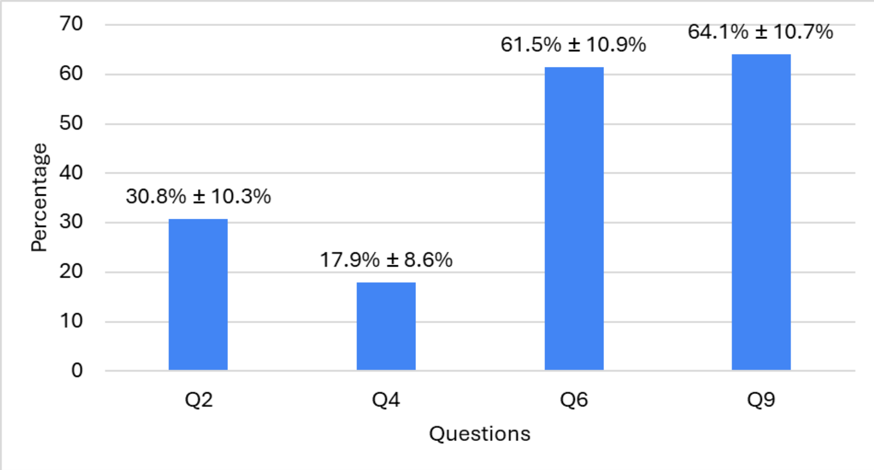

Figure 2. Percentage of correct responses for "reading between the data" items

Performance on reading between the data items

Figure 2 displays the correct response rates for the four items in the "reading between the data" category. This level requires students to move beyond simple data extraction and instead identify relationships, make comparisons, and interpret trends within graphical information (Friel et al., 2001). The results showed a wide range of performance across these four items. Item Q9, which asked students to calculate how many more men than women were represented among 100 patients with a particular disease based on an icon array, had the highest correct response rate in this category at approximately 64%. Item Q6, which required students to judge at which time interval the increase in the prevalence of an imaginary disease called Adeolitis was steeper, based on a line chart, was answered correctly by around 61% of students. In contrast, item Q2, which asked students to calculate the difference between the recovery rates of two different cancer treatments as shown on a bar chart, was answered correctly by only approximately 31% of students. The most poorly answered item in this category was Q4, with a correct response rate of only around 18%. Item Q4 required students to sum several specific slices of a pie chart to determine the combined percentage of deaths from colon cancer, breast cancer, and prostate cancer. This type of task demands both accurate visual parsing of the pie chart and a basic arithmetic operation, and the very low correct response rate suggests that many students struggled with this combined demand. For comparison, Galesic and Garcia-Retamero (2011) reported correct response rates of approximately 68% and 75% for Q4 in Germany and the United States, respectively, indicating that the Indonesian health science students in this study performed at less than a quarter of those benchmarks on this particular item.

