JKPK (JURNAL KIMIA DAN PENDIDIKAN KIMIA), Vol. 10, No.1, 2025 pp. 179-202
Chemistry Education Study Program, Universitas Sebelas Maret ISSN 2503-4146
https://jurnal.uns.ac.id/jkpk ISSN 2503-4154 (online)

Unveiling Students' Understanding of Ammonia as a Weak Base through Scaffolding-Based Chemical Reasoning Assessment

Yuda Faturohman, Endang Susilaningsih*, Harjono, Murbangun Nuswowati, Ella Kusumastuti, Dimas Gilang Ramadhani

Chemistry Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia

ARTICLE INFO

Keywords:
Ammonia Ionization;
Chemical Reasoning;
Clustering;
Scaffolding-Based Assessment;
XGBoost.

Article History:
Received: 2025-03-25
Accepted: 2025-04-23
Published: 2025-04-30
doi:10.20961/jkpk.v10i1.100779

ABSTRACT

Reasoning is a basic cognitive ability in science learning, especially in chemistry, in which students must connect macroscopic, symbolic, and microscopic levels. However, most students seem to have difficulty learning chemical reasoning, especially in the ionization of weak bases (examples: NH_3). This study uses a scaffold-based assessment to evaluate students' explanations for ammonia as a base. A paper-and-pencil test was applied to 91 first-year preservice chemistry students to test them on phenomenological, mechanical, and structural types of reasoning. Two raters rated responses, and scoring reliability was assessed using Cohen's Kappa (0.925). The data analysis consisted of descriptive statistics, correlation analysis, clustering (K-Means and t-SNE), and regression prediction with XGBoost. The results demonstrate that structural reasoning exhibits the highest level, but phenomenological reasoning has the most variation. There appears to be a high correlation between phenomenological empirical generalization and structural reasoning (r = 0.35+). Clustering outputs show three categories of students: high (R3), moderate (R2), and low (R1) reasoning, and most of the students are categorized at the moderate reasoning level, indicating some misconceptions. The XGBoost model performs well in predicting high-reasoning students but not in the moderate-reasoning classification. This paper indicates the power of scaffolding-embedded assessment for deducing reasoning patterns and misconceptions in ammonia ionization. The results can guide adaptive learning approaches for improving students' chemical reasoning.

*Corresponding Author: endangsusilaningsih@mail.unnes.ac.id
How to cite: Y. Faturohman, E. Susilaningsih, H. Harjono, M. Nuswowati, E. Kusumastuti, D. G. Ramadhani, "Unveiling Students' Understanding of Ammonia as a Weak Base Through Scaffolding-Based Chemical Reasoning Assessment," *Jurnal Kimia dan Pendidikan Kimia (JKPK)*, vol. 10, no. 1, pp. 179-202, 2025. [Online]. Available: http://dx.doi.org/10.20961/jkpk.v10i1.100779.

INTRODUCTION

Chemical reasoning is a form of critical thinking in chemistry education, which in turn helps a student develop conceptual links between the observed phenomenon and behavior at the molecular level and with symbolic representation in the domain of chemistry, which the Hong Kong postsecondary admission test for chemistry aims at [1]. The development of chemical reasoning is a key aspect of the high school chemistry curriculum, and it helps students grasp important topics like acid–base reactions, equilibrium, and molecular forces [2]. Students must fluidly navigate macroscopic observations, submicroscopic processes, and symbolic equations to develop an integrated understanding of chemical phenomena. Without these linkages, knowledge becomes fragmented, and misconceptions can interfere with scientific reasoning. Similarly, proficient chemical thinking also supports problem solving, critical thinking, and other forms of engagement with more complex scientific ideas at the college level. Strong chemical reasoning is, therefore, recognized as a critical building block of meaningful learning of chemistry and the development of science literacy. However, despite its centrality, students' difficulties creating coherently integrated, scientifically accurate explanations remain a key curricular challenge in chemistry education.

One of such ideas, which provides continuous trouble to students, is regarding the behavior of ammonia (NH_3) as a weak base. The notion that all bases necessarily generate high pH values is a common misunderstanding due to confusion about solution ionization strength and equilibrium behavior [3], [4]. Students tend to take an overly condensed view of the term basicity on pH, without considering the molecule's level of ionization or nature. This is shallow reasoning and suggests ignorance of the mechanism of the impact of molecular structure on chemical behavior. Moreover, didactic methods that emphasize algorithmic problem solving at the expense of conceptual inquiry can continue to support superficial learning and the failure of learners to understand the subtle behavior of weak bases and ammonia in particular [5]. Learners, however, often do not appreciate that when weak bases react with water, reversible equilibria result, meaning that there are lower concentrations of hydroxide than in strong bases and only partial acceptance of protons. Consequently, they explain too superficially, without using the structure of explanations, and they show rather persistent alternative conceptions, not responsive to traffic instruction.

To resolve these ongoing challenges, scaffolding-based pedagogical approaches have received growing interest in the science education literature. Scaffolding is defined as the instructional support, adapted to the learner's needs and knowledge, and coordinated with the instruction that is going on, whereby the support is progressively removed as students become more independent [6]. In reasoning with chemical phenomena, scaffolding might help students make sense of the demanding connections required between their observations, the particulate behavior, and symbolic representations. Scaffolding-based assessments, such as these, are especially useful as they enable step-by-step cognitive support, enabling students to systematically engage with phenomenological, mechanistic, and structural levels of thinking. In this way, students can be helped to develop more complex explanations that include evidence, causal mechanisms, and molecular structures. However, despite increasing interest, empirical studies on how scaffolding affects students' transitions from these reasoning dimensions are scarce, especially in explaining concretely observable weak base phenomena, such as ammonia ionization [5], [7]. Knowledge of this process is needed to develop effective interventions that foster more successful conceptual change in chemistry.

