Identifying Learning Gaps: A Diagnostic Chemistry Test for Secondary Students

Raihana M. Mangilala[1], Edna B. Nabua[2]

Mindanao State University-Iligan Institute of Technology, Philippines

DOI: https://dx.doi.org/10.47772/IJRISS.2026.1026EDU0027

Received: 24 December 2025; Accepted: 30 December 2025; Published: 14 January 2026

ABSTRACT

This study sought to develop and psychometrically validate a Grade 10 Chemistry achievement instrument intended for diagnostic assessment and to determine students' least mastered Chemistry competencies. A researcher-constructed 50-item multiple-choice instrument was developed based on selected Grade 10 Chemistry learning competencies prescribed in the curriculum. The instrument underwent readability analysis and content validation by three experts in Chemistry education to establish content relevance, clarity, and alignment with intended learning outcomes. Following expert review, the instrument was pilot tested with 150 Grade 10 students to evaluate reliability and perform item analysis. Reliability estimation using Cronbach's alpha yielded a coefficient of 0.846, indicating high internal consistency. Item difficulty and discrimination indices were examined, resulting in the retention of 30 items that satisfied acceptable psychometric standards. The validated 30-item achievement instrument was subsequently administered to 88 Grade 10 students from two purposively selected sections at MSU–University Training Center, Marawi City. Student performance was analyzed using a criterion-referenced framework to determine mastery levels across Chemistry learning competencies. Findings revealed differential levels of mastery, with Chemical Reactions emerging as the least mastered competency, while topics related to Solutions, Acids and Bases, and Gases and Gas Laws were moderately mastered. Overall, the results provide empirical evidence that the developed instrument demonstrates satisfactory validity and reliability and functions effectively as a diagnostic assessment tool. The study highlights the critical role of validated diagnostic assessments in identifying learning gaps and informing targeted instructional interventions to enhance the teaching and learning of Chemistry at the secondary level.

Keywords: diagnostic assessment, Grade 10 Chemistry, least mastered competencies,

INTRODUCTION

Chemistry is a fundamental component of secondary science education, playing a critical role in the development of students' scientific reasoning, problem-solving abilities, and evidence-based decision-making skills. At the Grade 10 level, Chemistry instruction encompasses a broad range of foundational concepts that underpin more advanced studies in science and related disciplines. Attaining the intended learning outcomes at this level requires not only effective pedagogical strategies but also assessment practices that accurately capture students' conceptual understanding and mastery of key competencies.

Assessment serves a central function in monitoring student learning, diagnosing misconceptions, and informing instructional decision-making. In Chemistry education, this function is particularly significant because many core topics involve abstract ideas, symbolic representations, and multilevel conceptualizations that pose persistent challenges for learners (Johnstone, 2006; Taber, 2018). Well-designed assessments provide essential evidence of students' conceptual progress and enable teachers to identify areas requiring targeted instructional support.

Achievement tests are among the most widely used tools for evaluating students' mastery of Chemistry competencies. However, the effectiveness of these assessments depends largely on their psychometric quality, including content validity, reliability, readability, and appropriate item functioning (DeVellis, 2017; Tavakol & Dennick, 2011). Instruments that lack systematic validation may yield results that misrepresent students' true level of understanding, thereby limiting their diagnostic and instructional value. Despite the widespread use of achievement tests in secondary Chemistry classrooms, locally developed and empirically validated diagnostic instruments remain limited.

In response to this gap, the present study focused on the development and validation of a Grade 10 Chemistry achievement test designed for diagnostic purposes. By establishing the psychometric properties of the instrument

and applying it to identify students' least mastered Chemistry competencies, the study aimed to provide a reliable assessment tool that can support data-informed instructional planning and improve learning outcomes in secondary Chemistry education.

Research Objectives. The primary objective of this study was to develop and psychometrically validate a Grade 10 Chemistry achievement test and to identify the least mastered Chemistry topics among Grade 10 students. Specifically, the study aimed to construct a 50-item multiple-choice achievement test aligned with selected curriculum-based learning competencies, establish the content validity and readability of the instrument through expert evaluation, and determine its internal consistency reliability using Cronbach's alpha coefficient. Following the validation process, the study further sought to utilize the finalized 30-item achievement test to diagnose students' levels of mastery across Chemistry learning competencies and to identify areas of least mastery for instructional support.

