Evaluating Learners' Performance in Secondary Chemistry Across Contextualized Stimuli and Cognitive Domains

Julyza A. Sayre, Edna B. Nabua

Mindanao State University University - Iligan Institute of Technology, Philippines

DOI: https://doi.org/10.47772/IJRISS.2026.10100098

Received: 30 December 2025; Accepted: 05 January 2026; Published: 23 January 2026

ABSTRACT

Contemporary reforms in science education emphasize the assessment of learners' ability to apply knowledge and reason within authentic, real-world contexts. Despite this shift, assessment practices in secondary chemistry classrooms often remain decontextualized and focused on recall. Addressing this gap, the present study examined the performance of Grade 11 learners in secondary chemistry using a contextualized, stimulus-based assessment. Specifically, the study described learners' overall achievement, analyzed mastery across chemistry competencies, and examined performance patterns across different stimulus types and levels of cognitive demand. A descriptive research design was employed involving 72 Grade 11 learners from a secondary school in Iligan City, Philippines. Data were gathered using a researcher-developed, validated stimulus-based chemistry assessment consisting of 47 multiple-choice items anchored in experimental, technology-related, environmental or societal, and everyday-life contexts. Learners performance was analyzed using descriptive statistics as well as the Friedman test to examine the differences across stimulus types and cognitive domains. Results indicated that a majority of learners did not meet expected performance levels when assessed through contextualized tasks. Mastery across chemistry competencies was generally at an average level, with relatively higher performance observed in Chemical Kinetics and Chemistry in Everyday Life and Environment, and lower mastery in more abstract competencies such as Chemical Bonding and Atomic Structure. Performance also varied across stimulus types, with higher performances in experiment-based and environmental contexts and lower scores in technology-related and everyday-life stimuli. Across cognitive domains, learners demonstrated average performance, with comparatively higher outcomes in Applying and Creating than in Understanding and Analyzing. These findings suggest that learner performance in stimulus-based chemistry assessments is associated with both contextual characteristics and cognitive demand. The study provides classroom-based evidence that contextualized assessment can reveal important patterns in learners' chemistry understanding and highlights the need for closer instructional and assessment alignment to support conceptual understanding and knowledge transfer in secondary chemistry.

Keywords: Contextualized assessment, Cognitive demand, Stimulus-based assessment, Secondary Chemistry, Student Performance

INTRODUCTION

Recent reforms in science education emphasize assessment practices that capture learners' ability to apply knowledge, reason with evidence, and solve problems in authentic contexts rather than merely recall isolated facts. Consequently, contextualized and stimulus-based assessments, such as scenario-based tasks anchored in experiments, data sets, technological applications, and socio-environmental issues, have gained prominence as tools for eliciting meaningful science learning and higher-order thinking (Nordine et al., 2020; Ruiz-Primo & Li, 2016). These approaches align with contemporary assessment frameworks that prioritize the measurement of "knowledge-in-use," particularly in cognitively demanding disciplines such as chemistry (OECD, 2015; Seah et al., 2024).

Within chemistry education, the movement toward contextualized assessment is especially consequential given the abstract, symbolic, and representational character of many core concepts (Chi et al., 2018). Empirical evidence consistently demonstrates that while learners may succeed on routine, algorithmic tasks, they often struggle when required to interpret empirical data, integrate contextual information, or transfer conceptual understanding to real-world situations information, or transfer concepts to real-world situations (Gholiyah & Lutfi, 2021; Galymova et al., 2024). Although the integration of real-life contexts has been shown to enhance relevance and learner engagement, context-rich assessment tasks may simultaneously heighten cognitive and literacy demands. As a result, student performance becomes sensitive not only to content knowledge but also to the nature of the stimulus and the level of cognitive processing required (Nordine et al., 2020).