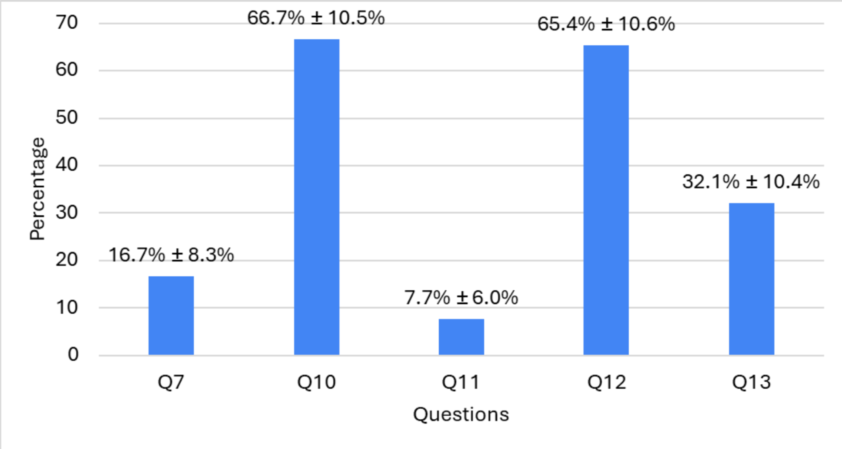

Figure 3. Percentage of correct responses for "reading beyond the data" items

Performance on reading beyond the data items

Figure 3 presents the correct response rates for the five items assessing the highest level of graph comprehension, described as "reading beyond the data". These items require students to make inferences, extrapolate trends, and draw conclusions that extend beyond what is explicitly shown in the graph, representing the most cognitively demanding form of graph literacy (Friel et al., 2001; Galesic & Garcia-Retamero, 2011). Performance across these five items was notably uneven. Items Q10 and Q12 were answered correctly by approximately 67% and 65% of students, respectively. Item Q10 asked students to compare the effectiveness of two different heart disease drugs whose data were presented on two separate bar charts with different scale ranges, requiring students to notice that the visual difference between bars was misleading due to differing axes. Item Q12 required students to compare the rate of increase of two imaginary diseases, Coliosis and Tiosis, displayed on the same line chart, by correctly attending to the slope rather than the height of the lines. The relatively higher correct response rates on these two items, compared to the other items at this level, may suggest that students performed better when both datasets were presented in the same visual space, as in item Q12, or when the task involved a direct comparison with visible reference points. In contrast, item Q13, which asked students to read the percentage of cancer patients who died after chemotherapy from a bar chart whose bars fell between labelled gridlines rather than exactly on them, was answered correctly by only around 32% of students. Most notably, item Q11, which asked students to compare two line charts for different psoriasis treatments where the y-axis scales were entirely unlabeled, was answered correctly by only approximately 8% of students. The correct answer for item Q11 was "can't say", requiring students to recognize that a valid comparison was impossible due to missing scale information. This finding is consistent with results reported by Galesic and Garcia-Retamero (2011), where item Q11 was also the most difficult item in both Germany (16% correct) and the United States (20% correct). The fact that Indonesian health science students performed even more poorly on this item, at less than half the rate seen in Western general population samples, suggests particular difficulty in tolerating ambiguity and recognizing the limits of graphical information, which corresponds to the highest cognitive demand of the "reading beyond the data" level.

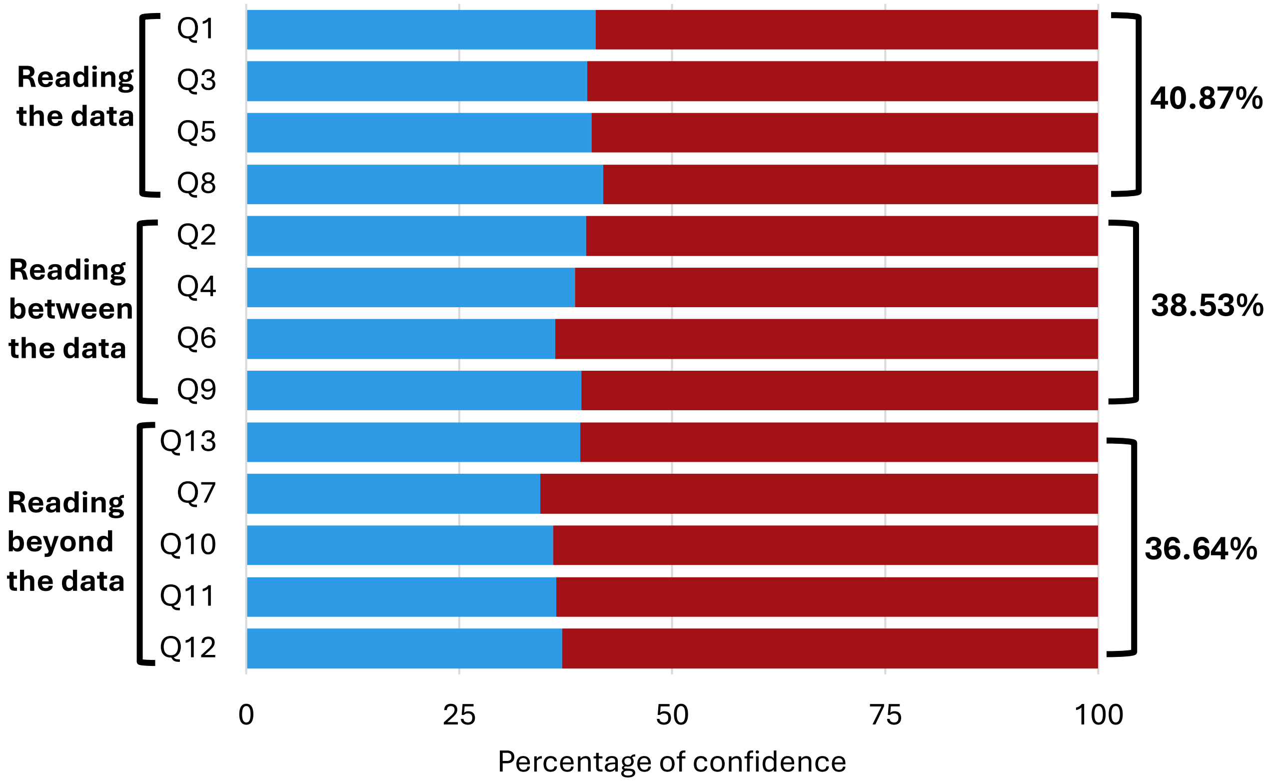

Figure 4. Mean percentage of confidence ratings by questions and categories for the graph literacy scale. Blue = mean confidence %, red = remainder to 100%.