METHODOLOGY

Research Design. This study employed a descriptive research design incorporating test development and diagnostic assessment components. This design was appropriate because the study focused on the systematic construction, validation, and refinement of an achievement test and its use in describing students' mastery of Chemistry competencies, rather than on manipulating variables or evaluating the effects of an instructional intervention.

The research process involved several sequential phases: test construction based on selected Grade 10 Chemistry learning competencies, readability analysis, expert content validation, pilot testing, reliability estimation, and item analysis. Following the establishment of acceptable psychometric properties, the validated achievement test was administered to a group of Grade 10 students to generate descriptive data on their performance across Chemistry topics. Student scores were subsequently analyzed using a criterion-referenced approach to determine levels of mastery and to identify the least mastered competencies.

The descriptive test development and diagnostic framework enabled the generation of empirical evidence regarding the quality of the assessment instrument while simultaneously providing meaningful insights into students' learning gaps. This methodological approach was well-suited to achieving the study's objectives of producing a valid and reliable Chemistry achievement test and demonstrating its utility as a diagnostic tool in secondary science education.

Table 1 presents the development and validation phases of the Chemistry achievement questionnaire, including the number of items, participants, and the purpose of each phase.

Table 1. Development and Validation Phases of the Chemistry Achievement Test

Phase	Number of Items	Number of Participants	Purpose
Test Construction	50	---	Initial Item Development
Pilot Testing	50	150	Reliability and Item Analysis
Final Implementation	30	88	Diagnostic Assessment

Participants. The participants of this study comprised two distinct groups of Grade 10 students selected through purposive sampling to address different phases of the research. During the pilot testing phase, 150 Grade 10 students participated in the administration of the initial 50-item Chemistry achievement test. Data from this group were used to establish the reliability of the instrument and to conduct item analysis. For the final implementation phase, the 88 Grade 10 students drawn from two Grade 10 sections at MSU–University Training Center, Marawi City, were selected to complete the validated 30-item Chemistry achievement test for diagnostic purposes. These sections were identified as having average academic performance based on school records. The deliberate selection of average-performing sections was intended to represent typical Grade 10 Chemistry learners and to generate meaningful diagnostic data on students' mastery levels and least mastered Chemistry topics.

Research Instruments. The primary research instrument employed in this study was a researcher-developed Grade 10 Chemistry achievement test designed to assess students' mastery of selected curriculum-based learning competencies. The instrument was initially constructed as a 50-item multiple-choice test, with each item consisting of one correct answer and three plausible distractors. Test items were developed to represent key content areas in the Grade 10 Chemistry curriculum, including matter and its properties, atomic structure and the periodic table, chemical bonding, chemical reactions, gases and gas laws, solutions, acids and bases, thermochemistry and kinetics, chemistry in everyday life, organic and biochemistry, and nuclear chemistry.

To ensure the appropriateness of the instrument for Grade 10 learners, the test items underwent readability analysis using an open-source readability assessment tool and content validation by three experts in Chemistry education. The experts evaluated each item in terms of clarity, relevance, and alignment with the intended learning competencies, and the instrument was revised based on their recommendations. The revised 50-item test was subsequently pilot tested with 150 Grade 10 students to evaluate its psychometric properties. Reliability analysis using Cronbach's alpha yielded a coefficient of 0.846, indicating high internal consistency. Item analysis focusing on item difficulty and discrimination indices was then conducted, resulting in the retention of 30 items that met acceptable psychometric criteria. The finalized 30-item Chemistry achievement test served as the validated instrument for the study. It was administered to the selected Grade 10 students to determine levels of mastery across Chemistry learning competencies and to identify the least mastered topics.