Meanwhile, although diagnostic tools have been designed for identifying misconceptions in chemistry education [8]-[10], few studies have taken machine learning approaches to categorize and compare reasoning profiles of students. Novel machine learning methods such as XGBoost provide significantly more powerful tools to model cognition development and predict reasoning outcomes from student responses. Combining scaffolding-based assessments with predictive modeling improves the affordance for diagnosing conceptual difficulty and creates new opportunities for personalized adaptive support. Machine learning methods can detect hidden structures in student thinking, which is not easily revealed from traditional assessments, and offer a deeper understanding of learning trajectories [11]-[15]. In addition, due to real-time feedback and analysis, predictive models are well-suited to dynamic classroom settings. Although these benefits exist, chemistry studies combining scaffolding supports and machine learning analysis are also limited. This discrepancy emphasizes the necessity of research integrating theoretical scaffolding, analysis of levels of reasoning development, and predictive modeling [16].

In this context, this study explores students' chemical reasoning patterns about the ionization of ammonia as a weak base. In particular, this paper aims to investigate the connections between phenomenological, mechanical, and structural reasoning levels and common misconceptions related to the weak base character of ammonia. Finally, the effectiveness of the scaffolding-based assessments in conjunction with XGBoost classification for revealing students' cognitive profiles and reasoning transitions will also be tested. Through the use of such an integrated approach and iteratively interacting with it, a deeper understanding of the dynamics of chemical reasoning is believed to arise, which can ultimately lead to more targeted and flexible instructional strategies. Finally, the outcomes are expected to add to the ongoing debate on data-driven innovation in chemistry education in the wider community. This work stresses integrating theoretical scaffolding and machine learning techniques to treat classic, challenging chemical education reasoning questions. In the end, the project aims to contribute to developing evidence-based pedagogical resources for improving students' understanding of and reasoning about high school chemistry.

METHODS

1. Research Design

The study used a descriptive quantitative research design relevant to systematically document and analyze students' chemical reasoning patterns when explaining the ionization of ammonia (NH_3) as a weak base. While the quantitative side makes reasoning measurable in a large dataset, the descriptive side enables researchers to identify new cognitive profiles by applying clustering, correlation studies, and predictive modeling. This aligns with the study's purpose of diagnosing, categorizing, and foretelling students' reasoning levels from empirical data.

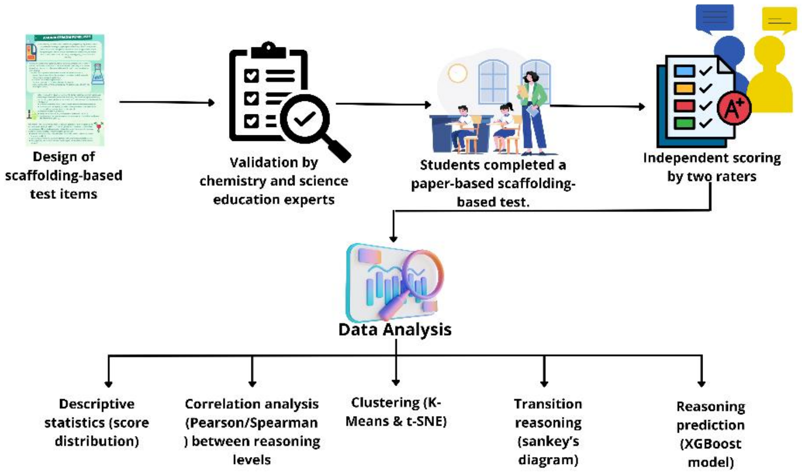

Figure 1. Research Design.

2. Participants

The participants were pre-service teachers in their first year, consisting of 91 people at one of the universities in Semarang City, Indonesia. All the students had been instructed on concepts of acid-base and the Brønsted–Lowry theory in prior studies; hence, they had acquired the basic knowledge necessary for performing chemical reasoning tasks. A convenience sampling method was used, and the participants were recruited from already identified classes in the school. Although not completely random, the sampling was designed to represent a broad range of student abilities and cognitive levels to support deeper analysis of reasoning patterns.

3. Instrument Development

A five-item scaffolding-based paper test was prepared to measure three levels of chemical reasoning: phenomenology, mechanism, and structure. Item development was based on extant theoretical frameworks in previous literature [1], [4], and was consistent with representational reasoning theory. The steps of the activities were scaffolded to help students move from concrete to abstract thought.

Two-stage validation of the instrument was performed by three experts in the field of chemistry (two chemistry education researchers and a senior chemistry teacher, who also has a background in curriculum development). Expert feedback informed the rewording of questions, scoring rubrics, and the logic of scaffolding, thus ensuring the appropriateness and legibility of the content.

4. Data Collection

The data were obtained with a paper-based test constructed within a scaffolding-based assessment frame. This instrument included five task items that progressively led students from simple observations to more complex abstract reasoning along the phenomenological, mechanistic, and structural levels (as identified in Table 1).

The test was conducted under controlled classroom conditions in which the students were taking their regular classes and had a time limit of 60 minutes to complete the test. Examinations were proctored by the classroom teacher and by a research team member to standardize the conditions under which the test was administered and reduce extraneous distractions.