A central concern in the design and interpretation of contextualized or stimulus-based assessments is the extent to which observed performance reflects variations in content mastery versus differences in item characteristics. Prior research indicates that features such as stimulus format (e.g., narrative descriptions, graphical data sets, or real-world case scenarios) and associated cognitive demand (e.g., recall, application, or analysis) can systematically influence student responses. Findings from large-scale and international science assessments further suggest that both item format and cognitive domain are significantly associated with performance outcomes, implying that reliance on aggregated scores may obscure meaningful patterns across contextual and cognitive dimensions (Pey-yan et al., 2020)

In the Philippine education context, national policy initiatives explicitly advocate for the development of higher-order thinking skills and the integration of contextualized learning experiences in science instruction. Despite these policy directions, assessment practices in secondary chemistry classrooms remain predominantly traditional and decontextualized, thereby constraining opportunities to examine learners' reasoning processes and their ability to apply chemical concepts in authentic situations (Sanchez, 2019; Bello et al., 2023). Furthermore, empirical school-based studies that systematically disaggregate chemistry assessment outcomes by stimulus type and cognitive demand are scarce. Addressing this gap, the present study investigates the performance of Grade 11 learners in secondary chemistry using a contextualized, stimulus-based assessment, with particular emphasis on overall achievement, mastery across secondary chemistry competencies, variation across stimulus types, and differences across levels of cognitive demand.

METHODOLOGY

This study adopted a descriptive research design to examine the performance of Grade 11 learners in secondary chemistry using a contextualized, stimulus-based assessment. A descriptive approach was appropriate because the investigation aimed to document existing performance patterns across competencies, stimulus types, and cognitive domains without manipulating instructional variables or inferring causal relationships. In addition to descriptive statistical measures, inferential analysis using the Friedman test was employed, in accordance with its assumptions, to determine whether observed differences in learners' performance across stimulus types and levels of cognitive demand were statistically significant.

The participants comprised 72 Grade 11 learners enrolled in a private secondary school in Iligan City, Philippines, representing both the Science, Technology, Engineering, and Mathematics (STEM) and Humanities and Social Sciences (HUMSS) strands. Inclusion of students from these academic tracks enabled the analysis to capture diverse academic orientations and varying levels of exposure to formal chemistry content.

Data were collected using a researcher-developed stimulus-based chemistry assessment consisting of 47 multiple-choice items anchored on shared contextualized stimuli. These stimuli reflected laboratory experiments, technology-related applications, environmental and societal issues, and everyday-life situations. The instrument employed a scenario-based format, wherein multiple test items were linked to a common stimulus to elicit higher-level processes such as application, interpretation, and scientific reasoning. This structure is consistent with established stimulus-based assessment frameworks commonly used in large-scale such as PISA and classroom-based science assessments.

Assessment items were systematically classified according to competencies, stimulus type and cognitive demand. Cognitive levels were aligned with a Bloom's Taxonomy–based framework encompassing lower-order thinking skills (Remembering and Understanding), moderate-order thinking skills (Applying and Analyzing), and higher-order thinking skills (Evaluating and Creating).

The instrument underwent content and technical validation by experts in chemistry education, followed by pilot testing and item analysis to evaluate item difficulty and discrimination indices. Based on the results of the item

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue I January 2026

analysis, three items were removed, resulting in a final instrument with a Kuder–Richardson Formula 20 (KR-20) reliability coefficient of 0.86, indicating high internal consistency. After obtaining institutional approval, the finalized assessment was administered during regular class hours. Student responses were scored, and data analysis initially employed descriptive statistical measures, including mean, mean percentage score (MPS), frequency, percentage distribution, and standard deviation, to summarize overall learner performance. To examine performance patterns across chemistry competencies, contextualized stimulus types, and levels of cognitive demand, MPS values were computed for each stimulus category and cognitive domain. In addition, inferential analysis using the Friedman test, applied in accordance with its assumptions, was conducted to determine whether observed differences in performance across stimulus types and cognitive domains were statistically significant. This combined descriptive and inferential approach enabled a more nuanced examination of how contextual features and cognitive demand are associated with learner performance.