Table 2. Percentage Score Ranges and Corresponding Levels of Mastery

Percentage Score	Level of Mastery
81-100%	Highly Mastered
61-80%	Moderately Mastered
41-60%	Fairly Mastered
21-40%	Least Mastered
0-20%	Not Mastered

Mastery Scale. The mastery scale employed in this study, as presented in Table 2, was criterion-referenced and researcher-defined, drawing on established principles of educational measurement and classroom-based diagnostic assessment. Students' percentage scores were categorized into ordered mastery levels ranging from highly mastered to not mastered. The use of percentage score ranges to describe mastery levels is consistent with recommended practices for interpreting achievement data in diagnostic and formative assessment contexts, where the primary objective is to identify learning gaps and inform instructional decision-making rather than to rank learners normatively (Nitko & Brookhart, 2014; Popham, 2017; McMillan, 2018). This criterion-referenced approach allowed for meaningful interpretation of students' performance relative to predefined learning expectations and facilitated the identification of the least mastered Chemistry competencies.

Limitations of the Study. The findings of this study should be interpreted in light of several limitations inherent in test development and initial validation research. First, the implementation of the validated instrument involved a relatively small sample of 88 Grade 10 students drawn from two purposively selected average-performing sections in a single school. This sampling approach, while appropriate for diagnostic purposes, limits the generalizability of the findings to broader student populations and educational contexts. Instrument validation literature emphasizes that results from initial applications are often context-specific and require replication using larger and more diverse samples to establish broader applicability (Morgado et al., 2017).

Second, although the achievement test was initially constructed to ensure balanced representation across Chemistry learning competencies, the final validated instrument consisted of an unequal number of items per competency as a consequence of the item analysis process. Only items that satisfied acceptable psychometric criteria for difficulty, discrimination, and reliability were retained. As a result, some competencies were represented by fewer items than others. Unequal item distribution may reduce the precision of competency-level mastery estimates, particularly when a competency is measured by only one or two items, and therefore necessitates cautious interpretation of the results (Taber, 2018; Tavakol & Dennick, 2011). Accordingly, the identification of least mastered competencies in this study should be regarded as indicative rather than definitive.

Third, the study relied primarily on Classical Test Theory (CTT) methods, including Cronbach's alpha and item difficulty and discrimination indices, to establish the psychometric properties of the instrument. CTT-based

estimates are known to be sample-dependent, such that item functioning and reliability coefficients may vary across different groups of test takers (Soeharto, 2021). Consequently, the instrument is considered initially validated, and further validation studies involving additional samples and alternative measurement models, such as Item Response Theory, are recommended.

Despite these limitations, the study contributes to a systematically developed and empirically supported diagnostic assessment tool for identifying learning difficulties in Grade 10 Chemistry. Future research may strengthen the instrument by developing additional validated items for underrepresented competencies and by administering the test to broader and more diverse student populations.

Data Gathering Procedure. Data collection followed a systematic and sequential process of test development, validation, and implementation. The procedure began with securing formal permission from school principals and administrators to conduct the study. This was followed by the construction of the initial 50-item Chemistry achievement test based on selected Grade 10 learning competencies, readability analysis, and expert content validation. After revisions based on expert feedback, the instrument was pilot tested to establish reliability and conduct item analysis. Items meeting acceptable psychometric criteria were retained to form the finalized 30-item validated test. The validated instrument was then administered to the selected Grade 10 students for diagnostic assessment. The overall sequence of these procedures is illustrated in Figure 1, which summarizes the stages from initial approval to final test administration.

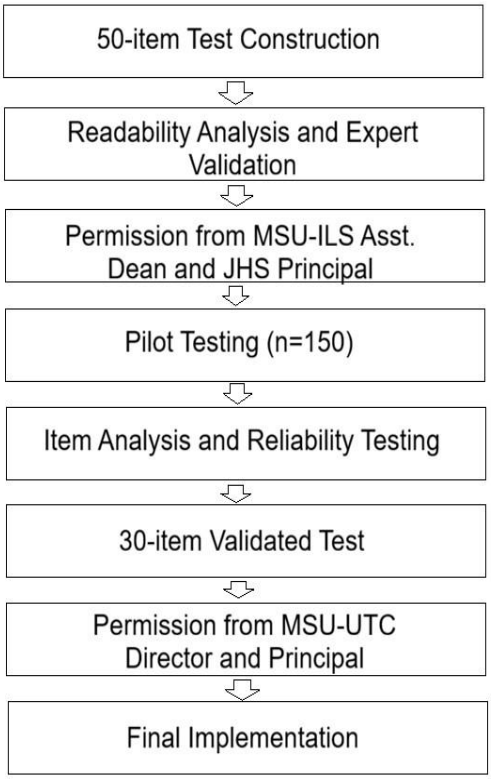

Figure 1. Flow of the data gathering procedure

Data Analysis. The data collected from the pilot testing and final implementation of the Grade 10

The chemistry achievement test was analyzed using descriptive and psychometric statistical techniques, consistent with standard practices in educational test development and validation (DeVellis, 2017; Nitko & Brookhart, 2014). Separate analyses were conducted for the pilot testing and final implementation phases of the study.