Student confidence in graph literacy tasks

The confidence levels of the 78 undergraduate health science students across all 13 items of the Graph Literacy Scale are shown in figure 4. Overall, students demonstrated consistently low confidence across all three categories. The mean confidence percentages were 40.87% for "reading the data", 38.53% for "reading between the data", and 36.64% for "reading beyond the data". These figures indicate that, on average, students were uncertain about their responses across all levels of graph comprehension, with confidence declining gradually as the cognitive demands of the tasks increased. This general pattern of low confidence is notable in itself, as it suggests that students were broadly aware of their limited ability to engage with graphical information, regardless of the task type.

Within the "reading the data" category, the mean confidence of 40.87% was the highest of the three groups, yet it remained well below the midpoint of the confidence scale. When viewed alongside the actual performance data, a clear mismatch is apparent. Students answered Q8, which involved reading an icon array, correctly at a rate of approximately 87%, yet their overall category confidence sat at only 40.87%, suggesting that the relatively strong performance on Q8 was not sufficient to raise students' perceived certainty across the group. More strikingly, items Q1 and Q3, which were answered correctly by only 23% and 15% of students respectively, appear to have contributed to suppressing confidence in this category. A similar pattern emerged in the "reading between the data" category, where the mean confidence of 38.53% was slightly lower. Correct response rates on Q2 and Q4 were only 31% and 18%, and the overall confidence level for this category reflects students' limited ability to identify relationships and make comparisons from graphical displays. The confidence level for "reading beyond the data" was the lowest of all three categories at 36.64%, which corresponds with the most variable performance observed in this group, where correct response rates ranged from as low as 8% on Q11 to as high as 67% on Q10.

Taken together, the confidence data reveal a pattern that is broadly consistent with students' actual performance, in that confidence was lowest in the most cognitively demanding category and highest in the most foundational one. However, even in the category where performance was relatively strongest, confidence remained low in absolute terms. This overall pattern of low confidence combined with low accuracy suggests that the students in this study were, to some extent, aware of the limits of their graph literacy skills. This stands in contrast to the overconfidence pattern documented in related health literacy research, where students maintained high confidence despite poor performance (Bustang et al., 2026). A more detailed examination of calibration scores, which quantify the alignment between confidence and accuracy at the individual level, is presented in the following section.

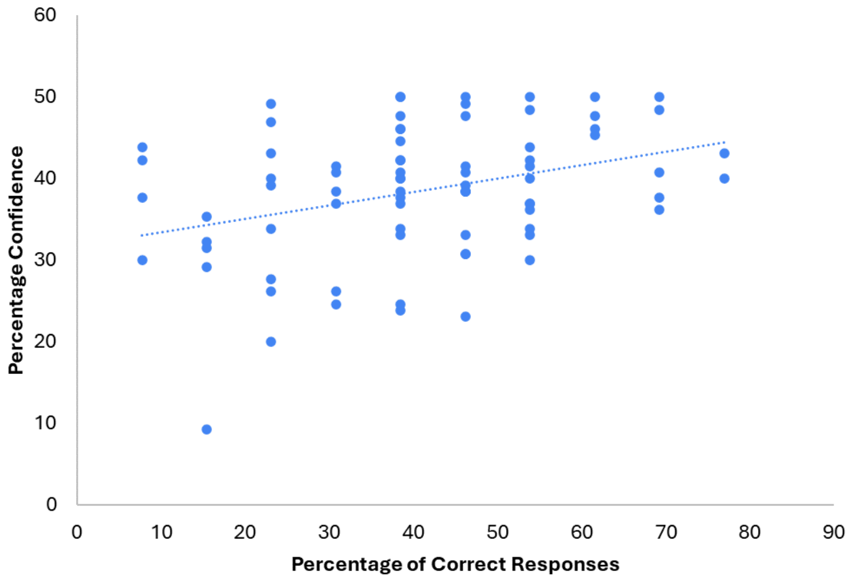

Figure 5. Confidence against graph literacy competence for each health sciences students