Students were asked not to complete the task using their textbooks or digital resources. Responses were de-identified and collected directly following the session. Responses that were unclear or incomplete were excluded from analyses conducted for individual item responses, but remained in the dataset when calculating reliabilities, as appropriate.

5. Data Analysis

Data analysis was performed in multiple steps to gain deeper insights into students' ideas of chemical reasoning. First, descriptive statistics (mean, median, standard deviation, and range) were estimated for each reasoning category (phenomenological, mechanistic, and structural). The distribution of scores was presented with histograms and boxplots to visualize trends and anomalies [11], [12].

Pearson correlation tests were performed to examine the correlations between reasoning categories. Data assumed to be normal was tested using Pearson's test. A heatmap was created to show the strength and direction of these relationships [13].

Table 1. Chemical reasoning test instrument

Chemical Reasoning	indicator	Example Questions
Phenomenological Referring to insights gained empirically about substances and chemical reactions.	Classification Systems	NH_3 is recognized as a weak base. According to the Brønsted-Lowry theory, explain how NH_3 functions as a base in its reaction with water.
	Empirical Generalization	A 0.01 M solution of NH_3 has a pH of approximately 11. Calculate the concentration of OH^- ions in this NH_3 solution.
Mechanical Focused on evaluating the dynamics of each model component and the resulting changes in their positions and motions	Static-Deterministic	When NH_3 dissolves in water, the NH_3 molecule, which has a lone pair of electrons on the nitrogen atom, accepts a proton from a water (H_2O) molecule. This reaction results in the formation of NH_4^+ and OH^- ions, thereby increasing the concentration of OH^- ions in the solution, making it essential. The lone pair of electrons on the nitrogen atom in NH_3 allows it to bond with a proton (H^+) from the water molecule, leading to the formation of NH_4^+ and OH^- ions.
	Dynamic-Probabilistic	When the temperature of the NH_3 solution is increased, how does the temperature affect the movement of NH_3 and H_2O molecules? Explain how the rise in temperature impacts the probability of effective collisions between NH_3 and H_2O.
Structural Based on an examination of the makeup and structure of chemical compounds shown through structural representations	Structural	NH_3 has a trigonal pyramidal structure with a lone pair of electrons on the nitrogen atom. This lone pair allows NH_3 to accept a proton from water. However, because NH_3 is a weak base, ionization does not occur completely, resulting in a solution with a pH above seven but not as high as a strong base. Describe the structure of NH_3 in its neutral state and after accepting a proton (forming NH_4^+). Please explain how the trigonal pyramidal structure of NH_3 influences its basic properties.

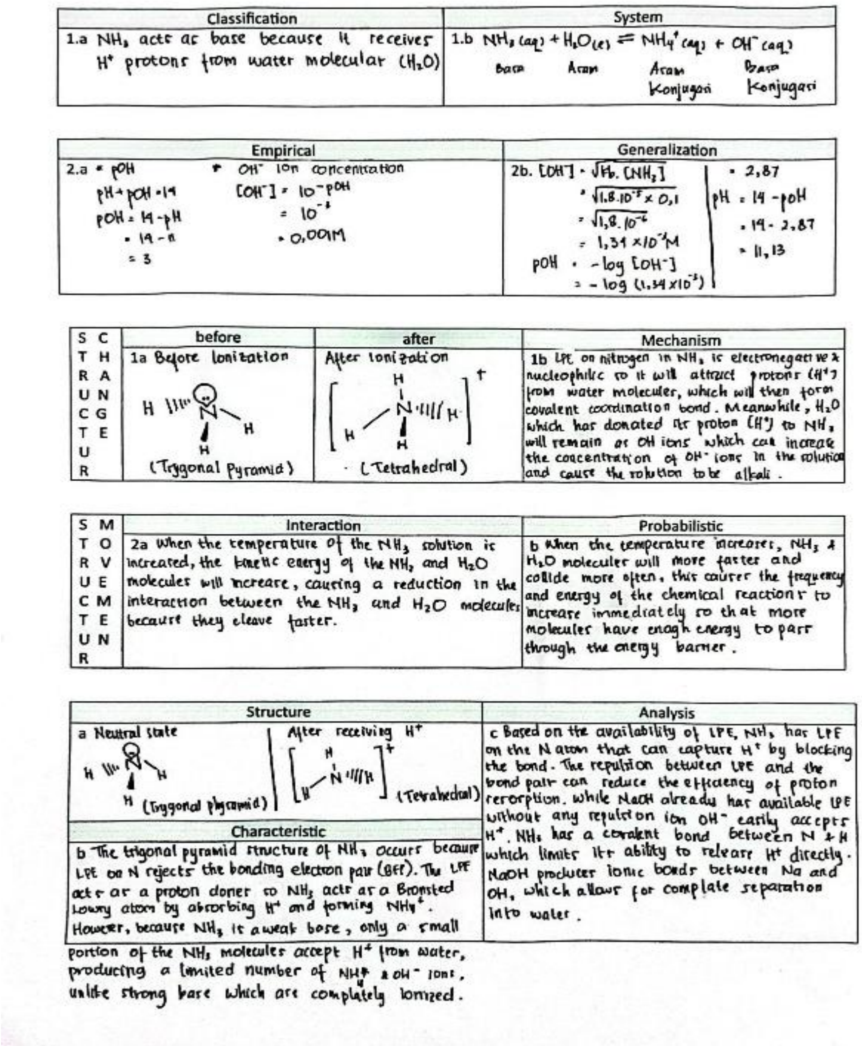

Figure 2. Paper-Based Scaffolding Assessment

To help identify patterns, K-Means clustering was used to cluster students into low, medium, and high reasoning groups. The clustering results were visualized using dimensionality reduction through t-distributed stochastic neighbor embedding (t-SNE) to facilitate visualization and interpretation in a low-dimensional representation. The choice of these tools is based on their widespread use for unsupervised learning and dimension reduction in educational data [14].