For the pilot testing phase, responses from 150 Grade 10 students were analyzed to determine the psychometric properties of the initial 50-item test. Internal consistency reliability was estimated using Cronbach's alpha, which is widely used to assess the consistency of items within an achievement test (Tavakol & Dennick, 2011). Item analysis was also performed to examine item difficulty and item discrimination indices, following Classical Test Theory procedures. The results of the reliability analysis and item analysis guided decisions on item retention and revision, leading to the selection of 30 items that met acceptable psychometric criteria (Ebel & Frisbie, 1991).

For the final implementation phase, data from 88 Grade 10 students who answered the validated 30-item test were analyzed descriptively. Students' raw scores were converted into percentage scores and interpreted using a criterion-referenced approach, which focuses on describing students' level of mastery of specific learning competencies rather than comparing performance across individuals (Nitko & Brookhart, 2014; Popham, 2017). To identify the least mastered Chemistry topics, the mean percentage score for each learning competency was computed. Competencies with the lowest mean percentage scores were identified as the least mastered areas. This approach is appropriate for diagnostic assessment, particularly when learning competencies are represented by an unequal number of test items (McMillan, 2018). All data were analyzed and reported in aggregate form to ensure confidentiality. The results of the analyses were presented using tables to facilitate clear interpretation and discussion of findings.

Ethical Considerations. Ethical considerations were observed throughout the conduct of this study to ensure the protection of the participants and the integrity of the research process. Prior to data collection, formal permission was obtained from the school principals and school administrators of the participating schools to conduct the study. Coordination with the teachers concerned was also undertaken to ensure that the data-gathering activities did not disrupt regular classroom instruction.

Participation in the study was voluntary, and the Grade 10 students were informed of the purpose of the study and the nature of their participation. Students were assured that their responses would be used solely for research purposes and would not affect their academic grades or standing. To protect participants' privacy, no identifying information, such as names or student numbers, was collected, and all test results were treated with strict confidentiality.

The data collected from both the pilot testing and the final implementation were handled responsibly and used only for analysis related to the objectives of the study. The results were reported in aggregate form, ensuring that no individual participant or class could be identified. All procedures were conducted in accordance with generally accepted ethical standards for educational research.

RESULTS AND DISCUSSIONS

This chapter presents and discusses the results of the study in relation to its stated objectives. The findings are organized to reflect the sequential process of test development, validation, and implementation. Specifically, the chapter reports the results of the construction of the 50-item Grade 10 Chemistry achievement test, the outcomes of content validity and readability evaluation, and the reliability analysis conducted during pilot testing. It further presents the results of the final administration of the validated 30-item test, focusing on students' performance across selected Chemistry learning competencies. The discussion emphasizes the identification of the least mastered Chemistry topics using a criterion-referenced approach, with interpretations grounded in the descriptive and diagnostic purpose of the study.

Table 3. Summary of the Content Validity

Evaluation Criterion	Description	Mean Score	Qualitative Interpretation
Content Validity	Accurately represents key concepts.	4	Excellent
Clarity of Wording	Questions and options are clearly, concisely, and grammatically stated.	4	Excellent
Appropriateness of Difficulty	Well-matched to students' level; neither too easy nor too hard.	4	Excellent
Quality of Distractors	All distractors are plausible and conceptually sound.	4	Excellent
Format and Consistency	Consistent format and parallel structure.	4	Excellent
Alignment with Learning Objective	Directly align with the intended learning outcome.	4	Excellent
Score: 4-Excellent, 3-Good, 2-Fair, 1-Needs Improvement			

Table 3 presents a summary of the results of the content validity evaluation of the Chemistry achievement test as assessed by subject-matter experts. Across all evaluation criteria—namely content validity, clarity of wording, appropriateness of difficulty level, quality of distractors, format and consistency, and alignment with learning objectives—the instrument obtained a mean rating of 4, corresponding to a qualitative interpretation of Excellent. These results indicate a high level of agreement among the experts that the test items accurately represent essential Chemistry concepts, are clearly and grammatically stated, and are appropriate for the cognitive level of Grade 10 learners.