Calibration analysis of students' graph literacy

To assess the degree of alignment between actual performance and self-assessed ability, a calibration score was computed for each of the 78 undergraduate health science students as an index of metacognitive monitoring accuracy. This score reflects the statistical correspondence between objective graph literacy competence, expressed as the percentage of correct responses across 13 items, and subjective confidence judgments, expressed as mean confidence ratings standardized to a percentage scale. Calibration scores were derived by subtracting each student's mean competence percentage from their mean confidence percentage, yielding a signed difference score. A negative value indicates that a student rated their confidence below their actual performance (*underconfidence*), whereas a positive value signals that confidence exceeded performance (*overconfidence*). The mean calibration score for the sample was −2.31% (SD = 16.04), suggesting that, on average, students were marginally underconfident in their graph literacy ability, though the wide standard deviation reflects considerable variability in metacognitive accuracy across individuals.

To determine whether the observed mean calibration score departed systematically from a state of perfect calibration, a one-sample t-test was conducted with a test value of zero, representing exact correspondence between confidence and competence. The analysis yielded a non-significant result, $t(77) = -1.27$, $p = .208$, 95%CI [−5.92, 1.31], Cohen's d = −0.14. Although the sample mean calibration score was −2.31% (SD = 16.04), indicating a slight directional tendency toward underconfidence, this deviation did not reach statistical significance at the $\alpha = .05$ threshold. The negligible effect size ($|d| < 0.20$) further confirms that the magnitude of miscalibration was trivial at the group level. The ANOVA analysis further showed no significant difference in calibration scores across the four study programs ($F(4, 73) = 1.97$, $p = .109$, $\omega^2 = .047$), suggesting that the calibration pattern was consistent regardless of academic discipline, and that the tendency toward slightly conservative self-assessment was not specific to any single program of study. Taken together, these findings indicate that the cohort, as a whole, exhibited adequate metacognitive calibration. Their aggregate confidence judgments were not significantly biased relative to actual graph literacy performance. Nevertheless, the wide dispersion of individual calibration scores (SD = 16.04; range: −36.92% to 36.15%) underscores that group-level calibration accuracy masks substantial individual variability, with a considerable proportion of students holding systematically inflated or deflated confidence relative to their demonstrated competence. This heterogeneity in metacognitive accuracy warrants targeted instructional attention and suggests that calibration training may benefit particularly from individualized feedback mechanisms.

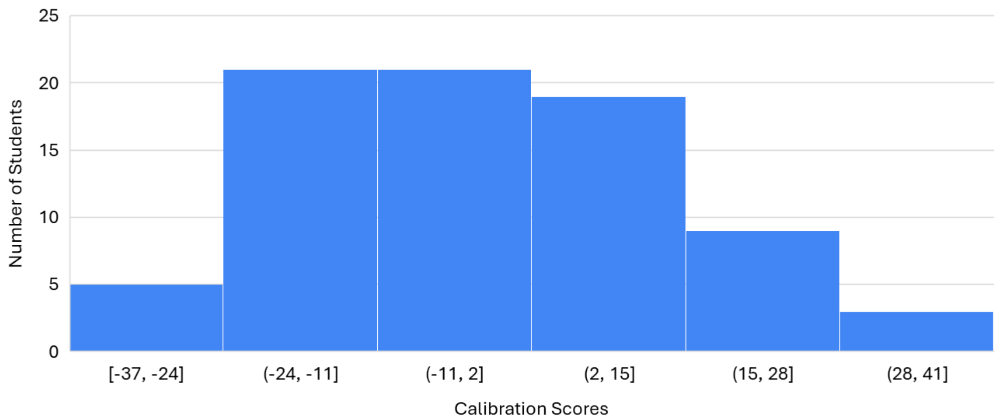

Figure 6. Distribution of calibration scores

Examining the distribution of calibration patterns reveals that the majority of students demonstrated underconfidence: 43 out of 78 students (55.1%) produced negative calibration scores, meaning their confidence consistently fell short of their measured competence (see Figure 6). In contrast, 34 students (43.6%) were overconfident, perceiving themselves as more capable than their performance indicated. Only one student (1.3%) achieved a calibration score of exactly zero, representing perfect alignment between confidence and competence. The mean competence score across the sample was 40.83% (SD = 16.88), while the mean confidence score was 38.52% (SD = 8.14). Notably, the substantially smaller standard deviation for confidence compared to competence suggests that students tended to cluster their confidence estimates within a narrower range than the true spread of their performance, a pattern consistent with regression toward the mean in confidence judgments, where both low and high performers moderate their self-assessments toward an intermediate value.