To facilitate interpretation of results, the study utilized the Department of Education (DepEd) Mean Percentage Score (MPS), a standard metric employed in Philippine large-scale and classroom-based assessments. The use of MPS was deemed appropriate because the instrument was standardized and designed to measure student performance across multiple content areas, stimulus types, and cognitive levels. Applying DepEd's established mastery level descriptors allowed.

Table I: Mastery/Achievement Level

Mean Percentage Score	Descriptive Equivalent
96-100%	Mastered
86-95%	Closely Approximating Mastery
66-85%	Moving Towards Mastery
35-65%	Average
15-34%	Low
5-14%	Very Low
0-4%	Absolutely No Mastery

In addition, the interpretation of overall learner performance was guided by the DepEd performance index system, which provides standardized descriptors for summarizing and reporting overall assessment results.

Table II: Interpretation of Learners' Performance

Index	Descriptors	Range	Interpretation
41-47	Outstanding	90-100	Passed
39-40	Very Satisfactory	85-89	Passed
36-38	Satisfactory	80-84	Passed
34-35	Fairly Satisfactory	75-79	Passed
0-33	Did Not Meet Expectations	74 Below	Failed

RESULTS AND DISCUSSION

This section presents the distribution of Grade 11 learners' scores in secondary chemistry as measured by a contextualized, stimulus-based assessment. The accompanying table summarizes learner performance across established proficiency levels, reporting the frequency and percentage of learners who met or did not meet the passing criterion, along with the overall mean and standard deviation. This descriptive overview provides a general profile of learner achievement when chemistry concepts are assessed through real-world contexts that require interpretation, application, and scientific reasoning, and serves as a foundation for the subsequent analysis and discussion of results.

Table III: Overall Assessment of Grade 11 Learners in General Chemistry

Index	Frequency	Percentage	Interpretation	Remarks
41-47	5	6.9%	Outstanding	Passed

39-40	4	5.6%	Very Satisfactory	Passed
36-38	13	18%	Satisfactory	Passed
34-35	12	16.7%	Fairly Satisfactory	Passed
0-33	38	52.8%	Did Not Meet Expectations	Failed
Total	95			
Mean 30.39		Standard Deviation 7.79	Interpretation Majority of Learners Did Not Meet Expectation	More than half of Learners Failed

The findings indicate that the overall performance of Grade 11 learners in secondary chemistry was below expected proficiency levels when measured using a contextualized, stimulus-based assessment. More than half of the respondents (52.8%) were classified under the "Did Not Meet Expectations" category, indicating that a substantial proportion of learners encountered difficulty demonstrating adequate understanding of chemistry concepts when these were embedded in real-world contexts and required higher-level processes such as interpretation, application, and scientific reasoning.

Only a limited proportion of learners attained passing performance levels, with 6.9% achieving an Outstanding rating, 5.6% classified as Very Satisfactory, 18.0% as Satisfactory, and 16.7% as Fairly Satisfactory. While these results suggest that a small subset of learners was able to effectively engage with the cognitive and contextual demands of the stimulus-based tasks, the majority struggled to integrate foundational content knowledge with contextual information. Moreover, the mean score of 30.39, which falls within the "Did Not Meet Expectations" performance band, further substantiates the overall low achievement of the cohort. The standard deviation of 7.79 reflects moderate score variability, suggesting notable differences in learners' capacity to manage the analytical and reasoning demands imposed by the assessment format.

A. Mastery of Learners Across Different Competencies

Table IV presents the mastery levels of learners across various chemistry competencies, as measured by their Mean Percentage Scores (MPS) and corresponding descriptive equivalents. The table ranks the competencies from highest to lowest based on learner performance, providing an overview of areas where learners demonstrate stronger understanding as well as competencies that require further improvement.