The uniformly high ratings further suggest that the distractors were plausible and conceptually sound, thereby reducing the likelihood of random guessing and enhancing the diagnostic value of the instrument. Moreover, the excellent rating for alignment with learning objectives confirms that the test items are directly linked to the intended curriculum outcomes, which is a fundamental requirement for establishing content validity in achievement testing. Collectively, these findings provide strong evidence that the instrument demonstrates satisfactory content validity and readability, supporting its suitability for pilot testing and subsequent reliability and item analyses.

These results are consistent with the existing literature that emphasizes the central role of expert judgment in establishing content validity during instrument development. Zamanzadeh et al. (2015) underscored that expert evaluation of item clarity, relevance, and representativeness is a critical step in ensuring that an assessment instrument adequately measures its intended construct. High levels of expert agreement, as reflected in excellent mean ratings, are widely regarded as strong indicators of content validity. Similarly, Boateng et al. (2018) highlighted that clearly worded items with appropriate difficulty levels and well-designed distractors contribute substantially to the overall quality and validity of achievement tests, and that expert review serves to identify ambiguities and misalignment prior to empirical testing.

Further support is provided by Kalkbrenner (2021), who emphasized that structured expert evaluation yielding consistently high ratings across multiple criteria signals an instrument's readiness for pilot testing. In particular, readability and alignment with instructional objectives were identified as critical considerations in school-based assessments, as they help ensure that test performance reflects students' content knowledge rather than linguistic complexity. Taken together, these studies corroborate the present findings and affirm that the developed Chemistry achievement test possesses strong content validity and readability, thereby justifying its use for pilot testing and diagnostic assessment purposes.

Table 4. Reliability of the 50-item Chemistry Achievement Questionnaire

Test Version	Number of Items	*Cronbach's Alpha*	Qualitative Interpretation
Pilot Test	50	0.846	High Reliability

The reliability results of the 50-item Chemistry achievement questionnaire obtained from the pilot testing phase are presented in Table 4. The computed Cronbach's alpha coefficient of 0.846 indicates a high level of internal consistency, suggesting that the test items function coherently in measuring students' achievement in Chemistry. This reliability coefficient implies that the items are sufficiently interrelated and consistently reflect the underlying construct being assessed. In educational measurement, Cronbach's alpha values exceeding 0.80 are generally interpreted as indicative of acceptable to high reliability, particularly during the initial stages of instrument development. Accordingly, the reliability result demonstrates that the test scores are stable and dependable, supporting the use of the instrument for subsequent item analysis and test refinement. The findings provide empirical justification for retaining the 50-item test as a basis for further validation procedures, including the selection of items for the final validated instrument.

The reliability coefficient obtained in this study is consistent with findings reported in the educational measurement literature. Taber (2018) emphasized that Cronbach's alpha values above 0.80 reflect strong internal consistency, especially in science education assessments where constructs may be multidimensional yet conceptually related. Such values indicate that test items collectively measure the intended domain without excessive redundancy. Similarly, Hair et al. (2019) noted that reliability coefficients ranging from 0.70 to 0.90 are desirable during early-stage test or scale development, as they represent an appropriate balance between item consistency and content coverage, thereby supporting the instrument's use for diagnostic and research purposes.

Further support is provided by Kyriazos and Stalikas (2018), who argued that high reliability coefficients obtained during pilot testing offer strong empirical grounds for proceeding with item refinement rather than redeveloping the instrument. They underscored that a Cronbach's alpha coefficient exceeding 0.80 indicates that an assessment tool is sufficiently reliable for continued psychometric evaluation and application in educational settings. Taken together, these findings corroborate the reliability results presented in Table 4 and affirm that the Chemistry achievement test demonstrates strong internal consistency, rendering it suitable for item analysis and subsequent diagnostic implementation.

