International Journal of Contextual Science Education

e-ISSN. 3046-8876
Vol. 4, No. 2, June 2026
pp. 72-84

https://jurnalpasca.unram.ac.id/index.php/ijcse/index

Initial Psychometric Assessment of a Multidimensional Concept Test Addressing Ethnomedicine, Phytochemistry, and DPPH–IC_{50} Antioxidant Assay Analysis

Faizul Bayani[*1,2], Joni Rokhmat[1,3], Aliefman Hakim[1,4], AA Sukarso[1,5]

[1]Doctoral Program in Natural Science Education, Graduate School, Mataram University, Indonesia
[2]Bachelor's Program in Natural Science Education, Qamarul Huda Badaruddin Bagu University, Indonesia
[3]Bachelor's Program in Physics Education, Faculty of Teacher Training and Education, Mataram University, Indonesia
[4]Bachelor's Program in Chemistry Education, Faculty of Teacher Training and Education, Mataram University, Indonesia
[5]Bachelor's Program in Biology Education, Faculty of Teacher Training and Education, Mataram University, Indonesia

*Corresponding author e-mail: faizulbayani0@gmail.com

Accepted: May 26th 2026, Approved: June 6th 2026, Published: June 7th 2026

Abstract— Multidisciplinary learning in ethnomedicine and ethics, phytochemistry, and antioxidant bioassay analysis requires students to integrate ethical reasoning, chemical concepts, and quantitative data interpretation. This study conducted an initial psychometric evaluation of a multidomain concept test and described students' pre-intervention baseline performance across these three domains. The assessment was administered to fifth-semester undergraduate pharmacy students enrolled in the Phytochemistry course at the Faculty of Health and Science, UNIQHBA Bagu. A 28-item mixed-format instrument, consisting of multiple-choice, constructed-response, and numerical items, was administered to 41 students, with a maximum possible score of 32 points. Descriptive statistics and Classical Test Theory indices were calculated, including Cronbach's alpha, item difficulty, D27 discrimination, and corrected item–total correlations. The mean total score was 19.90 out of 32, equivalent to 62.20% ± 13.05%, with a median of 20 and a score range of 11–29. Only 9 students, or approximately 22.0%, achieved the ≥70% criterion. Domain-level performance was uneven. Ethnomedicine and ethics reached 7.88 ± 1.25, or 78.80% of the domain maximum; phytochemistry reached 7.29 ± 2.15, or 72.90%; and DPPH–IC50 interpretation reached 4.73 ± 2.54, or 39.40%. The instrument showed acceptable internal consistency, with Cronbach's alpha of 0.756. However, item diagnostics revealed floor and ceiling effects, as well as several challenging DPPH–IC50 items. These findings indicate that the instrument is suitable for preliminary diagnostic profiling and that DPPH–IC50 quantitative interpretation is the main baseline learning gap. The study provides initial evidence for targeted instructional scaffolding and systematic item refinement to improve diagnostic differentiation across the student ability spectrum.

Keywords— Ethnomedicine, Phytochemistry, DPPH assay, IC50, Concept test.
How to Cite— Bayani, F., Rokhmat, J., Hakim, A., & Sukarso, AA. (2026). Initial Psychometric Assessment of a Multidimensional Concept Test Addressing Ethnomedicine, Phytochemistry, and DPPH–IC_{50} Antioxidant Assay Analysis. *International Journal of Contextual Science Education*, *4*(2), 72-84. https://doi.org/10.29303/ijcse.v4i2.1617

1. Introduction

Background and rationale

The integration of ethnomedicine, phytochemistry, and antioxidant bioassay analysis is increasingly relevant in undergraduate pharmacy and science education. Students are expected not only to understand community-based medicinal knowledge and its ethical implications, but also to connect this knowledge with chemical principles, secondary metabolites, extraction procedures, and quantitative interpretation of experimental data. These domains require different but complementary competencies, including ethical reasoning, conceptual understanding of phytochemical principles, and the ability to interpret dose–response relationships in antioxidant assays.

Concept tests and concept inventories are widely used to identify misconceptions, describe baseline understanding, and support targeted instruction. From a constructivist perspective, learning requires students to connect new information with prior

knowledge and revise inaccurate conceptions through active engagement [1]. Concept inventories can help reveal conceptual errors and guide instructional improvement [2]. Similarly, authentic research-based activities can strengthen engagement and deepen conceptual understanding in science learning contexts [3]. These perspectives support the development of multidomain assessments that evaluate student performance across related areas rather than focusing on a single content domain.

In phytochemistry education, students often experience persistent misconceptions related to secondary metabolites, extraction logic, and structure–activity relationships [4][5]. They may also struggle to connect chemical representations and laboratory procedures with biological or health-related contexts [6]. The inclusion of ethical reasoning adds another layer of complexity because students must evaluate the use of traditional knowledge, community participation, informed consent, and benefit-sharing within scientific inquiry [7]. Therefore, assessment in this field should capture not only factual knowledge, but also reasoning across scientific, ethical, and contextual dimensions.

Quantitative literacy is also essential for interpreting antioxidant assays such as DPPH. Students may make errors when calculating percentage inhibition, converting concentrations, or interpreting the relationship between concentration and antioxidant activity [8]. They may also misunderstand IC50 as a comparative index, although it represents the concentration required to inhibit 50% of DPPH radicals [9]. In antioxidant analysis, measured activity generally changes with concentration, but students may not fully understand how IC50 values are used to compare antioxidant capacity across samples [10]. Educational research further emphasizes that quantitative literacy in laboratory learning involves interpreting dose–response curves, applying interpolation or regression, and reasoning under uncertainty [11][12][13]. Formative assessment is therefore useful for identifying errors in data interpretation and providing feedback that supports deeper learning [14][15].

Ethnomedicine and ethical competencies are increasingly emphasized in science and health education. These competencies include the ability to understand informed consent, community engagement, cultural sensitivity, benefit-sharing, and the responsible use of shared knowledge. Scott describes the integration of ethical, legal, and social implications into life-science curricula as a way to strengthen students' understanding of research responsibility [16]. Goodwin also highlight the importance of community involvement and explicit discussion of cultural dynamics in health-related contexts [17]. In addition, research on ethics in educational settings shows the importance of clearly defining ethical responsibilities when participants and community knowledge are involved [18]. These issues indicate the need for assessment tools that are aligned with domain-specific ethical and scientific competencies.

Research problem and study purpose

Although ethnomedicine, phytochemistry, and DPPH–IC50 analysis are closely connected in pharmacy and science education, courses often proceed without a clear baseline profile of students' initial competencies. In addition, few assessments provide preliminary psychometric evidence showing whether an instrument is suitable for diagnostic profiling across these domains. Diagnostic pretests are useful for identifying baseline knowledge and locating specific learning gaps [19]. They can also support targeted remediation and help instructors adjust learning activities according to students' needs [20][21]. The Force Concept Inventory provides an example of how pre-instruction concept assessment can reveal persistent misconceptions and inform teaching decisions [22].

The central problem addressed in this study is whether students demonstrate adequate foundational proficiency in ethnomedicine and ethics, phytochemistry, and DPPH–IC50 interpretation, and whether a multidomain concept test provides sufficient preliminary measurement evidence for baseline diagnostic use. This focus is important because these three domains represent different forms of reasoning. Ethnomedicine and ethics require contextual and normative reasoning, phytochemistry requires conceptual understanding of chemical principles, and DPPH–IC50 interpretation requires quantitative reasoning and data interpretation.

This study was conducted in an authentic course context involving fifth-semester undergraduate pharmacy students enrolled in the Phytochemistry course at FKES UNIQHBA Bagu. The study provides an initial psychometric evaluation of a multidomain concept test designed to assess conceptual and quantitative reasoning across ethnomedicine, ethics, phytochemistry, and DPPH–IC50 antioxidant assay analysis. The instrument consisted of 28 items in mixed formats, including multiple-choice, constructed-response, and numerical items, with a maximum score of 32 points.