A Pearson correlation analysis was conducted to examine the relationship between graph literacy competence and confidence ratings across the 78 students. The results revealed a weak-to-moderate positive correlation ($r = 0.34$, $p = 0.002$), indicating that students with higher objective competence tended to report higher confidence, and vice versa. This association was statistically significant at the $\alpha = 0.05$ level. However, the coefficient of determination ($r^2 = 0.12$) indicates that competence accounts for only approximately 12% of the variance in confidence scores, underscoring that the majority of variation in students' self-assessed ability was not explained by their actual performance. This modest correlation suggests that, while there is a general directional alignment between competence and confidence, the relationship is far from perfect, and that many students held inaccurate metacognitive beliefs about their graph literacy skills. Such findings highlight the importance of integrating structured self-assessment activities into health science curricula to improve students' metacognitive calibration and support more accurate monitoring of their own graph literacy development.

Discussion

Graph literacy competence of indonesian health science students

The first research question asked about the level of graph literacy competence among undergraduate health science students at a public Indonesian university. The mean score of 5.31 points (SD = 2.19) out of 13 indicates a relatively low level of graph literacy competence. This finding is considerably below the scores reported by Galesic and Garcia-Retamero (2011) for German and United States general population samples, which averaged 9.4 and 9.3 points respectively, and is also far below the 11.42 points reported among German medical students by Soler Wenglein et al. (2025). The gap is substantial and cannot be easily attributed to differences in instrument design, since all three studies used the same 13-item Graph Literacy Scale. Several explanations may account for this gap, and a single cause is unlikely. One contributing factor may be differences in secondary mathematics and statistics curricula. Indonesian secondary education places comparatively less emphasis on data interpretation and graphical reasoning than systems in Germany or the United States, which may limit students' prior exposure to the conventions of statistical visualization. Digital learning experiences and visual data environments in Indonesian schools may also differ substantially from those of Western peers, providing less incidental exposure to graphical information. Bustang (2022) previously noted that Indonesian students' mathematical reasoning patterns, including their engagement with data-based tasks, can differ significantly from those observed in Western educational contexts. Furthermore, Bustang (2026) argued that health numeracy, which encompasses the ability to interpret graphical health data, remains underdeveloped in Indonesian health science education more broadly. The low competence scores in this study appear to support those arguments, while also highlighting that the problem is likely systemic rather than confined to a single cause.

The relationship between confidence and competence: Calibration

The second research question examined the relationship between students' confidence and their actual graph literacy performance, and specifically how well calibrated their self-assessments were. The confidence data revealed that students maintained consistently low confidence across all three categories, with mean confidence levels of 40.87%, 38.53%, and 36.64% for "reading the data", "reading between the data", and "reading beyond the data", respectively. These low confidence levels, viewed alongside the correspondingly low performance scores, suggest a pattern of broadly accurate self-awareness. This finding differs from what has been reported in some related studies. For instance, Knof et al. (2024) observed a pronounced Dunning-Kruger effect among medical students, where low academic performance was strongly correlated with significant overestimation of ability. Bustang et al. (2026), in their investigation of health numeracy among Indonesian nutrition students, found a significant overconfidence gap, where students maintained high confidence despite poor performance. The pattern in the present study is somewhat different: students appeared to sense their own limitations, and their confidence largely moved in the same direction as their performance across the three categories. However, the degree of calibration alignment varied by category and by item. Certain items, particularly those at the "reading beyond the data" level, showed wide variation in performance, with correct response rates ranging from 8% to 67%, yet confidence across this category remained uniformly low. This suggests that students were not always discriminating between items they could and could not answer, but rather applying a general low-confidence stance across the board. Schraw (2009) described this kind of non-discriminating self-assessment as a sign of underdeveloped metacognitive monitoring, where learners lack the ability to accurately judge their competence at the item level. Kruger and Dunning (1999) similarly noted that limited domain knowledge often prevents individuals from distinguishing between easier and harder tasks within the same domain. These theoretical perspectives help explain why confidence remained uniformly low even on items where performance was relatively strong.