Table 5. Summary of Item Retention after Item Analysis

Test Version	Number of Items
Pilot Test	50
Validated Test	30

The results of the item retention process following the pilot testing of the Chemistry achievement test are summarized in Table 5. From the initial pool of 50 items, a total of 30 items were retained for inclusion in the validated version of the instrument. This reduction reflects the systematic application of item analysis procedures to identify items that met acceptable psychometric standards, particularly with respect to item difficulty and discrimination indices. Retaining only those items that demonstrated satisfactory statistical performance indicates that the final test is composed of items that function effectively in measuring students' Chemistry achievement. The removal of poorly performing items enhances the overall quality of the instrument by improving its validity, internal coherence, and interpretability. Consequently, the validated 30-item test is more appropriate for diagnostic purposes, particularly for identifying the least mastered Chemistry competencies among Grade 10 students.

The refinement of test items through empirical item analysis is a widely recognized and recommended practice in educational measurement. Haladyna and Rodriguez (2013) emphasized that item analysis is essential for identifying weak, ambiguous, or non-discriminating items and for ensuring that retained items contribute meaningfully to score interpretation. They further noted that item reduction following pilot testing strengthens test quality and measurement accuracy. Similarly, Magno (2017) reported that empirically guided reduction of test items is a critical step in achievement test development, as it helps eliminate items that fail to differentiate adequately among levels of student performance, thereby enhancing the diagnostic and evaluative value of the instrument.

Additional support is provided by Kline (2021), who underscored the importance of grounding item retention decisions in psychometric evidence. The author argued that retaining only items with acceptable statistical properties results in more efficient assessment instruments without compromising measurement validity. This consideration is particularly salient in educational diagnostic contexts, where precision and clarity of measurement are paramount. Taken together, these findings affirm that the reduction of items from 50 to 30 following item analysis reflects sound measurement practice and contributes to the development of a more valid and reliable Grade 10 Chemistry achievement test.

The results align with prior research emphasizing the importance of systematic test validation in producing reliable and meaningful diagnostic assessments (Furr & Bacharach, 2014). Consistent with findings in Chemistry education literature, students continue to experience greater difficulty with abstract topics, particularly chemical reactions, due to challenges in coordinating symbolic, particulate, and conceptual representations (Bain et al., 2019; Sevian & Talanquer, 2021). These findings underscore the critical role of diagnostic assessments in identifying least mastered competencies and guiding evidence-based instructional improvement.

Table 6. Mean Percentage Scores and Mastery Levels by Chemistry Learning Competency

Learning Competency	Number of Items	Mean Percentage of Correct Responses	Level of Mastery	Rank
Matter and Its Properties	1	58	Fairly Mastered	6
Atomic Structure and Periodic Table	4	62.5	Moderately Mastered	3

Chemical Bonding	3	51.89	Fairly Mastered	9
Chemical Reactions	1	48	Fairly Mastered	10
Gases and Gas Laws	4	64.20	Moderately Mastered	2
Solutions, Acids, and Bases	3	64.77	Moderately Mastered	1
Thermochemistry and Kinetics	2	53.98	Fairly Mastered	8
Chemistry in Everyday Life	4	61.36	Moderately Mastered	4
Organic and Biochemistry	5	58.86	Fairly Mastered	5
Nuclear Chemistry	3	54.55	Fairly Mastered	7

Analysis of students' performance across selected Grade 10 Chemistry competencies using the validated 30-item achievement test revealed notable variability in mastery levels (Table 6). Solutions, Acids, and Bases (64.77%) and Gases and Gas Laws (64.20%) achieved the highest mean scores, both interpreted as Moderately Mastered, suggesting that students were more proficient in topics reinforced through practical applications, problem-solving exercises, and laboratory experiences. In contrast, Chemical Reactions emerged as the least mastered competency, with the lowest mean score (48%), indicating substantial difficulty in integrating conceptual understanding, symbolic representations, and particulate-level reasoning. Competencies such as Chemical Bonding, Thermochemistry and Kinetics, Nuclear Chemistry, and Organic and Biochemistry were classified as Fairly Mastered, reflecting partial understanding and persistent learning gaps. Collectively, these results indicate that while students showed moderate competence in applied topics, abstract and conceptually demanding areas remain challenging, highlighting the diagnostic value of the developed test for informing targeted instructional interventions.