Thesis and contribution

This study positions the multidomain concept test as a preliminary diagnostic instrument rather than as a final validated measure. Its contribution is twofold. First, it provides baseline evidence about students' strengths and weaknesses across conceptually connected but cognitively different domains. Second, it identifies item-level priorities for instrument refinement, especially items showing extreme difficulty, extreme easiness, or limited discrimination.

By combining domain-level performance, internal consistency, item difficulty, discrimination, and corrected item–total correlations, this study offers an empirical basis for improving both instruction and assessment. Instructionally, the findings can guide targeted scaffolding in areas that require quantitative interpretation, particularly DPPH–IC50 calculation, curve reading, and IC50 reasoning. Psychometrically, the findings can guide item revision so that the instrument provides better diagnostic information across the student ability spectrum. Future validation using posttest data, larger samples, expert judgment, response-process evidence, and structural analyses will be needed to strengthen the interpretation and use of the instrument.

2. Method

Study design and participants

This study used an exploratory quantitative design to conduct an initial psychometric assessment of a multidomain concept test. The instrument was designed for diagnostic use before the development and implementation of instructional materials. The

assessment focused on students' baseline conceptual and quantitative understanding across three related domains: ethnomedicine and ethics, phytochemistry, and DPPH–IC50 antioxidant assay interpretation.

The concept test was administered to fifth-semester undergraduate pharmacy students at the Faculty of Health and Science (FKES), UNIQHBA Bagu, during the Phytochemistry course. The analyzed sample consisted of 41 students. The study examined total scores, domain subscores, item difficulty, item discrimination, and corrected item–total correlations to evaluate the preliminary diagnostic usefulness of the instrument.

Instrument development, domains, and cognitive mapping

The multidomain concept test consisted of 28 items with a maximum possible score of 32 points. The content was organized into three domains. The Ethnomedicine and Ethics domain measured students' understanding of community-based medicinal knowledge, informed consent, benefit-sharing, and ethical reasoning. The Phytochemistry domain measured students' understanding of secondary metabolites, extraction principles, and structure–activity relationships. The DPPH–IC50 domain measured students' ability to interpret antioxidant assay data, including percentage inhibition, concentration–response relationships, and IC50 reasoning. The maximum score was distributed across the three domains as follows: Ethnomedicine and Ethics = 10 points, Phytochemistry = 10 points, and DPPH–IC50 antioxidant assay interpretation = 12 points. This distribution reflected the multidomain structure of the instrument and the need to assess both conceptual and quantitative reasoning.

Each item was mapped to a cognitive level based on Bloom's taxonomy, ranging from C1 to C6. This mapping was used to check whether item demands aligned with the intended construct and to support interpretation of item difficulty. Lower-level items assessed recall and comprehension, whereas higher-level items assessed application, analysis, evaluation, or construction of explanations. The item diagnostic tables report the cognitive level of each item to make this alignment transparent.

Item formats and scoring procedure

The instrument used a mixed-format design to capture different forms of student reasoning. Multiple-choice items assessed conceptual recognition and discrimination among alternatives. Numerical items assessed calculation and quantitative interpretation. Constructed-response items assessed students' ability to explain, justify, or integrate concepts in written form.

Most multiple-choice and numerical items were scored from 0 to 1 point. Constructed-response items were scored from 0 to 2 points. A score of 0 indicated an incorrect, irrelevant, or missing response. A score of 1 indicated a partially correct response that showed limited reasoning or incomplete use of relevant concepts. A score of 2 indicated a complete and accurate response that demonstrated appropriate conceptual understanding and reasoning. This scoring approach follows the recommendation that mixed-format assessments should use explicit rubrics to support scoring consistency [23]. Partial-credit scoring was used because it can capture different levels of understanding in constructed-response and numerical tasks and provide more useful diagnostic information than dichotomous scoring alone [24].

Because this study represents an initial psychometric assessment based on PRE data, the analysis focused on internal score consistency and item-level diagnostics. Inter-rater reliability was not used as a primary source of evidence in the present analysis. However, because constructed-response items were included, future validation should examine scoring consistency using inter-rater agreement indices such as percentage agreement or Cohen's kappa [25].

Data preparation and quality checks

Before analysis, all item scores and total score fields were converted into numeric format. The dataset was checked for missing values and irregular entries. Total scores were recalculated from item-level scores and compared with the maximum possible score of 32 points to ensure scoring accuracy. Domain subscores were also computed according to the predefined score allocation for Ethnomedicine and Ethics, Phytochemistry, and DPPH–IC50.

Statistical analysis and psychometric evaluation

Baseline performance was summarized using descriptive statistics, including mean, standard deviation, median, minimum, maximum, and percentage of the maximum score. Total score distribution was examined to describe the spread of student performance. Domain-level achievement was calculated by dividing the mean score of each domain by its maximum possible score and expressing the result as a percentage. The psychometric evaluation followed Classical Test Theory. Internal consistency of the 28-item instrument was estimated using Cronbach's alpha. Reporting reliability is a common practice in initial instrument evaluation [26], and coefficients above 0.70 are commonly considered acceptable for preliminary educational measurement purposes [27]. However, alpha was interpreted cautiously because the instrument included multiple domains and mixed item formats.

Item difficulty was calculated as the proportion of students who answered each item correctly and was expressed as a percentage. Items with difficulty values between 0.30 and 0.70 were considered more informative for diagnostic purposes because they reduce extreme floor or ceiling effects while still differentiating among students [23]. Items with very low p values were interpreted as potentially too difficult, whereas items with very high p values were interpreted as potentially too easy.

Item discrimination was evaluated using the D27 index, which compares performance between the upper and lower 27% of students. Values above approximately 0.30 were interpreted as indicating useful discrimination between higher- and lower-performing students [23]. Corrected item–total correlations were also calculated to examine the alignment between each item and overall test performance after removing the item's own contribution from the total score.

To support diagnostic refinement, items were classified according to their difficulty and discrimination patterns. Items with extreme p values below 10% or above 90%, especially when combined with low discrimination, were identified as priorities for revision [28]. For multiple-choice items, distractor effectiveness analysis is recommended in future refinement to identify distractors that fail to represent plausible misconceptions [29]. Response-process evidence, such as think-aloud protocols or cognitive interviews, should also be used in later validation phases to examine how students interpret item prompts and construct their answers [30]. After item revision, reliability should be recalculated to determine whether the revised instrument maintains or improves its

psychometric quality [31].

Graphical diagnostics were used to support item-level interpretation. The analysis included a difficulty curve by item number, a discrimination curve by item number, and a joint diagnostic map plotting item difficulty against D27 discrimination. These visualizations were used to identify items that were too easy, too difficult, weakly discriminating, or diagnostically informative.

Considerations for subsequent validation phases

This study should be interpreted as an exploratory psychometric assessment based on a limited sample. Therefore, additional validity evidence is needed before the instrument can be used for stronger inferential purposes. Future studies should include expert review to strengthen content validity, response-process methods to examine how students understand and answer the items, and larger samples to support structural analyses.

For evaluating internal structure, exploratory factor analysis may be more appropriate than confirmatory approaches during early-stage validation, especially when sample size is limited [32]. Bootstrapping may also be used to produce more stable confidence intervals for estimated parameters in small-sample contexts [33]. If domain subscores are used for diagnostic interpretation, each domain must remain theoretically grounded, and correlations among domains should be interpreted carefully to avoid overfitting or unsupported subscale claims [34].

3. Result and Discussion

Results

Overall PRE performance and score distribution

The analyzed cohort consisted of 41 fifth-semester undergraduate pharmacy students enrolled in the Phytochemistry course at FKES UNIQHBA Bagu. All students completed the PRE concept test, which had a maximum score of 32 points. Table 1 summarizes the descriptive statistics of the total scores.

Table 1. Descriptive Statistics of Total Scores

N	Maximum Score (points)	Mean (points)	SD (points)	Median (points)	Min–Max (points)	Mean (%)	SD (%)
41	32	19.90	4.18	20.00	11–29	62.20	13.05

The mean total score was 19.90 out of 32, equivalent to 62.20% of the maximum score. The standard deviation was 4.18 points, or 13.05% in percentage-score terms. The median score was 20.00, and the observed score range was 11–29 points. These results indicate moderate baseline performance with substantial variation among students.