The finding of broadly adequate group-level calibration, with a slight directional tendency toward under confidence, warrants theoretical interpretation. Stankov and Lee (2014) demonstrated in a cross-cultural study across 33 nations that overconfidence is particularly pronounced in populations with lower cognitive performance on the tasks assessed, as a form of unconscious self-protective mechanism. However, their data also showed that some populations, particularly those from East and South-East Asian educational contexts, showed more conservative self-assessments. Morony et al. (2013) similarly found that students from Confucian Asian countries, including those in South-East Asia, tended to have lower self-concept in mathematically oriented tasks despite sometimes performing comparably to or above their Western peers. Although Indonesia is not typically classified as a Confucian society, it shares certain educational and cultural values, including respect for authority, collective modesty, and an educational culture that discourages public expressions of confidence or self-promotion. These cultural norms may promote more conservative, restrained self-assessments, even in contexts where students do not perceive themselves as highly competent. The calibration pattern in this study, where most students (55.1%) were underconfident rather than overconfident, may therefore reflect not only limited metacognitive skill but also culturally shaped response tendencies. This interpretation is speculative, and future research should directly investigate the role of cultural factors in confidence calibration among Indonesian health science students.

Specific areas of difficulty in graph literacy

The third research question asked which specific areas of graph literacy presented the greatest challenges for students. The item-level findings reveal a pattern that is more complex than a simple hierarchy of difficulty. While performance was generally weakest at the highest level of graph comprehension, some foundational "reading the data" items, specifically Q1 and Q3, produced correct response rates of only 23% and 15%, respectively. These items required students to read the height of a bar from a fully labelled bar chart and to identify a proportion from a pie chart, tasks that Galesic and Garcia-Retamero (2011) reported as being answered correctly by over 80% of both German and United States general population participants. The very poor performance on these basic items in the Indonesian context is particularly concerning. Friel et al. (2001) argued that the ability to read the data is a prerequisite for higher-order graph comprehension, so fundamental weaknesses at this level are likely to compound difficulties at the more advanced levels. Within the "reading beyond the data" category, Q11, which required students to recognize that a meaningful comparison between two graphs was impossible due to unlabeled axes, was answered correctly by only 8% of participants. This item specifically tests the ability to tolerate ambiguity and resist drawing conclusions from incomplete information. Soler Wenglein et al. (2025) reported a similarly low correct response rate for Q11 among German medical students, and they interpreted this weakness as reflecting a tendency toward base-rate neglect and low ambiguity tolerance. The finding that Indonesian health science students performed even more poorly on this item than Western medical students and general population samples suggests that critical data appraisal skills are a particularly urgent area for curricular development. Durand et al. (2020) similarly found that graph literacy was the strongest predictor of health information comprehension in their study, reinforcing the practical relevance of addressing these specific skill gaps in health science education.

Implications and future directions

The findings of this study collectively point to the need for deliberate and targeted integration of graph literacy instruction into Indonesian health science curricula. The combination of low competence, broadly low confidence, and specific weaknesses at both the foundational and advanced levels of graph comprehension suggests that the problem is not confined to a single skill area. Rather, it reflects a more systemic gap in students' preparation to engage with graphical health information. Rothman et al. (2008) and Bustang (2026) both emphasised that the ability to interpret quantitative and graphical health data is inseparable from safe and effective clinical decision-making, which gives these findings direct practical significance beyond the academic context. Gaissmaier et al. (2012) further demonstrated that individuals with low graph literacy are likely to misinterpret even well-designed visual health information, suggesting that simply improving the quality of health communication materials is insufficient without also building the literacy skills of the intended audience.

Concretely, curricula should incorporate explicit graph interpretation training as a structured component of health statistics and research methods courses, rather than assuming students will develop these skills incidentally (Gardner et al., 2024). Activities might include guided analysis of real-world health graphs drawn from clinical guidelines and epidemiological reports, with structured discussion of how to read, compare, and critically evaluate graphical data across the three comprehension levels identified by Friel et al. (2001). Visual data reasoning tasks, where students are required to interpret ambiguous or incomplete graphs and justify their conclusions, would specifically address the critical weakness observed in items such as Q11, where students struggled to recognise the limits of available information. Formative assessment with immediate and targeted feedback has been shown to support the development of both competence and metacognitive accuracy (Schraw, 2009), and could be embedded within weekly coursework rather than reserved for summative examinations to optimize continuous learning in integrated health science curricula (Pai, 2025).