The difficulty with chemical reactions aligns with prior research documenting students' challenges in linking macroscopic observations to symbolic and sub-microscopic representations, which often results in fragmented understanding and misconceptions (Talanquer & Pollard, 2010; Cooper, Underwood, & Hilley, 2012; Stowe & Cooper, 2019). Such representational and reasoning demands render chemical reactions among the most conceptually complex topics in secondary Chemistry, reinforcing the need for diagnostic assessment to identify areas of weakness. Moreover, research suggests that conceptually oriented instruction significantly improves student understanding of reactions compared with algorithmic approaches, particularly when instructional strategies explicitly integrate macroscopic, symbolic, and particulate-level representations (Kozma & Russell, 2005; Dori & Hameiri, 2003). The early identification of misconceptions through diagnostic tools further enables teachers to implement targeted interventions that enhance conceptual learning (Gulacar, Overton, & Bowen, 2014).

Taken together, these findings confirm that abstract topics, particularly chemical reactions, pose persistent challenges for Grade 10 students. They underscore the importance of validated diagnostic assessments and concept-focused instructional strategies in addressing learning gaps and improving mastery in secondary Chemistry.

CONCLUSION

This study successfully developed and validated a Grade 10 Chemistry achievement questionnaire that demonstrated strong content validity, readability, and high internal consistency. Expert evaluation confirmed that the instrument items were clear, relevant, and well aligned with the intended learning competencies, while pilot testing yielded a Cronbach's alpha coefficient of 0.846, indicating reliable measurement of students' Chemistry achievement. Item analysis further refined the instrument, resulting in a validated 30-item achievement instrument suitable for diagnostic use. The final implementation of the instrument revealed varying levels of mastery across Chemistry learning competencies. Although students demonstrated moderate mastery of applied topics such as Solutions, Acids, and Bases, and Gases and Gas Laws, Chemical Reactions emerged as the least mastered competency, indicating substantial learning gaps in this area. These findings highlight the utility of the validated achievement instrument as an effective diagnostic tool for identifying specific areas of difficulty and informing targeted instructional interventions in Grade 10 Chemistry.

RECOMMENDATIONS

Based on the findings of the study, it is recommended that greater instructional emphasis be placed on Chemical Reactions, which was identified as the least mastered competency among Grade 10 students. Instructional strategies that prioritize conceptual understanding—such as the explicit integration of macroscopic observations, symbolic representations, and particulate-level explanations—may be strengthened to address students' learning difficulties in this area. The validated 30-item Chemistry achievement test developed in this study may also be utilized by teachers as a diagnostic assessment tool to identify specific learning gaps and to guide targeted remediation efforts.

Furthermore, future research may enhance the instrument by developing additional validated items for underrepresented competencies and by administering the test to larger and more diverse student populations to improve its diagnostic precision and generalizability. Collectively, these findings support the recommendation to employ diagnostic tools and concept-focused teaching strategies to improve mastery of challenging Chemistry competencies.

ACKNOWLEDGMENT

The researcher expresses sincere gratitude to Mindanao State University–Integrated Laboratory School (MSU–ILS) for permitting the pilot testing of this study and to Mindanao State University–University Training Center (MSU–UTC) for allowing its implementation. Deep appreciation is extended to Dr. Ruben Leo D. Alabat for his guidance, valuable insights, and continuous support throughout the research process, and to Dr. Edna B. Nabua for her encouragement and assistance that greatly contributed to the completion of this paper. The researcher also acknowledges the support of friends and family for their motivation and understanding. Above all, heartfelt thanks are given to the researcher's husband, Javier M. Usop, whose unwavering support and encouragement provided strength throughout the conduct of this study.