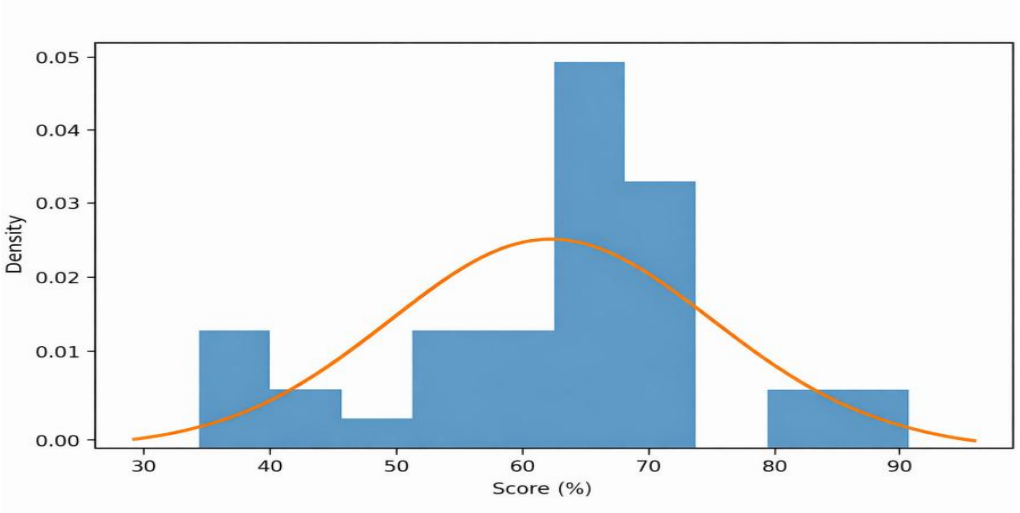

Figure 1. Histogram of score percentages (%) with a normal curve as a visual reference

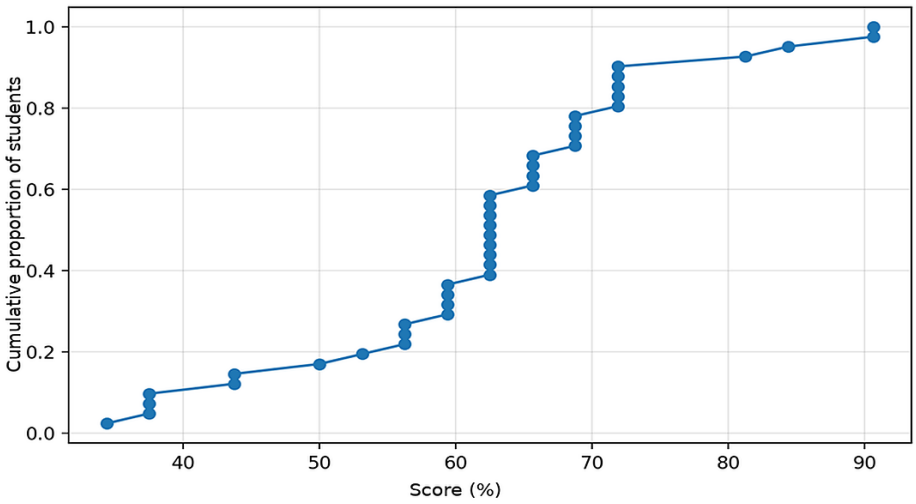

Figure 2. Empirical cumulative distribution function (ECDF) showing the proportion of students scoring below a given threshold.

Figures 1 and 2 show the distribution of PRE percentage scores. The histogram provides a visual overview of score spread, while the ECDF shows the cumulative proportion of students below each score threshold. Together, these diagnostics indicate that students entered the course with uneven levels of initial competence.

Mastery threshold profile

A mastery criterion of ≥70% was used to contextualize baseline achievement. Based on this criterion, 9 students, or approximately 22.0% of the cohort, reached the mastery threshold. The remaining 32 students, or approximately 78.0%, did not reach the threshold. This pattern shows that most students had not yet achieved the expected level of baseline competence before targeted instructional support.

The mastery threshold should be interpreted diagnostically rather than as a definitive indicator of competence. Because the study used an initial PRE design and the instrument still requires further validation, the threshold mainly serves as a practical reference for identifying students who may need additional support.

Domain-level performance and subscore profile

Table 2 summarizes mean attainment across the three domains, and Figure 3 presents the domain-level comparison as a percentage of the maximum score.

Table 2. Mean attainment by domain

Domain	Maximum Score	Mean (points)	SD	Mean (% of maximum)
Ethnomedicine / Ethics	10	7.88	1.25	78.80
Phytochemistry	10	7.29	2.15	72.90
DPPH–IC50	12	4.73	2.54	39.40

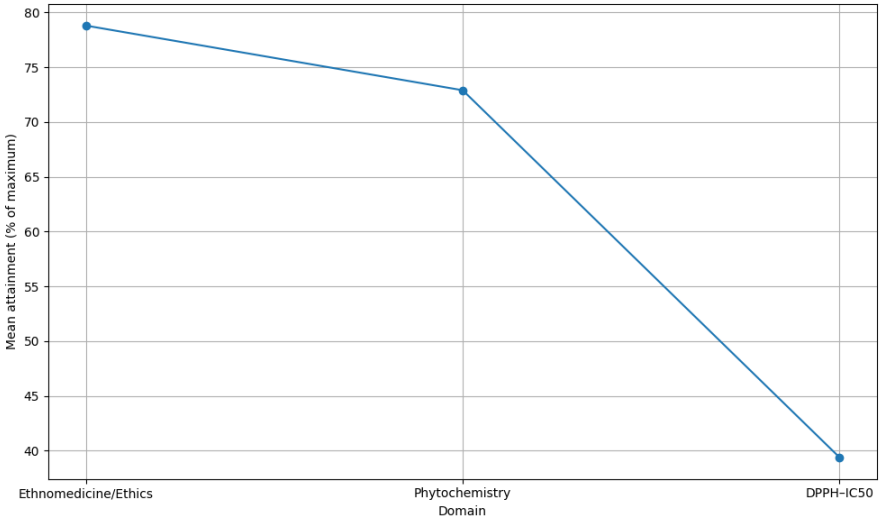

Figure 3. Curve of mean attainment by domain (% of maximum score)

The domain profile revealed a clear imbalance. Students performed relatively well in Ethnomedicine and Ethics, with a mean score of 7.88 out of 10, or 78.80% of the domain maximum. They also showed moderate-to-high performance in Phytochemistry, with a mean score of 7.29 out of 10, or 72.90%. In contrast, performance in DPPH–IC50 interpretation was

considerably lower, with a mean score of 4.73 out of 12, or 39.40%.

This pattern identifies DPPH–IC50 interpretation as the main baseline learning gap. The relatively high standard deviation in this domain also suggests uneven quantitative reasoning skills among students.

Reliability of the multidomain concept test

The internal consistency of the 28-item instrument was acceptable for preliminary diagnostic use, with Cronbach's α = 0.756. This value indicates that the instrument had sufficient internal consistency for initial baseline profiling. However, because the instrument contains multiple domains and mixed item formats, this coefficient should not be interpreted as complete validity evidence. Further reliability estimates at the domain level, as well as complementary indices such as McDonald's omega and the standard error of measurement, should be considered in future studies.

Item difficulty and discrimination diagnostics

Item-level diagnostics were calculated using Classical Test Theory. Item difficulty was expressed as the proportion of correct responses, or p value. Item discrimination was examined using the D27 index, which compares the upper and lower 27% of students. Corrected item–total correlations were also calculated to assess the alignment of each item with overall test performance.