The flow of students across levels of reasoning was analyzed using Sankey diagrams. This technique was selected to graphically highlight how moves were used and how ideas evolved from one category to another, providing an image of the structure of students' movement among categories of justification [15].

A predictive model, based on an XGBoost (extreme gradient boosting) algorithm [16], was configured to leverage the robustness and relatively high performance for small-to-medium scale educational datasets. The input features in the model were the students' scores on each of the reasoning domains. Performance of the model was analyzed in terms of confusion matrices and feature importance plots to determine the accuracy and the reasoning categories that most influenced classification.

All statistical and machine learning analyses were performed in Python (pandas, seaborn, scikit-learn, XGBoost) [17].

6. Validity and Reliability

a. Inter-Rater Reliability

To establish scoring consistency and objectivity, two trained raters rated students' solutions independently using an instrument designed to operationalize phenomenological, mechanistic, and structural reasoning levels. The scoring raters participated in a standardization session with sample responses before scoring to ensure consistent interpretation of the rubric criteria [18].

The inter-rater reliability was determined to be good with Cohen’s Kappa coefficient of 0.925. According to Landis and Koch (1977), this value indicates almost perfect agreement, supporting the reliability and robustness of the scoring process in all reasoning dimensions [19].

b. Construct Validity

The face validity of the scaffolding-based assessment instrument was established through expert validation and pre-testing. Three chemistry education experts (two from a university and one from a senior high school) were invited to examine each test item's cognitive alignment with the chemical reasoning framework. Two rounds of iterative validation were applied, with feedback on both the appropriateness of the content in terms of cognitive level and the clarity of the scaffolding structure. Adjustments were made accordingly to improve item quality and alignment.

A pilot test among a similar group of high school students was also conducted to ascertain the clarity, timing, and level of difficulty of the instrument. Pilot feedback influenced final revisions before full implementation, ensuring the instrument validly captured students' representational reasoning dimensions [20].

RESULTS AND DISCUSSION

1. Distribution of Reasoning Score

The scoring of the analysis of students’ reasoning according to the five chemical reasoning aspects provides a complete picture of their cognitive comprehension, mainly for explaining the ionization of ammonia. As presented in Figure 3, participants presented the best performance on Phenomenological System Classification with a significant peak of score 5.0 representing performance levels on classifying observable chemical phenomena [21].

This trend is also reflected in Table 2, where this domain has an average of 4.75, a median of 5.0, a S.D. of 0.58, and a skewness of -3.10, indicating that nearly all students achieved the maximum score on the scale. These performances were possible due to learning these chemical activity instruments in chemistry education, using indicators for the recognition of acids and bases [22]–[24].

Meanwhile, although still unimodal and right-leaning, Structural Reasoning has a lower central tendency with mean = 4.03, median = 4.0, and IQR = 0.0, as reflected in Figure 4 and Table 2. This indicates that even though most students can use chemical formulae and molecular structures more or less equally at a medium-to-high level, they hardly arrive at the highest level of representational integration. The absence of dispersion may also suggest overscaffolding or mechanistic symbolic procedure, but not genuine conceptual development [25].

Table 2. Distribution Score

Indicator	Mean	Median	SD	Skewness	IQR
Phenomenological system-Classification	4.75	5.0	0.58	-3,10	0.0
Phenomenological Empirical-Generalization	4.62	5.0	0.54	-1,29	0.75
Mechanistic static-Deterministic	4.18	4.0	0.58	-0.29	0.5
Mechanistic Dynamic-Probabilistic	4.48	4.5	0.65	-1.63	1.0
Structural-Structural	4.03	4.0	0.43	-1.46	0.0