Table IV: Mastery of Competencies

Rank	Competencies	MPS	Descriptive Equivalent
1	Chemical Kinetics	69.60	Moving Towards Mastery
2	Chemistry in Everyday Life and Environment	66.67	Moving Towards Mastery
3	Matter and Properties	63.19	Average
4	Solutions and Bases	61.11	Average
5	Gases and Gas laws	57.50	Average
6	Thermochemistry	56.67	Average
7	Organic Chemistry	56.39	Average
8	Atomic Structure and the Atomic Table	55.28	Average
9	Chemical Bonding	54.44	Average

The results reveal that learners generally demonstrate moderate mastery of the competencies, with most areas falling under the descriptive equivalent *Average*. Chemical Kinetics obtained the highest MPS (69.60), followed by Chemistry in Everyday Life and Environment (66.67), both described as *Moving Towards Mastery*. These results suggest that learners perform better in competencies that are more application-based and closely related to real-life experiences, making the concepts easier to understand and apply.

The rest of the competencies were described as *Average*, indicating that learners have a basic understanding but have not yet achieved a high level of mastery. Matter and Properties (63.19) and Solutions and Bases (61.11)

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue I January 2026

ranked relatively higher within this category, while Gases and Gas Laws (57.50), Thermochemistry (56.67), Organic Chemistry (56.39), Atomic Structure and the Atomic Table (55.28), and Chemical Bonding (54.44) showed lower mastery levels. These topics are generally more abstract and concept-heavy, which may explain the difficulties encountered by learners.

From a curriculum perspective, these findings are consistent with concerns addressed in the Department of Education's Learning Activity Sheets (LAS) and the MATATAG Curriculum, both of which emphasize strengthening foundational understanding and supporting learners through scaffolded, competency-based activities. The predominance of average mastery levels, particularly in abstract chemistry competencies, suggests that learners may benefit from more structured learning sequences and guided practice aligned with LAS design principles. Such alignment may help learners consolidate essential concepts before engaging with more cognitively demanding, contextualized assessment tasks envisioned in the MATATAG curriculum framework.

B. Learners' Performance Across Different Stimulus

To describe how learners performed under each type of stimulus, mean percentage scores (MPS) were computed and interpreted using descriptive statistics. The use of MPS allowed for comparison of the relative level and distribution of learner performance across the four stimulus types. However, because descriptive measures alone do not indicate whether observed differences reflect consistent patterns rather than random variation, a non-parametric repeated measures analysis using the Friedman test was subsequently employed. This inferential approach was used to determine whether differences in learner performance across stimulus types were statistically meaningful within the same group of learners. Tables V and VI therefore present complementary descriptive and inferential evidence, providing a more robust basis for interpreting variation in performance across instructional stimuli.

Table V: Learners' Performances in Different Stimulus

Types of Stimuli	**Mean %**	**Interpretation**
Experiment-Based Stimuli	64.59	Average
Technology-related Stimuli	56.39	Average
Everyday Life/Material-Based Stimuli	55.42	Average
Environmental/Societal Context Stimuli	63.58	Average

Table V presents the descriptive statistics of learners' performance across four types of stimuli, namely experiment-based, technology-related, everyday life/material-based, and environmental/societal context stimuli. The results show that learners demonstrated an average level of performance across all stimulus types. Among the four, experiment-based stimuli obtained the highest mean percentage score (M = 64.59), followed closely by environmental/societal context stimuli (M = 63.58). In comparison, technology-related stimuli (M = 56.39) and everyday life/material-based stimuli (M = 55.42) yielded relatively lower mean scores, although they were still interpreted as average, able to engage meaningfully with chemistry concepts embedded in real-world environmental issues. Such contexts may enhance perceived relevance and provide intuitive conceptual cues that support reasoning, particularly for tasks requiring interpretation of data or evaluation of outcomes.