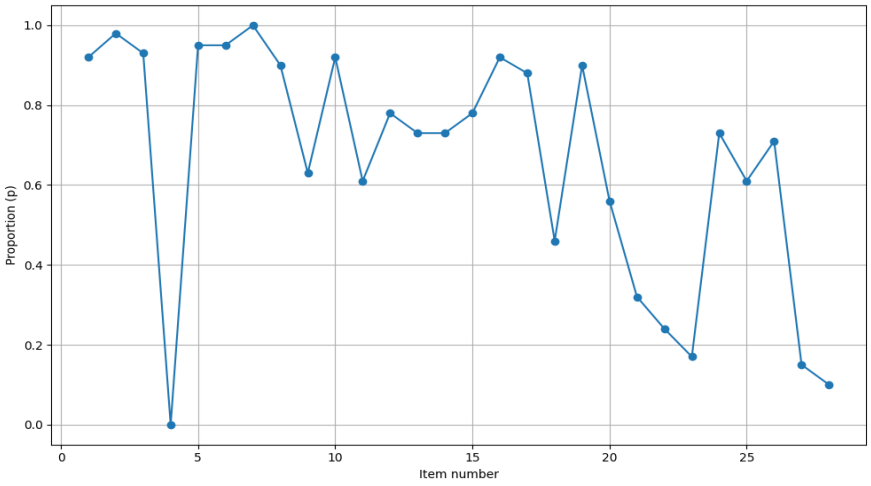

Figure 4. Item difficulty curve (p) by item number

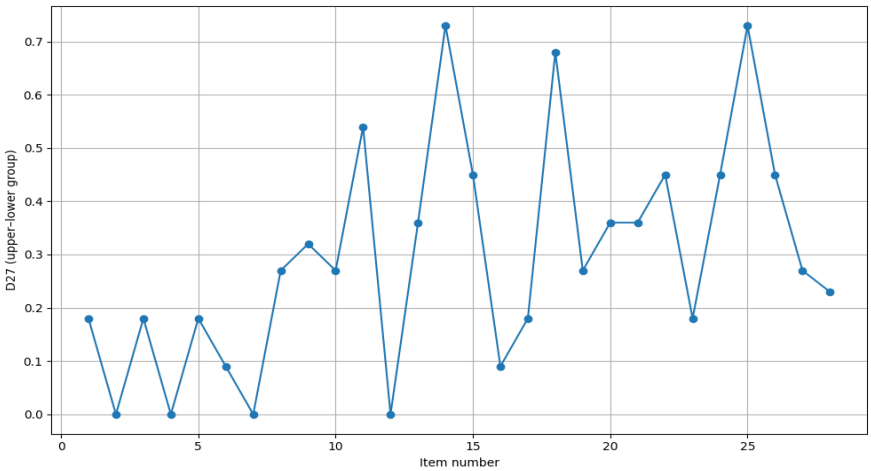

Figure 5. D27 discrimination curve by item number

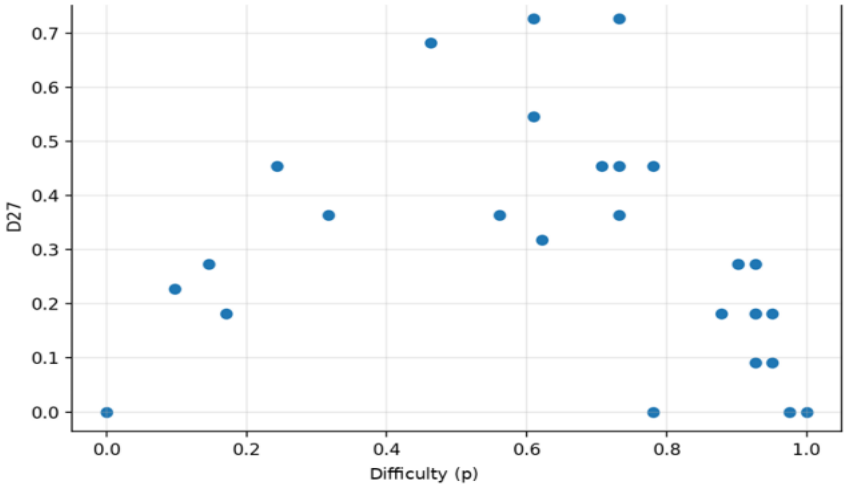

Figure 6. Item diagnostic map: difficulty (p) versus discrimination (D27)

Figures 4–6 provide item-level diagnostic information. Figure 4 shows the difficulty pattern across the 28 items, Figure 5 shows discrimination values across items, and Figure 6 combines difficulty and discrimination to identify items that are diagnostically useful, too easy, too difficult, or weakly discriminating.

Most difficult items

Table 3 shows the eight most difficult items based on the lowest p values.

Table 3. Eight most difficult items (lowest p values, %)

Item	Domain	Cognitive Level	Format	Max Points	Proportion Correct (p, %)	D27	Corrected Item–Total Correlation
4	Ethnomedicine / Ethics	C4	MCQ	1	0.00	0.00	NaN
28	DPPH–IC50	C6	CR	2	9.80	0.23	0.37
27	DPPH–IC50	C5	CR	2	14.60	0.27	0.43
23	DPPH–IC50	C3	NUM	1	17.10	0.18	0.29
22	DPPH–IC50	C3	NUM2	1	24.40	0.45	0.43
21	DPPH–IC50	C3	NUM	1	31.70	0.36	0.28
18	Phytochemistry	C6	CR	2	46.30	0.68	0.45
20	DPPH–IC50	C2	MCQ	1	56.10	0.36	0.34

The most extreme result was Item 4, which had p = 0.00%, D27 = 0.00, and a non-estimable corrected item–total correlation. This result suggests that the item should be reviewed for possible problems in wording, keying, content alignment, or cognitive demand.

Most difficult items were concentrated in the DPPH–IC50 domain. Items 28, 27, 23, 22, and 21 showed relatively low p values, indicating that students struggled with quantitative assay interpretation. However, not all difficult items were psychometrically weak. Item 22 showed useful discrimination, with D27 = 0.45 and corrected r = 0.43. Item 18 in the Phytochemistry domain also showed strong discrimination, with D27 = 0.68 and corrected r = 0.45. These items should not be removed automatically; instead, they should be reviewed for clarity while preserving their diagnostic value.

Easiest items

Table 4 shows the eight easiest items based on the highest p values.

Table 4. Eight easiest items (highest p values, %)

Item	Domain	Cognitive Level	Format	Max Points	Proportion Correct (p, %)	D27	Corrected Item–Total Correlation
7	Ethnomedicine / Ethics	C3	MCQ	1	100.00	0.00	NaN
2	Ethnomedicine / Ethics	C2	MCQ	1	97.60	0.00	−0.00
5	Ethnomedicine / Ethics	C2	MCQ	1	95.10	0.18	0.22
6	Ethnomedicine /	C4	MCQ	1	95.10	0.09	0.08

1	Ethnomedicine / Ethics	C1	MCQ	1	92.70	0.18	0.28
3	Ethnomedicine / Ethics	C3	MCQ	1	92.70	0.18	0.28
10	Phytochemistry	C1	MCQ	1	92.70	0.27	0.47
16	Phytochemistry	C3	MCQ	1	92.70	0.09	0.09

The easiest items were mainly located in the Ethnomedicine and Ethics domain. Items 7 and 2 showed strong ceiling effects, with p = 100.00% and p = 97.60%, respectively. Both items also had D27 = 0.00, indicating that they did not distinguish higher-performing students from lower-performing students. These items should be revised by increasing cognitive demand, improving distractor plausibility, or adding contextual complexity.

Some easy items still showed useful alignment with the overall score. For example, Item 10 had p = 92.70%, D27 = 0.27, and corrected r = 0.47. This item may be retained with minor revision, especially if it assesses an essential foundational concept.

Discussion

Overview of findings

This exploratory study provides an initial diagnostic profile of students' baseline competence and the preliminary psychometric quality of a multidomain concept test. The main finding is a clear imbalance across domains. Students showed relatively stronger performance in Ethnomedicine and Ethics and Phytochemistry, but substantially weaker performance in DPPH–IC50 interpretation.

The overall PRE result was moderate, with a mean score of 19.90 out of 32, or 62.20% ± 13.05%. The observed score range of 11–29 points indicates substantial variation in students' initial understanding. Similar variability is often found in concept-test and diagnostic-assessment contexts, where differences in prior knowledge, conceptual preparation, and learning experience influence pretest performance [35][36][37][38][39][40].