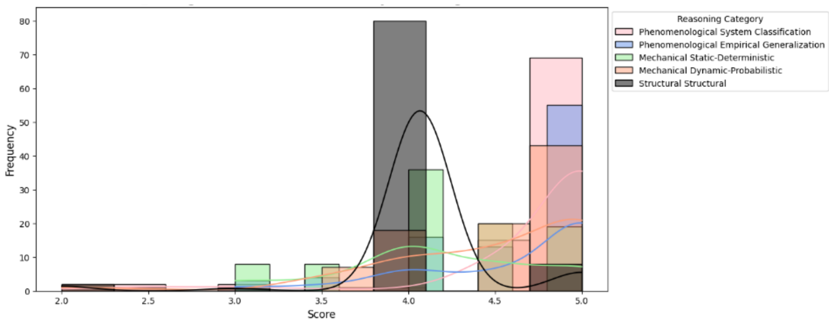

Figure 3. Histogram of Score Distribution

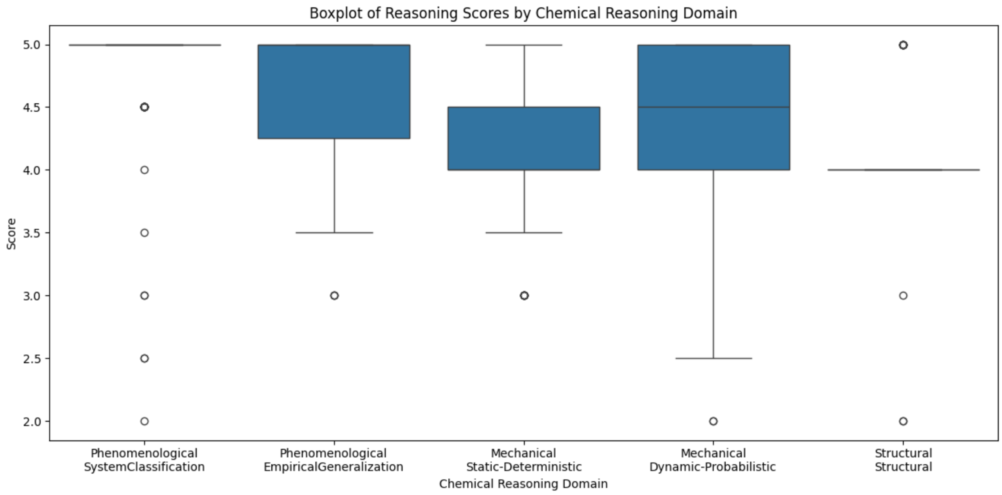

Figure 4. Boxplot Comparison of Reasoning

Outside of these areas of stability, Phenomenological Empirical Generalization entails greater cognitive diversity. Although this result also has a median of 5.0, its IQR (0.75) and skewness (–1.29) indicate that for some students it is possible to generalize the pattern they observe, while for others it is not. This result is congruent with past work, revealing the tendency of many students to be good at surface-level trend recognition but facing generalization challenges, especially when teaching does not involve explicit demonstration of inductive reasoning [3].

The mechanistic justification types exhibit the most distinguished pattern in students' cognitive patterns. Mechanistic Static-Deterministic Reasoning has a mean of 4.18, median of 4.0, IQR = 0.5, and similar slight negative skewness (–0.29), reflecting consistent procedural reasoning, typical of algorithmic-type work like calculating pH or predicting products. The students seem to rely on straightforward or rote rules but may lack an understanding of causality.

In contrast, Mechanistic Dynamic-Probabilistic Reasoning shows the greatest range of scores (mean = 4.48, IQR = 1.0, skewness = –1.63). This trend, as observed in Figure 4, indicates a distinct cognitive barrier in students' capacity to model chemical systems' reversible, inconclusive, and time-related nature, such as the equilibrium reactions participating in weak base ionization. These activities go beyond mechanical rule application and require conceptual flexibility for dynamically relating symbolic, macroscopic, and submicroscopic representations [26], [1].

These results support Talanquer's (2022) representational learning model, stating that chemical reasoning develops from recognition at the phenomenological level to explanation at the mechanistic level and finally reaches abstraction at the structural level. Yet the data also show that this growth is not linear or homogeneous. Some students may excel at classification but not generalization or dynamic reasoning, which indicates incomplete understanding [27], [5].

From an instructional perspective, the findings may underline the need for domain-adequate scaffolding. Although tasks involving multiple representations may be well-established in today's curricula, students need explicit guidance for transfer within cognitive tools, such as scaffolding in the form of structured observation and explanation cycles, dynamic visualizations, or simulations to facilitate the understanding of the constructs of equilibrium and reversibility, and cross-representational activities to establish the connection of formulas, particle models, and macrological outcomes.

The strong negative skewness and large IQR in empirical and dynamic domains indicate the demand for flexible instructional interventions in these domains because they are critical leverage points for orchestrating a shift from descriptive to analytical reasoning, provided that appropriate cognitive tools and pedagogical support are given [5].

2. Correlation Between Reasoning Levels

The Pearson correlation studies of Figures 5 and 6 present the degree of the relationship of students' scores over five chemical reasoning domains. Although all are positive, their strengths and statistical significance vary, suggesting important differences in how students link different kinds of reasoning. The most highly correlated pair is Phenomenological System Classification and Phenomenological Empirical Generalization ($r = 0.50$, $p < 0.001$), which implies that chemistry learners with high proficiency in chemical phenomena classification are likely to perform well in forming generalizations of empirical data. This profile illuminates a robust developmental connection between surface-form recognition and inductive reasoning—a key precursor of conceptual understanding [28], [29].