This study also makes a broader theoretical contribution by demonstrating the value of integrating graph literacy assessment with metacognitive calibration measurement in health professional education. Most existing studies have examined graph literacy competence in isolation, treating performance scores as the sole indicator of students' readiness to engage with visual health data. By simultaneously assessing confidence and computing calibration scores, this study reveals that the relationship between what students know and what they think they know is an important and independent dimension of health literacy preparation. The integration of these two constructs advances current understanding of quantitative literacy in health education by showing that calibration patterns, not just competence levels, vary across educational contexts and may reflect culturally shaped response tendencies. Future research should test whether targeted instructional interventions, including calibration feedback exercises, can improve both competence and metacognitive accuracy among Indonesian health science students. Longitudinal and comparative studies across multiple institutions would also strengthen the generalizability of these findings.

Conclusion

This study set out to examine the graph literacy competence and confidence of undergraduate health science students at a public Indonesian university, using the validated 13-item Graph Literacy Scale developed by Galesic and Garcia-Retamero (2011). The findings indicate that these students demonstrated a notably low level of graph literacy competence, performing substantially below both Western general population benchmarks and comparable health science student samples reported in the international literature. Confidence levels were consistently low across all three categories of graph comprehension, and the alignment between confidence and performance suggested that students had a general sense of their own limitations, though their self-assessments were not sufficiently discriminating at the item level to reflect genuine metacognitive monitoring. Item-level analysis further revealed that challenges were not confined to the most cognitively demanding tasks, but also extended to foundational skills such as reading values directly from bar charts and pie charts. These findings collectively suggest that graph literacy, as both a competence and a metacognitive skill, has not yet been adequately developed within the health science curriculum in the Indonesian context studied here.

The theoretical contribution of this study lies in demonstrating that simultaneously measuring graph literacy competence and metacognitive calibration provides a completer and more nuanced picture of students' quantitative literacy than performance scores alone. This integrated approach extends existing international research by applying it in a non-Western educational context that has been largely absent from the literature, and by revealing that calibration patterns may be shaped by cultural factors that differ meaningfully from those documented in Western populations. Practically, the findings underscore the need for explicit graph literacy instruction, including visual data reasoning activities and calibration feedback mechanisms, to be integrated into Indonesian health science curricula. Several limitations should be acknowledged. The convenience sample was drawn from a single faculty at one public university, and the predominantly female composition limits generalizability. The cross-sectional design precludes conclusions about how graph literacy develops over students' undergraduate years. Future research should employ larger and more diverse samples across multiple institutions, longitudinal designs, and intervention studies to test the effectiveness of targeted instructional approaches in building both competence and metacognitive calibration.

Acknowledgment

The authors express their sincere gratitude to the students of the Faculty of Sport and Health Sciences at Universitas Negeri Makassar for their valuable participation in this study.

Declarations

Conflicts of Interest	:	The authors declare no conflict of interest.
Generative AI Statement	:	Generative AI tools (Claude) were employed solely for language editing and minor phrasing enhancements. All conceptualization, analysis, and scholarly content were independently developed and verified by the authors.
Funding Statement	:	This work was supported by an internal research grant (PNBP) from Universitas Negeri Makassar under contract number SP DIPA - 023.17.2.677523/2023.

Author Contributions : **Bustang Bustang:** Conceptualization, methodology, funding acquisition, formal analysis, writing - original draft, editing, and visualization; **Garry Kuan:** Writing - review & editing, and Validation.

References

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Bodén, U., & Stenliden, L. (2019). Emerging visual literacy through enactments by visual analytics and students. *Designs for Learning*, *11*(1), 40–51.

BPS-Statistics Indonesia. (2025). *Profil statistik kesehatan 2025 [Health statistics profile 2025]*. BPS-Statistics Indonesia.

Bustang, B. (2022). *Probability conceptions and metacognitive judgements of Indonesian secondary school students and in-service mathematics teachers* [PhD Thesis, Loughborough University]. https://doi.org/10.26174/THESIS.LBORO.19625229

Bustang, B. (2026). Konseptualisasi peran numerasi kesehatan dalam pengambilan keputusan klinis: Kajian teoritis integrasi matematika dalam domain kesehatan [Conceptualizing the role of health numeracy in clinical decision-making: A theoretical review of mathematics integration in the healthcare domain]. In *Kesehatan Longevity dalam Peningkatan Kualitas Hidup* (pp. 165–171). Akademia Pustaka.