The use of a ≥70% mastery threshold provided a practical reference for identifying students who may need additional support. However, mastery thresholds must be interpreted carefully because score meaning depends on instrument alignment, difficulty distribution, and validation evidence [41][42][43][44][45][46]. In this study, the threshold functions as a diagnostic indicator rather than a final standard of competence.

Domain imbalance and the DPPH–IC50 learning gap

The domain-level results showed a clear hierarchy. Ethnomedicine and Ethics reached 78.80% of the maximum score, Phytochemistry reached 72.90%, and DPPH–IC50 reached only 39.40%. This pattern suggests that students were more prepared for concept-based and contextual domains than for quantitative assay interpretation. Domain subscores are useful because they can reveal specific strengths and weaknesses that may be hidden by a total score [47][48]. However, their interpretation requires caution. Subscores should be supported by item-level evidence, theoretical construct alignment, and internal-structure evidence before they are used for strong diagnostic decisions [47][49].

The low DPPH–IC50 score likely reflects the complex nature of antioxidant assay interpretation. Students must calculate percentage inhibition, understand concentration–response relationships, interpret curves, and reason about IC50 as a comparative index. These tasks require both conceptual understanding and procedural quantitative fluency. The current pattern is consistent with studies showing that students often perform better on knowledge-based tasks than on quantitative reasoning or curve-interpretation tasks [50]. Therefore, future instruction should strengthen graph interpretation, mathematical reasoning, and practice-based interpretation of experimental data [51][52].

Instrument quality and psychometric interpretation

The instrument showed acceptable internal consistency for preliminary diagnostic profiling, with Cronbach's $\alpha = 0.756$. This value supports the use of the test for early-stage mapping of student competence. However, alpha should be interpreted cautiously because the instrument is multidomain and includes mixed item formats. Cronbach's alpha is widely used in educational measurement, but it does not by itself establish validity or dimensionality [53]. Future studies should report additional reliability evidence. McDonald's omega may provide a more appropriate estimate when item covariance structures are complex [54][55][56]. The standard error of measurement can also clarify score precision. For multidomain instruments, subscale reliability should be examined so that domain-specific interpretations are not based only on a single total-score coefficient [57][58][59].

Item difficulty, discrimination, and revision priorities

The item diagnostics provide clear guidance for instrument refinement. Items with p values between 0.30 and 0.70 are generally considered more informative for diagnostic purposes, whereas items below 0.30 may be too difficult and items above 0.70 may be too easy [60][61]. Discrimination indices such as D27 and corrected item–total correlations help determine whether an item distinguishes higher-performing students from lower-performing students [62][63]. Several items require revision because they showed extreme difficulty or extreme easiness. Item 4 is the strongest candidate for immediate review because no student answered it correctly. This may indicate a scoring-key error, unclear wording, content misalignment, or excessive cognitive demand. In contrast, Items 7 and 2 were answered correctly by nearly all students but showed no discrimination. These items should be made more cognitively demanding or revised with more plausible distractors.

The DPPH–IC50 items require a more nuanced approach. Some were difficult but still discriminative. Item 22, for example, had p = 24.40%, D27 = 0.45, and corrected r = 0.43. This suggests that the item was challenging but useful for differentiating student

ability. Similarly, Item 18 in Phytochemistry had strong discrimination, with D27 = 0.68 and corrected r = 0.45. Such items should generally be retained, with refinements to wording or scoring criteria if needed. Difficult but discriminative items can remain valuable when they align with the intended construct and are supported by appropriate instruction [66]. Items that are too easy and weakly discriminating should be revised by increasing contextual complexity, strengthening distractors, or requiring more reasoning rather than recall [63][64]. Items that are too difficult and weakly discriminating should be simplified through clearer wording, reduced irrelevant cognitive load, intermediate prompts, or segmentation into smaller steps [65]. The goal is not to make all items easier, but to improve the balance between difficulty, discrimination, and construct alignment.

Pedagogical implications

The PRE results show that students need targeted support in DPPH–IC50 quantitative interpretation. Instruction should begin with foundational calculations, such as percentage inhibition and concentration conversion, before moving to curve reading, interpolation, and IC50 comparison. Worked examples can reduce cognitive load by showing step-by-step reasoning and calculation procedures [1]. Scaffolding can then guide students from simple calculations to more complex assay interpretation tasks [2] [3].

Explicit instruction in graph and curve interpretation is also needed. Visual representations of concentration–inhibition relationships can help students connect numerical data with biological meaning. This is important because students often struggle to connect chemical concepts, representations, and biological contexts [4][5][6]. Peer discussion and inquiry-based activities may further support reasoning because students must explain, compare, and justify their interpretations [7]. Regular formative assessment can also help identify misconceptions early and provide timely feedback [8].

Aligning future evaluation with diagnostic use

Because this study used only PRE data, it cannot support claims about instructional effectiveness. The results show baseline competence and item performance, not learning gains. Future studies should include posttest data and, when possible, comparison groups. Pre–post comparisons can show whether students improve in quantitative assay interpretation after targeted instruction [9]. Rubric-based performance tasks can provide richer evidence of procedural reasoning and conceptual accuracy than total scores alone [10]. Longitudinal tracking may also show whether students retain and transfer quantitative interpretation skills across laboratory contexts [11].

Balancing content coverage and psychometric quality

A multidomain concept test must balance content representativeness and psychometric quality. The instrument should include items across relevant domains and cognitive levels, but it should also avoid excessive clustering of items at very low or very high difficulty levels. Balanced cognitive demand is important so that the instrument captures both foundational knowledge and higher-order reasoning [67]. At the same time, items must discriminate adequately among students with different levels of competence [68]. The current findings suggest two main refinement needs. First, Ethnomedicine and Ethics items with strong ceiling effects should be revised to better differentiate students. Second, DPPH–IC50 items should be reviewed to ensure that they measure the intended quantitative reasoning skills rather than unnecessary procedural overload. These revisions can improve diagnostic information across the full ability spectrum.

Roadmap for further validity evidence

The present findings provide preliminary evidence, but a stronger validity argument is still needed. Future validation should begin with expert content review to confirm that items match the intended constructs, indicators, and cognitive levels [69]. Cognitive interviews or other response-process methods should then be used to examine how students understand prompts and construct their answers [70]. With larger samples, structural analyses such as exploratory factor analysis, Item Response Theory, or Rasch modeling can be used to examine dimensionality and item functioning beyond Classical Test Theory [71]. Known-groups validity testing can determine whether the instrument distinguishes students with different levels of preparation [72]. Differential item functioning analysis can also help evaluate fairness across student groups [73].

Limitations

This study is limited by its PRE-only design. The results describe baseline competence and preliminary item performance, but they do not demonstrate instructional effectiveness or learning gains. The sample was also drawn from a single cohort of fifth-semester undergraduate pharmacy students in one institutional and course context. Therefore, generalization to other programs, semesters, or institutions should be made cautiously. The instrument also requires further validation. Although Cronbach's alpha was acceptable, additional evidence is needed for content validity, response processes, internal structure, subscore reliability, and scoring consistency for constructed-response items. Future research should include larger samples, posttest data, expert review, and more complete psychometric analyses.

Concluding synthesis

The PRE results reveal a clear domain-specific gap in DPPH–IC50 interpretation. Students showed stronger baseline performance in Ethnomedicine and Ethics and Phytochemistry, but weaker performance in quantitative antioxidant assay interpretation. This finding supports the need for targeted scaffolding in percentage inhibition calculation, curve reading, and IC50 reasoning.

The multidomain concept test showed acceptable internal consistency for preliminary diagnostic profiling. However, item diagnostics revealed floor and ceiling effects that should be addressed through systematic item revision. Overall, the findings provide a useful empirical basis for improving both instruction and instrument development in multidomain phytochemistry-related assessment.