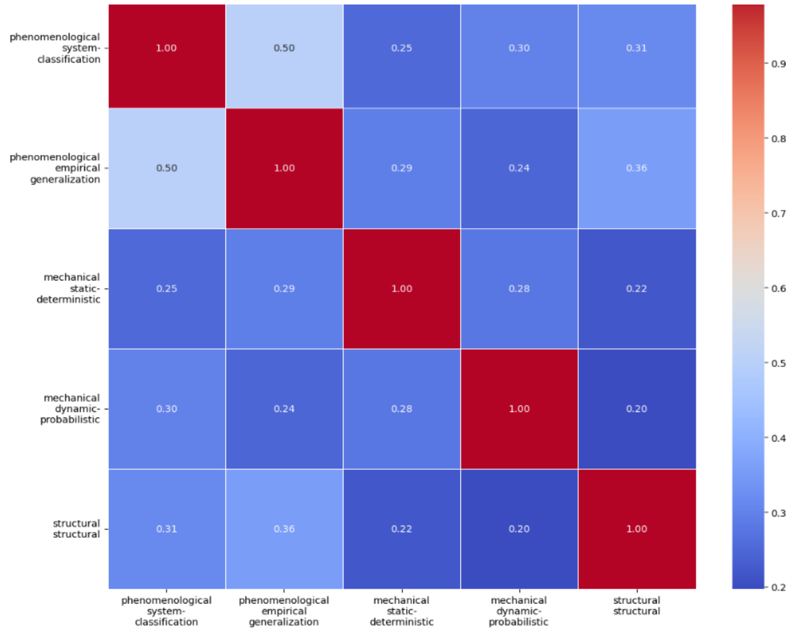

Figure 5. Pearson Correlation Heatmap Between Reasoning Levels

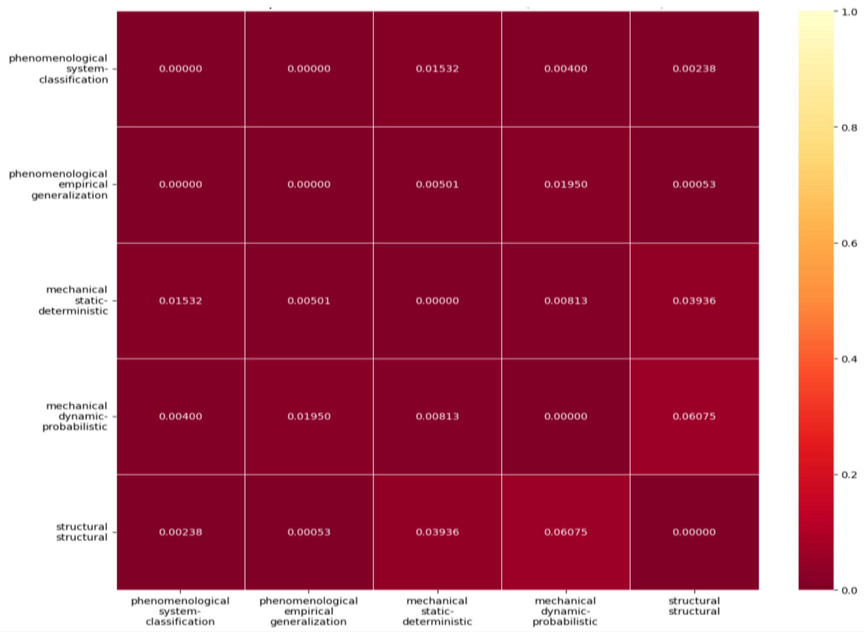

Figure 6. p-value Pearson Correlation Heatmap

A moderate and statistically significant relationship is also found between Empirical Generalization and Structural Reasoning (r = 0.36, p = 0.0005), suggesting that students who can generalize from data are likelier to link those patterns to symbolic and molecular level representations. This indicates an increasing representational proficiency, where students recontextualize macroscopic observations on a level of the submicroscopic and symbolic, as it is described for example in Johnstone's triplet model and Eduktion, and is also supported by research on representational fluency [28], [4].

A less conclusive relationship is found between mechanistic reasoning and structural reasoning. The relationship between Mechanistic Static-Deterministic Reasoning and Structural Reasoning, r = 0.22 (p =0.0393), and the correlation between Mechanistic Dynamic-Probabilistic Reasoning and Structural Reasoning, r = 0.20, is not significant (p = 0.0607). These results suggest that the students who reason causally about chemical processes, especially reversibility and uncertainty, may not translate this understanding into symbolic or structural representations [28].

This separation suggests that mechanistic reasoning is not implicitly linked to representational reasoning unless taught in an integrated manner. Students might understand what happens during a chemical change but struggle to describe it with structural pictures, chemical formulae, or symbolic representation. This fragmented view can hinder students' capacity to work with open-ended tasks and transfer ideas across settings [30].

The conclusion to be drawn from these findings is that students are quite competent when working with stories (phenomenological reasoning). Still, they struggle with mechanical reasoning, especially bringing these two representations together. Consistent with that argument, we find that mechanistic reasoning, especially of a dynamic nature, is more cognitively challenging and does not receive as much instructional emphasis. The weak or nonexistent connections between mechanistic and structural reasoning domains suggest that the latter cannot be treated as stand-alone skills, but must be seen instead as part and parcel of a larger frame of scientific understanding [1], [5].

The above cognitive cleavages must be addressed consciously to bridge the gap between mechanistic and structural thinking for the learners. A successful strategy is to provide instruction based on multiple representations (going back and forth between particle diagrams, reaction mechanisms, and symbolic equations). This enables them to observe how microscopic changes are expressed in macroscopic representations. Also, scaffolding causal explanations (why, how) of a chemical process to structural representations (e.g., molecular structure, or a balanced equation) can help children overcome conceptual hurdles. The visual explanation through molecular animations and visual simulations is also necessary to help students visualize how the structure of the substances changes during the chemical reactions, even in complex situations such as the ionization of weak bases and dynamic equilibria. Without these focused instructional supports, students can persist in performing relatively well on discrete reasoning tasks. Still, they struggle to achieve the deeper integration required for advanced scientific problem solving and conceptual transfer [4], [31].