Table VI: Friedman Test and Pairwise Comparison Across Stimulus

Variables		**p-value**
Experiment-Based vs. technology-Based		< 001
Experiment-Based vs. Everyday-Life		< 001
Experiment-Based vs. Environmental/Societal		.389
Technology-Based vs. Everyday-Life		.654
Technology-Based vs. Environmental/Societal		.006
Everyday-Life vs. Environmental/Societal		.002
Friedman ($X^{2)}$ 22.2	**Df** 3	**P-value** .001

While the MPS results suggest variation in learner performance across stimulus types, inferential analysis was necessary to establish whether these observed differences were statistically significant. The non-parametric repeated measures analysis using the Friedman test was conducted, considering deviations of data from parametric assumptions. The results revealed a significant difference in learners' performance across the four types of stimuli, $p < .001$. This finding indicates that the type of instructional stimulus significantly influenced learners' performance.

Further examination through Durbin–Conover pairwise comparisons clarified the specific sources of these differences. The results showed that experiment-based stimuli led to significantly higher performance compared to technology-related stimuli ($p < .001$) and everyday life/material-based stimuli ($p < .001$). However, no significant difference was found between experiment-based stimuli and environmental/societal context stimuli ($p = .389$), suggesting that these two approaches were equally effective.

Moreover, environmental/societal context stimuli significantly outperformed both technology-related stimuli ($p = .006$) and everyday life/material-based stimuli ($p = .002$). In contrast, there was no significant difference between technology-related and everyday life/material-based stimuli ($p = .654$), indicating comparable effects on learner performance for these two stimulus types.

Given that inferential analysis indicates statistically meaningful differences in learner performance across stimulus types, the observed variation has implications for the implementation of learner-centered and inquiry-oriented curriculum goals. The relatively higher performance associated with experiment-based and environmental or societal contexts aligns with MATATAG's emphasis on experiential learning, real-world relevance, and scientific inquiry. These findings suggest that assessment tasks grounded in hands-on investigation and socio-environmental issues may be more congruent with learners' existing cognitive and experiential frameworks, as envisioned in DepEd's curriculum reforms. At the same time, the lower performance observed in technology-related and everyday-life stimuli highlights the need for careful design of contextual materials to ensure that contextual richness does not inadvertently increase cognitive load beyond learners' current capabilities.

C. Learners' Performance Across Cognitive Domains

To further examine learners' performance in relation to the level of cognitive demand, the study analyzed learners' responses across different cognitive domains based on Bloom's taxonomy. These domains were grouped into Lower-Order Thinking Skills (LOTS), Middle-Order Thinking Skills (MOTS), and Higher-Order Thinking Skills (HOTS) to determine how learners performed in tasks requiring varying levels of thinking complexity. Descriptive statistics were used to compute the mean percentage scores for each cognitive domain, while a non-parametric repeated measures analysis was conducted to determine whether differences in performance across the domains were statistically significant. Tables VII and VIII present the descriptive and inferential results of learners' performance across the different cognitive domains.

Table VII: Learners' Performances in Different Stimulus

Cognitive Domain	Mean %	Interpretation
Remembering	62.9	**Average**
Understanding	56.84	
LOTS	**59.5**	
Applying	68.3	**Average**
Analyzing	56.6	
MOTS	**61.1**	
Evaluating	62.5	**Average**
Creating	76.4	
HOTS	**64%**	

Table VII presents the learners' performance across the different cognitive domains, classified into Lower-Order Thinking Skills (LOTS), Middle-Order Thinking Skills (MOTS), and Higher-Order Thinking Skills (HOTS), based on their mean percentage scores and corresponding interpretations. Overall, the results indicate that learners demonstrated an average level of performance across all cognitive domains, reflecting a moderate mastery of the assessed skills.

For Lower-Order Thinking Skills (LOTS), learners obtained mean scores of 62.9% in remembering and 56.84% in understanding, resulting in an overall LOTS mean of 59.5%, which was interpreted as average. This suggests that learners were relatively more capable of recalling information than demonstrating deeper conceptual understanding.

In terms of Middle-Order Thinking Skills (MOTS), learners achieved a mean score of 68.3% in applying and 56.6% in analyzing, yielding an overall MOTS mean of 61.1%, also interpreted as average. The higher performance in applying indicates that learners were more adept at using learned concepts in practical contexts than at analytically examining information.