4. Conclusion

This initial pre-intervention assessment shows that students' foundational proficiency in ethnomedicine and ethics, phytochemistry, and DPPH–IC50 interpretation was uneven. Although the overall mean score reached 62.20% of the maximum score, most students did not meet the ≥70% mastery threshold. Domain-level results identified DPPH–IC50 interpretation as the main learning gap, with students achieving only 39.40% of the domain maximum. This finding indicates that students need stronger support in quantitative assay interpretation, particularly in percentage inhibition calculation, dose–response interpretation, and IC50 reasoning.

The mixed-format instrument demonstrated acceptable internal consistency for preliminary diagnostic use, with Cronbach's alpha of 0.756. However, item diagnostics revealed floor and ceiling effects, as well as variation in item discrimination. These findings suggest that the instrument is useful for initial baseline profiling but still requires refinement before broader diagnostic application. Future development should revise extreme items, rebalance item difficulty across domains, and strengthen the measurement of DPPH–IC50 quantitative reasoning. Further validation should also include expert review, response-process evidence, inter-rater reliability for constructed-response items, structural analysis, and larger samples with posttest or comparison-group data. Overall, this study provides preliminary empirical evidence for a multidomain diagnostic assessment framework that integrates ethnomedicine, phytochemistry, and antioxidant assay interpretation in pharmacy education. The findings can guide both targeted instructional scaffolding and systematic instrument refinement.

Acknowledgment

Place acknowledgments, including information on grants received, before the references, in a separate section, and not as a footnote on the title page.

References

[1] D. K. Cirit, "Global environmental problems based on common knowledge construction model: Evaluation of 'exploring and categorizing' stage," International Online Journal of Educational Sciences, vol. 12, no. 3, 2020, doi: https://doi.org/10.15345/iojes.2020.03.016.

[2] O. S. Anderson, "Development of a concept inventory to investigate student learning gains in life cycle nutrition," Journal of Nutritional Health & Food Engineering, vol. 7, no. 3, 2017, doi: https://doi.org/10.15406/jnhfe.2017.07.00241.

[3] E. A. Godin, S. V. Wormington, T. Perez, M. M. Barger, K. E. Snyder, L. S. Richman, R. D. Schwartz-Bloom, and L. Linnenbrink-Garcia, "A pharmacology-based enrichment program for undergraduates promotes interest in science," CBE—Life Sciences Education, vol. 14, no. 4, p. ar40, 2015, doi: https://doi.org/10.1187/cbe.15-02-0043.

[4] M. Versteeg, M. H. van Loon, M. Wijnen-Meijer, and P. Steendijk, "Refuting misconceptions in medical physiology," BMC Medical Education, vol. 20, no. 1, 2020, doi: https://doi.org/10.1186/s12909-020-02166-6.

[5] N. A. Omilani and M. I. Idika, "An investigation into the reasoning of pre-service integrated science teachers when classifying matter into elements and compounds," Creative Education, vol. 11, no. 12, pp. 2512–2522, 2020, doi: https://doi.org/10.4236/ce.2020.1112184.

[6] V. Gkitzia, K. Σάλτα, and C. Tzougraki, "Students' competence in translating between different types of chemical representations," Chemistry Education Research and Practice, vol. 21, no. 1, pp. 307–330, 2020, doi: https://doi.org/10.1039/c8rp00301g.

[7] E. Koster and H. W. de Regt, "Science and values in undergraduate education," Science & Education, vol. 29, no. 1, pp. 123–143, 2019, doi: https://doi.org/10.1007/s11191-019-00093-7.

[8] A. Tripathi, M. Muztaba, P. K. Baghel, P. S. Tajane, S. Gupta, and P. Chitrapu, "Pharmacological effects of the seed extract of Trigonella foenum-graecum L. on cardiovascular and anxiety diseases," International Journal of Zoological Investigations, vol. 9, no. 1, pp. 530–537, 2023, doi: https://doi.org/10.33745/ijzi.2023.v09i01.058.

[9] S. M. Elhusseiny, T. S. El-Mahdy, M. F. Awad, N. S. Elleboudy, M. M. S. Farag, M. A. Yassein, and K. M. Aboshanab, "Proteome analysis and in vitro antiviral, anticancer and antioxidant capacities of the aqueous extracts of Lentinula edodes and Pleurotus ostreatus edible mushrooms," Molecules, vol. 26, no. 15, p. 4623, 2021, doi: https://doi.org/10.3390/molecules26154623.

[10] V. Pedan, N. Fischer, and S. Rohn, "An online NP-HPLC-DPPH method for the determination of the antioxidant activity of condensed polyphenols in cocoa," Food Research International, vol. 89, pp. 890–900, 2016, doi: https://doi.org/10.1016/j.foodres.2015.10.030.

[11] İ. Kurt-Celep et al., "An in-depth study on the metabolite profile and biological properties of Primula auriculata extracts: A fascinating sparkle on the way from nature to functional applications," Antioxidants, vol. 11, no. 7, p. 1377, 2022, doi: https://doi.org/10.3390/antiox11071377.

[12] E. Suchman, "The use of online pre-lab assessments compared with written pre-lab assignments requiring experimental result prediction shows no difference in student performance," Journal of Microbiology and Biology Education, vol. 16, no. 2, pp. 266–268, 2015, doi: https://doi.org/10.1128/jmbe.v16i2.895.

[13] M. J. Gierl, O. Bulut, Q. Guo, and X. Zhang, "Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review," Review of Educational Research, vol. 87, no. 6, pp. 1082–1116, 2017, doi: https://doi.org/10.3102/0034654317726529.

[14] A. Eitel et al., "The misconceptions about multimedia learning questionnaire: An empirical evaluation study with teachers and student teachers," Psychology Learning & Teaching, vol. 20, no. 3, pp. 420–444, 2021, doi:

https://doi.org/10.1177/14757257211028723.
[15] K. K. H. Chan, “Eliciting and working with student thinking: Preservice science teachers’ enactment of core practices when orchestrating collaborative group work,” Journal of Research in Science Teaching, vol. 60, no. 5, pp. 1014–1052, 2022, doi: https://doi.org/10.1002/tea.21823.
[16] C. T. Scott, “Backward by design: Building ELSI into a stem cell science curriculum,” The Hastings Center Report, vol. 45, no. 3, pp. 26–32, 2015, doi: https://doi.org/10.1002/hast.448.
[17] L. D. Goodwin, A. Jones, and B. Hunter, “Addressing social inequity through improving relational care: A social–ecological model based on the experiences of migrant women and midwives in South Wales,” Health Expectations, vol. 25, no. 5, pp. 2124–2133, 2021, doi: https://doi.org/10.1111/hex.13333.
[18] M. Beardsley, P. Santos, D. Hernández-Leo, and K. Michos, “Ethics in educational technology research: Informing participants on data sharing risks,” British Journal of Educational Technology, vol. 50, no. 3, pp. 1019–1034, 2019, doi: https://doi.org/10.1111/bjet.12781.
[19] Y. Li et al., “The impact of coupling assessments on conceptual understanding and connection-making in chemical equilibrium and acid–base chemistry,” Chemistry Education Research and Practice, vol. 21, no. 3, pp. 1000–1012, 2020, doi: https://doi.org/10.1039/d0rp00038h.
[20] K. A. Lawless and S. W. Brown, “Developing scientific literacy skills through interdisciplinary, technology-based global simulations: GlobalEd 2,” The Curriculum Journal, vol. 26, no. 2, pp. 268–289, 2015, doi: https://doi.org/10.1080/09585176.2015.1009133.
[21] J. S. Cetron et al., “Decoding individual differences in STEM learning from functional MRI data,” Nature Communications, vol. 10, no. 1, 2019, doi: https://doi.org/10.1038/s41467-019-10053-y.
[22] A. A. Piacsek, “A new pre/post test to assess student mastery of introductory level acoustics and wave mechanics,” The Journal of the Acoustical Society of America, vol. 144, no. 3_Supplement, pp. 1785–1786, 2018, doi: https://doi.org/10.1121/1.5067886.
[23] J. S. Ilgen, I. Ma, R. Hatala, and D. A. Cook, “A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment,” Academic Medicine, vol. 49, no. 2, pp. 161–173, 2015, doi: https://doi.org/10.1111/medu.12621.
[24] A. Mougias, F. Christidi, G. Kiosterakis, L. Messinis, and A. Politis, “Dealing with severe dementia in clinical practice: A validity and reliability study of Severe Mini-Mental State Examination in Greek population,” International Journal of Geriatric Psychiatry, vol. 33, no. 9, pp. 1236–1242, 2018, doi: https://doi.org/10.1002/gps.4915.
[25] D. G. Nielsen, S. L. Jensen, and L. D. O’Neill, “Clinical assessment of transthoracic echocardiography skills: A generalizability study,” BMC Medical Education, vol. 15, no. 1, 2015, doi: https://doi.org/10.1186/s12909-015-0294-5.
[26] L. Stolz et al., “Multimodular ultrasound orientation: Residents’ confidence and skill in performing point-of-care ultrasound,” Cureus, 2018, doi: https://doi.org/10.7759/cureus.3597.
[27] S. Ehteshami, “Translation and psychometric properties of the Autism Screening Instrument for Educational Planning 3 in Persian (Farsi),” British Journal of Occupational Therapy, vol. 89, no. 1, pp. 46–53, 2025, doi: https://doi.org/10.1177/03080226251361008.
[28] A. J. Cheruth, S. A. M. Al Baloushi, K. Karthishwaran, S. Maqsood, S. S. Kurup, and S. Sakkir, “Medicinally active principles analysis of Tephrosia apollinea (Delile) DC. growing in the United Arab Emirates,” BMC Research Notes, vol. 10, no. 1, 2017, doi: https://doi.org/10.1186/s13104-017-2388-0.
[29] Md. Moniruzzaman et al., “In vitro antioxidant and cholinesterase inhibitory activities of methanolic fruit extract of Phyllanthus acidus,” BMC Complementary and Alternative Medicine, vol. 15, no. 1, 2015, doi: https://doi.org/10.1186/s12906-015-0930-y.
[30] F. Z. Guergouri, W. Sobhi, and M. Benboubetra, “Antioxidant activity of Algerian Nigella sativa total oil and its unsaponifiable fraction,” The Journal of Phytopharmacology, vol. 6, no. 4, pp. 234–238, 2017, doi: https://doi.org/10.31254/phyto.2017.6406.
[31] O. J. Famurewa, “Antioxidant activities of different solvent extracts of Moringa oleifera seeds using DPPH assay,” Asian Journal of Chemical Sciences, vol. 13, no. 5, pp. 78–85, 2023, doi: https://doi.org/10.9734/ajocs/2023/v13i5255.
[32] J. Lee, A. S. Wu, D. Li, and K. Kulasegaram, “Artificial intelligence in undergraduate medical education: A scoping review,” Academic Medicine, vol. 96, no. 11S, pp. S62–S70, 2021, doi: https://doi.org/10.1097/acm.0000000000004291.
[33] M. K. Alam, Z. H. Rana, S. N. Islam, and M. Akhtaruzzaman, “Total phenolic content and antioxidant activity of methanolic extract of selected wild leafy vegetables grown in Bangladesh: A cheapest source of antioxidants,” Potravinarstvo Slovak Journal of Food Sciences, vol. 13, no. 1, pp. 287–293, 2019, doi: https://doi.org/10.5219/1107.
[34] M. Y. Kim, “Evaluation on antioxidant properties of ethanolic leaves and branches extracts of Nerium indicum from Jeju Island in Korea,” Research Journal of Pharmacy and Technology, pp. 4897–4901, 2025, doi: https://doi.org/10.52711/0974-360x.2025.00706.
[35] S. Buczinski and J. Vandeweerd, “Diagnostic accuracy of refractometry for assessing bovine colostrum quality: A systematic review and meta-analysis,” Journal of Dairy Science, vol. 99, no. 9, pp. 7381–7394, 2016, doi: https://doi.org/10.3168/jds.2016-10955.
[36] K. C. dos Santos Berni, A. V. Dibai-Filho, P. F. Pires, and D. Rodrigues-Bigaton, “Accuracy of the surface electromyography RMS processing for the diagnosis of myogenous temporomandibular disorder,” Journal of Electromyography and Kinesiology, vol. 25, no. 4, pp. 596–602, 2015, doi: https://doi.org/10.1016/j.jelekin.2015.05.004.

[37] H. Husseinzadeh, P. A. Gimotty, A. M. Pishko, M. Buckley, T. E. Warkentin, and A. Cuker, "Diagnostic accuracy of IgG-specific versus polyspecific enzyme-linked immunoassays in heparin-induced thrombocytopenia: A systematic review and meta-analysis," Journal of Thrombosis and Haemostasis, vol. 15, no. 6, pp. 1203–1212, 2017, doi: https://doi.org/10.1111/jth.13692.
[38] A. González-Robles et al., "A brief online transdiagnostic measure: Psychometric properties of the Overall Anxiety Severity and Impairment Scale (OASIS) among Spanish patients with emotional disorders," PLOS ONE, vol. 13, no. 11, p. e0206516, 2018, doi: https://doi.org/10.1371/journal.pone.0206516.
[39] Y.-S. Yi, "Probing the relative importance of different attributes in L2 reading and listening comprehension items: An application of cognitive diagnostic models," Language Testing, vol. 34, no. 3, pp. 337–355, 2016, doi: https://doi.org/10.1177/0265532216646141.
[40] Z. Zhang, Y. Hong, N. Liu, and Y. Chen, "Diagnostic accuracy of contrast enhanced ultrasound in patients with blunt abdominal trauma presenting to the emergency department: A systematic review and meta-analysis," Scientific Reports, vol. 7, no. 1, 2017, doi: https://doi.org/10.1038/s41598-017-04779-2.
[41] E. H. Steffensen et al., "Inclusion of sex chromosomes in noninvasive prenatal testing in Asia, Australia, Europe and the USA: A survey study," Prenatal Diagnosis, vol. 43, no. 2, pp. 144–155, 2023, doi: https://doi.org/10.1002/pd.6322.
[42] N. Hu, H. Cheng, K. Zhang, and R. L. Jensen, "Evaluating the prognostic accuracy of biomarkers for glioblastoma multiforme using the Cancer Genome Atlas data," Cancer Informatics, vol. 16, 2017, doi: https://doi.org/10.1177/1176935117734844.
[43] K. Imwattana, P. Putsathit, D. R. Knight, P. Kiratisin, and T. V. Riley, "Molecular characterization of, and antimicrobial resistance in, Clostridioides difficile from Thailand, 2017–2018," Microbial Drug Resistance, vol. 27, no. 11, pp. 1505–1512, 2021, doi: https://doi.org/10.1089/mdr.2020.0603.
[44] D. A. Millado-Riambon, E. Gallardo, and A. Tulay, "Effects of rapid influenza antigen test on antimicrobial management of pediatric patients with influenza-like illness in the emergency room," Pediatric Infectious Disease Society of the Philippines Journal, vol. 22, no. 2, pp. 73–82, 2021, doi: https://doi.org/10.56964/pidspj20212202010.
[45] S. Kim, E. J. Choi, Y.-S. Jung, and I. Jang, "Postoperative delirium screening tools for post-anaesthetic adult patients in non-intensive care units: A systematic review and meta-analysis," Journal of Clinical Nursing, vol. 32, no. 9–10, pp. 1691–1704, 2021, doi: https://doi.org/10.1111/jocn.16157.
[46] R. Lampignano et al., "Multicenter evaluation of circulating cell-free DNA extraction and downstream analyses for the development of standardized (pre)analytical work flows," Clinical Chemistry, vol. 66, no. 1, pp. 149–160, 2019, doi: https://doi.org/10.1373/clinchem.2019.306837.
[47] S. Nurjanah, M. Iqbal, Z. Zafrullah, M. Mahmud, D. S. F. Seran, I. K. Suardi, and L. Arriza, "Psychometric quality of multiple-choice tests under Classical Test Theory (CTT): AnBuso, Iteman, and R," Jurnal Penelitian dan Evaluasi Pendidikan, vol. 28, no. 2, pp. 161–172, 2024, doi: https://doi.org/10.21831/pep.v28i2.71542.
[48] S. Nolte, C. Coon, S. Hudgens, and M. G. E. Verdam, "Psychometric evaluation of the PROMIS® Depression Item Bank: An illustration of Classical Test Theory methods," Journal of Patient-Reported Outcomes, vol. 3, no. 1, 2019, doi: https://doi.org/10.1186/s41687-019-0127-0.
[49] S. M. Lieux, L. E. Crosby, and S. L. Siedlecki, "Psychometrics of an eight-item pulmonary artery catheter safe-care assessment tool for critical care nurses," Journal of Nursing Measurement, vol. 33, no. 3, pp. 425–431, 2024, doi: https://doi.org/10.1891/jnm-2023-0123.
[50] E. Apino, E. Istiyono, H. Retnawati, W. Widihastuti, and K. Hidayati, "Development and calibration of an instrument measuring attitudes toward statistics using classical and modern test theory," Journal of Pedagogical Research, 2024, doi: https://doi.org/10.33902/jpr.202427097.
[51] S. Martens, "Psychometric evaluation of the Modified Neonatal Resuscitation Program Adherence Assessment Tool when utilized for in situ simulation and telesimulation scenarios," American Journal of Perinatology, 2025, doi: https://doi.org/10.1055/a-2722-7228.
[52] A. Hamed et al., "Fabry Disease Patient-Reported Outcome (FD-PRO) demonstrates robust measurement properties for assessing symptom severity in Fabry disease," Molecular Genetics and Metabolism Reports, vol. 29, p. 100824, 2021, doi: https://doi.org/10.1016/j.ymgmr.2021.100824.
[53] K. S. Taber, "The use of Cronbach's alpha when developing and reporting research instruments in science education," Research in Science Education, vol. 48, no. 6, pp. 1273–1296, 2017, doi: https://doi.org/10.1007/s11165-016-9602-2.
[54] P. M. Gutierrez et al., "Evaluating the psychometric properties of the Interpersonal Needs Questionnaire and the Acquired Capability for Suicide Scale in military veterans," Psychological Assessment, vol. 28, no. 12, pp. 1684–1694, 2016, doi: https://doi.org/10.1037/pas0000310.
[55] C. Mulchay, W. R. Rice, and M. Adolf, "Virtual reality-based executive functioning test review: The Nesplora Ice Cream Test," Journal of Pediatric Neuropsychology, vol. 11, no. 1, pp. 34–38, 2025, doi: https://doi.org/10.1037/jpn0000004.
[56] M. Rebhi et al., "Reliability and validity of the Arabic version of the Game Experience Questionnaire: Pilot questionnaire study," JMIR Formative Research, vol. 7, p. e42584, 2023, doi: https://doi.org/10.2196/42584.
[57] L. Deng and W. Chan, "Testing the difference between reliability coefficients alpha and omega," Educational and Psychological Measurement, vol. 77, no. 2, pp. 185–203, 2016, doi: https://doi.org/10.1177/0013164416658325.
[58] A. Janssens et al., "Measurement properties of multidimensional patient-reported outcome measures in neurodisability: A systematic review of evaluation studies," Developmental Medicine & Child Neurology, vol. 58, no. 5, pp. 437–451, 2015, doi: https://doi.org/10.1111/dmcn.12982.