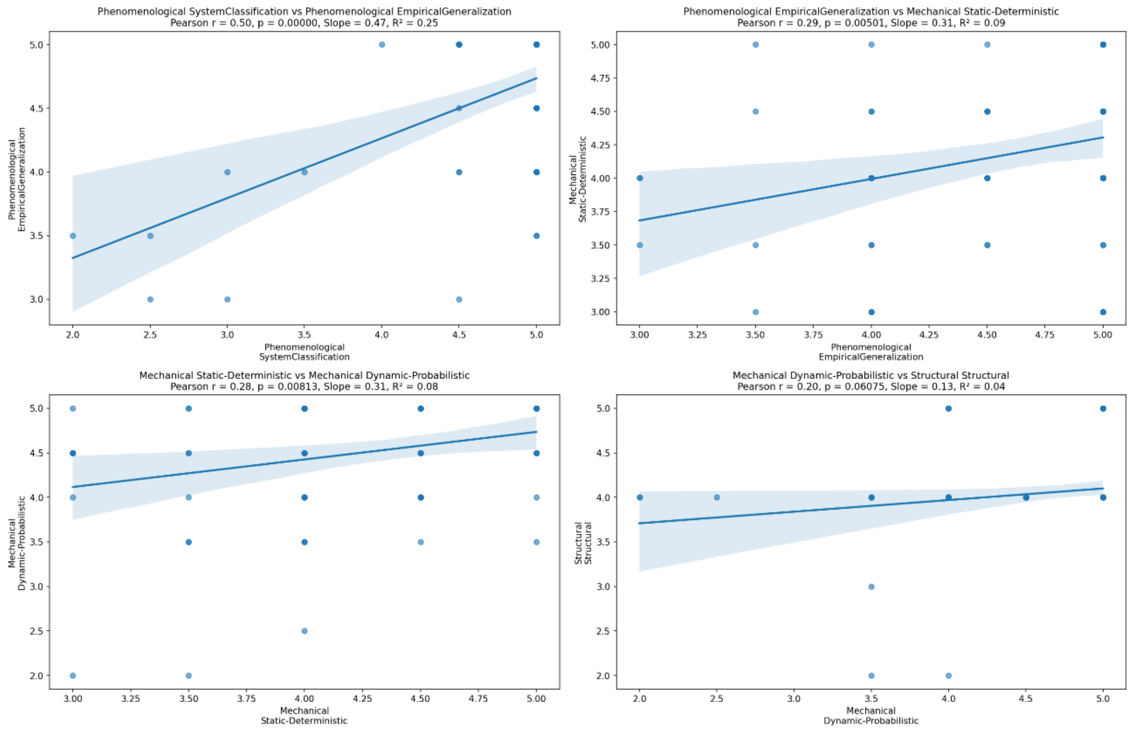

Figure 7. Scatter Plot of the Relationship Between Reasoning Categories

To support and visually strengthen the conclusions drawn in the Pearson correlation heatmap, we show in Figure 7 four scatter plots involving regression lines, the corresponding slopes, and R^2 values. These plots further demonstrate relations between the reasoning constructs and confirm the nature and magnitude of the associations identified in the heatmap earlier (Figure 5).

The upper left scatterplot between Phenomenological System Classification and Empirical Generalization reveals the tightest and consistent relationship between any pair of domains. The regression line has a slope of 0.47, $r = 0.50$, $p < 0.001$, and $R^2 = 0.25$, with classification ability accounting for 25% of the variance in students' empirical generalization scores. This map aligns with the heatmap, exhibiting a strong and moderate positive correlation in these two domains. The concentrated data points and fast upward trend also make the point that the ability to classify systematically goes with the ability to make empirical generalizations, supporting a developmental connection between recognition and inductive reasoning [32].

The top-right scatter plot, in which we examine the relationship between Empirical Generalization and Static-Deterministic Mechanistic Reasoning, reveals an even weaker trend ($r = 0.29$, slope = 0.31, $R^2 = 0.09$). The relationship is statistically significant ($p = 0.005$), but with the large spread and small slope, we again see what we noticed in the heatmap. A weak relation exists between students' ability to generalize from data and their procedural, rule-based reasoning. This implies that procedural proficiency develops independently from inductive reasoning and that success in algorithmic tasks does not necessarily imply a deep conceptual understanding [30].

The bottom-left graph (Static-Deterministic vs. Dynamic-Probabilistic Reasoning) also presents a small-to-moderate but significant ($r = 0.28$, $p = 0.008$, $R^2 = 0.08$). This is also in line with the heatmap and means that some successful students in the case of fixed, rule-based reasoning would start interacting with complex, probabilistic processes, though not surely nor reliably. The data points also reveal a great diversity of dynamics in the students' transition from static to dynamic mode [33].

Lastly, the bottom-right scatter plot that connects Dynamic-Probabilistic Reasoning to Structural Reasoning evidences the weakest correlation ($r = 0.20$, $p = 0.0607$, slope = 0.13, and $R^2 = 0.04$). That would be a flat regression line, and the data would be all over the place. The lack of significant correlation directly opposes this finding in the heatmap. It suggests that the visualization of students' capacity to reason about dynamic chemical processes (equilibrium, etc.) is not well related to their use of the symbolic or structural representations. This disparity represents a cognitive and pedagogical chasm [34].

The scatter plots and correlation matrix confirm this pattern: the development of students' reasoning entails some structured development (e.g., from classification to generalization, from static to dynamic causality), but these developments do not automatically lead to the development of their structural knowledge. Both sorts of analysis give us the same insight: structural reasoning is poorly integrated, particularly when combined with higher-order mechanistic reasoning.

This triangulated approach between statistical correlation and visual regression supports the robustness of our findings and overcomes the limitation of a single evidence. There is evidence for the same issue: students can reason well within domains but struggle connecting between them, especially between mechanistic reasoning and structural reasoning [35], [36].

3. Clustering of Reasoning Patterns

The elbow method presented in the WCSS plot (Figure 8) also supports that k=3 is the optimal number of clusters to choose. The approach shows a steep decrease of WCSS with an increase in k from k = 1 to k = 3, where there are few added gains. This corroborates the three-cluster solution, which combines computational efficiency with interesting data partitioning [37]. Adding clusters beyond k = 3 does not significantly improve clustering outcome, as indicated by the elbow in the plot, validating our choice to cluster students into three separate reasoning groups. These results highlight the need to integrate cluster analysis to inform instructional decisions, provide a more complex picture of students' reasoning abilities, and provide a foundation for growth in an organized learning environment [38].