Bustang, B., Upu, H., & Kuan, G. (2026). An investigation of health numeracy competence and confidence among nutrition students at an Indonesian university. *BIO Web of Conferences*, *217*, 03004. https://doi.org/10.1051/bioconf/202621703004

Danya, H., Yonekura, Y., & Nakayama, K. (2021). Effects of graphic presentation on understanding medical risks and benefits among Japanese adults. *Cogent Medicine*, *8*(1), 1907894. https://doi.org/10.1080/2331205X.2021.1907894

Degroot, A. M. B., Dannenburg, L., & Vanhell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*(5), 600–629. https://doi.org/10.1006/jmla.1994.1029

Durand, M.-A., Yen, R. W., O'Malley, J., Elwyn, G., & Mancini, J. (2020). Graph literacy matters: Examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population. *PLoS ONE*, *15*(11), e0241844. https://doi.org/10.1371/journal.pone.0241844

Foster, C. (2016). Confidence and competence with mathematical procedures. *Educational Studies in Mathematics*, *91*(2), 271–288. https://doi.org/10.1007/s10649-015-9660-9

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. https://doi.org/10.2307/749671

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A.-S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, *31*(3), 286–296. https://doi.org/10.1037/a0024850

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989X10373805

Gardner, S. M., Angra, A., Harsh, J. A., & Wilson, K. J. (2024). Supporting student competencies in graph reading, interpretation, construction, and evaluation. *CBE—Life Sciences Education*, *23*(1), fe1. https://doi.org/10.1187/cbe.22-10-0207

Gutierrez de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology*, *114*(4), 681–700. https://doi.org/10.1037/edu0000674

Knof, H., Berndt, M., & Shiozawa, T. (2024). Prevalence of Dunning-Kruger effect in first semester medical students: A correlational study of self-assessment and actual academic performance. *BMC Medical Education*, *24*(1), 1210. https://doi.org/10.1186/s12909-024-06121-7

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, *58*, 79–96. https://doi.org/10.1016/j.ijer.2012.11.002

Pai, G. (2025). Using formative assessment and feedback from Student Response Systems (SRS) to revise statistics instruction and promote student growth for all. *Journal of Statistics and Data Science Education*, *33*(1), 16–25. https://doi.org/10.1080/26939169.2024.2321241

Procter-Legg, T., Klassen, R. M., Snell, R. J. S., Stevenson, N., & Thompson, I. (2026). Conceptions of visual literacy in school-age education: A systematic review. *Journal of Visual Literacy*, *45*(2), 137–175. https://doi.org/10.1080/1051144X.2026.2680316

Rahmani, M. (2020). Medical trainees and the Dunning–Kruger Effect: When they don't know what they don't know. *Journal of Graduate Medical Education*, *12*(5), 532–534. https://doi.org/10.4300/JGME-D-20-00134.1

Reading Turchioe, M., & Mangal, S. (2024). Health literacy, numeracy, graph literacy, and digital literacy: An overview of definitions, evaluation methods, and best practices. *European Journal of Cardiovascular Nursing*, *23*(4), 423–428. https://doi.org/10.1093/eurjcn/zvad085

Roth, W.-M. (2002). Reading graphs: Contributions to an integrative concept of literacy. *Journal of Curriculum Studies*, *34*(1), 1–24. https://doi.org/10.1080/00220270110068885

Rothman, R. L., Montori, V. M., Cherrington, A., & Pignone, M. P. (2008). Perspective: The role of numeracy in health care. *Journal of Health Communication*, *13*(6), 583–595. https://doi.org/10.1080/10810730802281791

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Soler Wenglein, J., Heidenreich, A., & Friederichs, H. (2025). Assessment of graph literacy among German medical students– an observational cross-sectional survey study. *BMC Medical Education*, *25*(1), 612. https://doi.org/10.1186/s12909-025-07206-7

Stankov, L., & Lee, J. (2014). Overconfidence Across World Regions. *Journal of Cross-Cultural Psychology*, *45*(5), 821–837. https://doi.org/10.1177/0022022114527345

Valcheff, K. L., Pennington, D. S., Keaton, S. K., & Carter, K. F. (2026). Integrating the essentials core competencies related to health literacy into undergraduate curriculum: Tapping traditional and emerging education strategies. *Public Health Nursing*, *43*(2), 402–409. https://doi.org/10.1111/phn.70041

van Strien-Knippenberg, I. S., Timmermans, D. R. M., Weber, M., Okan, Y., van Gils, C. H., Broeders, M. J. M., & Damman, O. C. (2025). Assessing the impact of visualizations of quantitative and qualitative risk information on the comprehension of hypothetical risk-based breast cancer screening results. *Health Communication*, 1–16. https://doi.org/10.1080/10410236.2025.2582779