For Higher-Order Thinking Skills (HOTS), learners obtained mean scores of 62.5% in evaluating and a notably higher mean score of 76.4% in creating, resulting in an overall HOTS mean of 64%. Although all cognitive domains were interpreted as average, the relatively higher performance in creating suggests that learners were more capable of generating solutions or making reasoned decisions when contextual cues were sufficiently explicit.

Table VIII: Friedman Test and Pairwise Comparison Across Cognitive Domains

Variables		**p-value**
LOTS vs. MOTS		<.001
LOTS vs. HOTS		.001
MOTS vs. HOTS		.389
Friedman (X^2)	**df**	**P-value**
18.6	2	.001

To determine whether the observed differences among the cognitive domains were statistically significant, a non-parametric repeated measures analysis using the Friedman test was conducted. The results revealed a significant difference in learners' performance across the three cognitive domains, $\chi^2(2) = 18.60$, $p < .001$, indicating that the level of cognitive demand significantly influenced learners' performance.

Further analysis using Durbin–Conover pairwise comparisons showed that Higher-Order Thinking Skills (HOTS) differed significantly from both Lower-Order Thinking Skills (LOTS) ($p < .001$) and Middle-Order Thinking Skills (MOTS) ($p < .001$). These findings indicate that learners performed significantly better in higher-order cognitive tasks than in lower- and middle-order tasks. However, no significant difference was found between LOTS and MOTS ($p = .203$), suggesting comparable performance levels in these two domains.

Taken together, the descriptive and inferential analyses indicate that learner performance varies meaningfully across cognitive domains, suggesting that tasks differ in the cognitive demands they impose rather than reflecting uniform mastery across thinking levels. In relation to the MATATAG curriculum's emphasis on higher-order thinking skills (HOTS), the observed performance patterns across cognitive domains suggest important instructional considerations. While learners demonstrated relatively stronger performance in higher-order tasks involving creation, inferential results indicate significant differences across cognitive levels, with lower and middle-order skills remaining comparatively weaker. This pattern implies that learners may benefit from targeted and progressive HOTS development, wherein analytical and evaluative skills are explicitly scaffolded rather than assumed. Structured HOTS-oriented activities embedded within LAS, such as guided data interpretation, argumentation tasks, and stepwise problem analysis, may help bridge the gap between foundational understanding and higher-level reasoning expected in MATATAG-aligned assessments.

CONCLUSION

This study was conducted in response to the need for assessment approaches in secondary chemistry that move beyond decontextualized recall and provide clearer evidence of how learners demonstrate "knowledge-in-use" in authentic contexts. Specifically, it addressed a gap in school-based research by disaggregating learner performance across chemistry competencies, stimulus types, and levels of cognitive demand using a contextualized, stimulus-based assessment.

The results indicate that learner performance in chemistry cannot be adequately characterized through aggregate achievement indicators alone. When performance is examined across contextual and cognitive dimensions, meaningful variation becomes visible, suggesting that what learners are able to demonstrate depends not only on content mastery but also on how tasks are framed and what kinds of reasoning they require. In this sense, the study provides evidence consistent with concerns raised in the literature that contextualized assessment outcomes may reflect both conceptual understanding and item features such as stimulus structure and cognitive demand.

Inferential analyses further support the conclusion that differences observed across stimulus categories and cognitive domains are not trivial in this dataset. Rather than implying that particular contexts or cognitive levels cause better performance, these results underscore that stimulus-based assessments function differently depending on their contextual and cognitive characteristics, and that learners may experience uneven demands across task types.