REFERENCES

1. Bain, K., Rodriguez, J. M. G., Moon, A., & Towns, M. H. (2019). The characterization of college students' reasoning about chemical reactions using a mechanistic framework. Chemistry Education Research and Practice, 20(4), 741–755. https://doi.org/10.1039/C9RP00025A
2. Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. Frontiers in Public Health, 6, 149. https://doi.org/10.3389/fpubh.2018.00149
3. Cooper, M. M., Underwood, S. M., & Hilley, C. Z. (2012). Development and validation of the implicit information from Lewis structures instrument (IILSI): Do students connect structures with properties? Chemistry Education Research and Practice, 13(3), 195–200. https://doi.org/10.1039/C2RP00010E
4. DeVellis, R. F. (2017). Scale development: Theory and applications (4th ed.). Sage Publications.
5. Dori, Y. J., & Hameiri, M. (2003). Multidimensional analysis system for quantitative chemistry problems: Symbol, macro, micro, and process aspects. Journal of Research in Science Teaching, 40(3), 278–302. https://doi.org/10.1002/tea.10090
6. Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Prentice Hall.
7. Furr, R. M., & Bacharach, V. R. (2014). Psychometrics: An introduction (2nd ed.). Sage Publications. https://doi.org/10.4135/9781483390869
8. Gulacar, O., Overton, T., & Bowen, C. W. (2014). Development and application of a diagnostic assessment tool to evaluate students' understanding of thermodynamics concepts. Chemistry Education Research and Practice, 15(3), 447–458. https://doi.org/10.1039/C4RP00017H
9. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning. https://doi.org/10.1002/9781119409137
10. Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. Routledge. https://doi.org/10.4324/9780203850381
11. Johnstone, A. H. (2006). Chemical education research in Glasgow in perspective. Chemistry Education Research and Practice, 7(2), 49–63. https://doi.org/10.1039/B5RP90021B
12. Kline, R. B. (2021). Principles and practice of structural equation modeling (4th ed.). Guilford Press. https://doi.org/10.1037/0000157-000

13. Kalkbrenner, M. T. (2021). A practical guide to instrument development and score validation in the social sciences: The MEASURE approach. Practical Assessment, Research, and Evaluation, 26(1), Article 1. https://doi.org/10.7275/svg4-e671
14. Kozma, R. B., & Russell, J. (2005). Students becoming chemists: Developing representational competence. Visualization in Science Education, 1, 121–145. https://doi.org/10.1007/1-4020-3613-2_8
15. Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. Psychology, 9(11), 2531–2560. https://doi.org/10.4236/psych.2018.911145
16. Magno, C. (2017). Developing and validating achievement tests in educational settings. International Journal of Educational and Psychological Assessment, 18(2), 1–15. https://doi.org/10.2139/ssrn.2966335
17. McMillan, J. H. (2018). Classroom assessment: Principles and practice for effective standards-based instruction (7th ed.). Pearson Education.
18. Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. Psicologia: Reflexão e Crítica, 30(1), 3. https://doi.org/10.1186/s41155-016-0057-1
19. Nitko, A. J., & Brookhart, S. M. (2014). Educational assessment of students (7th ed.). Pearson Education.
20. Popham, W. J. (2017). Classroom assessment: What teachers need to know (8th ed.). Pearson Education.
21. Sevian, H., & Talanquer, V. (2021). Rethinking chemistry education: From teaching concepts to supporting sensemaking. Chemistry Education Research and Practice, 22(1), 8–15. https://doi.org/10.1039/D0RP00231F
22. Soeharto, S. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: A Rasch measurement approach. Journal of Turkish Science Education, 18(3), 351–370. https://doi.org/10.36681/tused.2021.78
23. Stowe, R. L., & Cooper, M. M. (2019). Practicing what we preach: Assessing “critical thinking” in chemistry. Journal of Chemical Education, 96(5), 837–846. https://doi.org/10.1021/acs.jchemed.8b00736
24. Taber, K. S. (2018). The nature of the chemical concept: Reconsidering chemical ideas. Royal Society of Chemistry. https://doi.org/10.1039/9781788012627
25. Taber, K. S. (2018). The use of Cronbach’s alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2
26. Talanquer, V., & Pollard, J. (2010). Let’s teach how we think instead of what we know. Chemistry Education Research and Practice, 11(2), 74–83. https://doi.org/10.1039/C005349J
27. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach’s alpha. International Journal of Medical Education, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
28. Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. Journal of Caring Sciences, 4(2), 165–178. https://doi.org/10.15171/jcs.2015.017