[59] R. S. Vaughan, D. Hanna, and G. Breslin, "Psychometric properties of the Mental Toughness Questionnaire 48 (MTQ48) in elite, amateur and nonathletes," Sport, Exercise, and Performance Psychology, vol. 7, no. 2, pp. 128–140, 2018, doi: https://doi.org/10.1037/spy0000114.
[60] N. Cittadini, D. D'Angelo, E. B. Zannetti, M. Celi, A. Pennini, and G. Rocco, "Development and testing of a new instrument to measure self-care in patients with osteoporosis: The Self-Care of Osteoporosis Scale," International Journal of Bone Fragility, vol. 1, no. 1, pp. 28–33, 2021, doi: https://doi.org/10.57582/ijbf.210101.028.
[61] L. A. Marsan, C. E. D'Arcy, and J. T. Olimpo, "The impact of an interactive statistics module on novices' development of scientific process skills and attitudes in a first-semester research foundations course," Journal of Microbiology and Biology Education, vol. 17, no. 3, pp. 436–443, 2016, doi: https://doi.org/10.1128/jmbe.v17i3.1137.
[62] S. Ayub, J. Rokhmat, A. Busyairi, and G. Afifah, "Kafah Science Test Model to improve the quality of prospective teachers," Jurnal Pendidikan Fisika dan Teknologi, vol. 9, no. 1, pp. 143–150, 2023, doi: https://doi.org/10.29303/jpft.v9i1.5028.
[63] S. H. Padliyyah, "Integration of self-diagnosis in Pascal Law learning using STEM approach," Dinamika Jurnal Ilmiah Pendidikan Dasar, vol. 12, no. 2, p. 104, 2020, doi: https://doi.org/10.30595/dinamika.v12i2.6397.
[64] Z. A. Munoz, "Roles of mathematics-related psychological factors in STEM sense of belonging and identity: A structural equation modeling analysis," International Journal of STEM Education, vol. 12, no. 1, 2025, doi: https://doi.org/10.1186/s40594-025-00586-8.
[65] G. J. B. Aligway et al., "Validity and reliability of concept inventory test in human physiology," JPBI (Jurnal Pendidikan Biologi Indonesia), vol. 10, no. 1, pp. 273–282, 2024, doi: https://doi.org/10.22219/jpbi.v10i1.29558.
[66] T. F. Fuller and J. N. Harb, "Concept inventories for electrochemistry and electrochemical engineering," ECS Meeting Abstracts, vol. MA2019-02, no. 48, p. 2176, 2019, doi: https://doi.org/10.1149/ma2019-02/48/2176.
[67] T. Braun, K. Ehrenbrusthoff, C. Bahns, L. Happe, and C. Kopkow, "Cross-cultural adaptation, internal consistency, test-retest reliability and feasibility of the German version of the Evidence-Based Practice Inventory," BMC Health Services Research, vol. 19, no. 1, 2019, doi: https://doi.org/10.1186/s12913-019-4273-0.
[68] M. Kamaruddin and M. E. E. M. Matore, "Development and validation of psychometric properties of the 10 IB Learner Profile Instrument (10iblp-I): A combination of the Rasch and classical measurement model," International Journal of Environmental Research and Public Health, vol. 18, no. 12, p. 6455, 2021, doi: https://doi.org/10.3390/ijerph18126455.
[69] C. A. C. Prinsen et al., "COSMIN guideline for systematic reviews of patient-reported outcome measures," Quality of Life Research, vol. 27, no. 5, pp. 1147–1157, 2018, doi: https://doi.org/10.1007/s11136-018-1798-3.
[70] J. Petrillo, S. Cano, L. McLeod, and C. D. Coon, "Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to evaluate patient-reported outcome measures: A comparison of worked examples," Value in Health, vol. 18, no. 1, pp. 25–34, 2015, doi: https://doi.org/10.1016/j.jval.2014.10.005.
[71] K. Tang, C. Hsiao, Y. Tu, G. Hwang, and Y. Wang, "Factors influencing university teachers' use of a mobile technology-enhanced teaching (MTT) platform," Educational Technology Research and Development, vol. 69, no. 5, pp. 2705–2728, 2021, doi: https://doi.org/10.1007/s11423-021-10032-5.
[72] Z. Yan, S. P. Brubacher, D. Boud, and M. B. Powell, "Psychometric properties of the Self-assessment Practice Scale for professional training contexts: Evidence from confirmatory factor analysis and Rasch analysis," International Journal of Training and Development, vol. 24, no. 4, pp. 357–373, 2020, doi: https://doi.org/10.1111/ijtd.12201.
[73] C. Lin, M. D. Griffiths, and A. H. Pakpour, "Psychometric evaluation of Persian Nomophobia Questionnaire: Differential item functioning and measurement invariance across gender," Journal of Behavioral Addictions, vol. 7, no. 1, pp. 100–108, 2018, doi: https://doi.org/10.1556/2006.7.2018.11.