Overall, the study contributes classroom-based evidence that contextualized, stimulus-based assessments can serve as diagnostic tools for identifying where learners encounter difficulty in demonstrating chemistry understanding, particularly when abstract concepts must be applied within context-rich situations. These findings are relevant to ongoing curriculum directions that emphasize higher-order thinking and contextualized learning, as they highlight the importance of aligning instructional support and assessment expectations with the conceptual and cognitive demands embedded in chemistry tasks. Future studies may extend this work by examining how instructional practices, learner characteristics, and specific stimulus features relate to performance patterns observed in contextualized chemistry assessments.

ACKNOWLEDGEMENT

Sincere appreciation is extended to the Grade 11 learners who willingly participated in the study and to the participating school in Iligan City for granting permission and support during the data collection process. Their cooperation and openness made the successful implementation of the stimulus-based assessment possible.

The researcher would also like to express sincere gratitude to Mindanao State University – Iligan Institute of Technology (MSU-IIT) for providing the academic environment and institutional support that made this study possible. Special appreciation is extended to the College of Education, whose commitment to research and academic excellence greatly contributed to the completion of this work.

The researcher likewise acknowledges the Department of Science and Technology (DOST) for the support and assistance provided, which greatly helped in sustaining the researcher's academic pursuits and research endeavors.

Finally, the researcher gratefully acknowledges all individuals and institutions who, in one way or another, contributed to the successful completion of this study. Their support and encouragement are deeply appreciated.

REFRENCES

1. Bello, J., Concon, L., Polache, M., Ayaton, M., Manlicayan, R. , Campomanes, J. & Saro, J. (2023). Contextualized and Localized Science Teaching and Learning Materials and Its Characteristics to Improve Students' Learning Performance. Psychology and Education: A Multidisciplinary Journal, *7*(1), 77-84.https://ejournals.ph/article.php?id=21051
2. Chi, S., Wang, Z., Luo, M., Yang, Y., Huang, Min., (2018). Students Progression on Chemical Symbol Representation Abilities at Different Grade Levels (Grades 10-12) Across Gender. Chemistry Education Research and https://doi.org/10.1039/C8RP00010G
3. Galymova, N. G., Mukatayeva, Z. S., Zhussupbekova, N., Orazbayeva, M. A., & Aharodnik, V. E. (2024). Methodology for integrating socio-humanitarian safety into the training of future chemistry

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue I January 2026

teachers. Journal of Turkish Science Education, 21(4), 749-774. https://doi.org/10.36681/tused.2024.041

4. Gholiyah, S. and Lutfi, A. (2021). Geometrichem Game to Improve Student Learning Autonomy in Molecular Shape Topic. Journal of Educational Chemistry (JEC), 3(1). https://doi.org/10.21580/jec.2021.3.1.7714
5. Nordine, J., Härtig, H., & Neumann, K. (2020). Contextualization in the Assessment of Learners' Learning About Science. Springer. https://doi.org/10.1007/978-3-030-27982-0_6
6. OECD, 2015. PISA 2015 RELEASED FIELD TRIAL COGNITIVE ITEMS - SCIENCE. OECD Programme for International Student Assessment 2015https://www.pi.ac.cy/keea/pisa2015/ReleasedMaterial/English/ScienceNew-2015.pdf
7. Pey-Yan and Bulut, Liou, and Okan, 2020.The Effects of Item Format and Cognitive Domain on Learners' Science Performance in TIMSS 2011. Research in Science Education https://doi.org/10.1007/s11165-017-9682-7
8. Ruiz-Primo, M. A., & Li, M. (2016). PISA science contextualized items: The link between cognitive demand and context. RELIEVE.https://doi.org/10.7203/relieve.22.1.8280
9. Seah, N.C., Lee, YJ., Ong, Y.S. (2024). The Cognitive Demands of Secondary Science Assessment Items: Refinements to a Classification Based on Semantic Gravity and Density. Springer, Singapore. https://doi.org/10.1007/978-981-97-2607-3_7
10. Sanchez, J. M. P. (2019). Indicators of Asian Achievement in Chemistry: Implications to the Philippine Setting. Kimika, 30(1), 18-30. https://doi.org/10.26534/kimika.v30i1.18-30