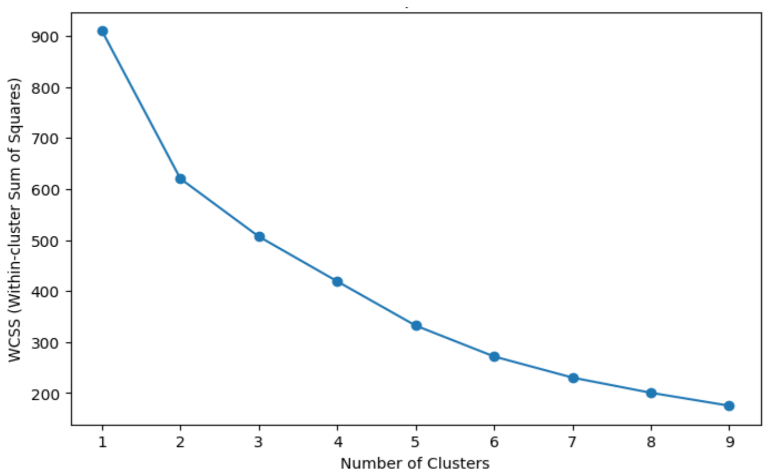

Figure 8. Elbow Method for optimal Cluster Selection

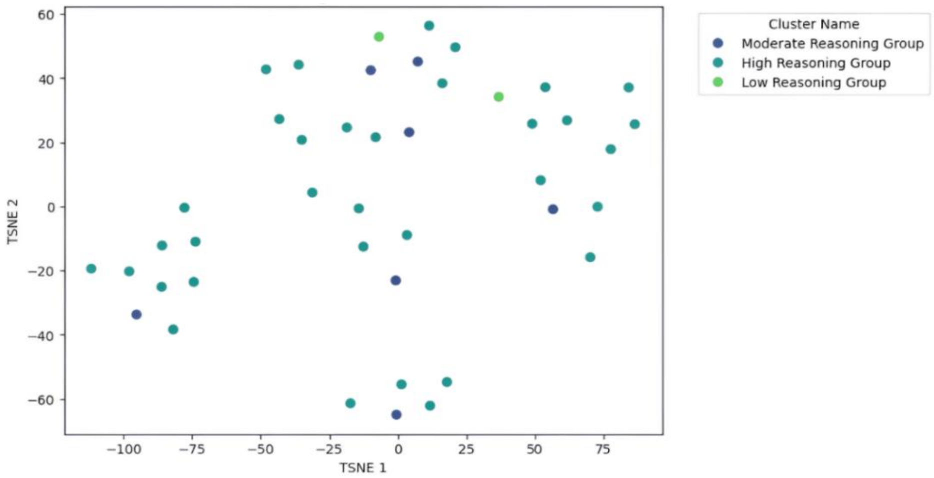

Figure 9. t-SNE Clustering

K-means clustering analysis, illustrated in Figure 9 through t-distributed Stochastic Neighbor Embedding (t-SNE), divides the 91 students into three different reasoning groups, namely the High Reasoning Group (teal), the Moderate Reasoning Group (dark blue), and the Low Reasoning Group (light green). The High Reasoning Group shows students who have consistently high-level abilities in each of the phenomenological, mechanistic, and structural reasoning categories, reflecting a well-developed ability to interrelate empirical observations, mechanistic logic, and structure-based representations—strands considered to be critical assets to fostering expertise in complex chemical ideas [39]. The tight grouping of this cluster indicates good cognitive stability and conceptual maturity [35], [36]. Conversely, the Moderate Reasoning Group shows prototypical cognitive profiles in transition, with some reasoning aspects well developed, others left weak. Their proximity to the cluster boundaries indicates the possibility of passing or failing according to the instructional support [40]. The Low Reasoning Group identifies students with low reasoning ability for whom successful interventions will be targeted and intensive.

Such overlap between the Moderate and High Reasoning Groups underscores the developmentally plastic nature of reasoning, and the importance of individualized teaching plans that reinforce weaknesses while spinning the gains of students into a leverage for their advance toward a more coherent and advanced reasoning [1], [5].

These results have critical implications for the chemistry education teaching and learning practices. The first observation from the clustering analysis is that students' logical reasoning ability is not homogeneous but falls into different clusters with different performance levels. This indicates that there is a need for an educational approach based on their knowledge level. High Reasoning Group: They should continue to stretch and deepen their thinking by giving them complex, real-life situations of chemical concepts. The Moderate Reasoning Group should focus on bolstering weak reasoning areas by scaffolding targeted activities, such as probe-and-improve tasks and relating empirical evidence to theoretical representations. For the Low Reasoning Group, extensive interventions are necessary, including: diagnostic tests to detect errors focusing on misconception, and instructional design centered on cognitive development of that reasoning foundation area [1], [4], [5].

4. Reasoning Transition Between Levels

Fig. 10 Sankey Diagram: flows of students' reasoning profiles from a narrow category (Structural Reasoning, Phenomenological Empirical Generalization, Phenomenological System Classification, Mechanical Static-Deterministic, Mechanical Dynamic-Probabilistic) to the general reasoning domain and then to one of the proficiency levels: High, Moderate, Low. The clearest, most direct bounces into High Reasoning come from the Structural Reasoning and the Phenomenological. In particular, in the transition from Structural Reasoning into the High Reasoning category, 88 students appear, and 78 move from the Phenomenological System Classification. This is supported by the observation that students who can generalize from empirical observations or classify phenomena phenomenologically in this way, who demonstrate strong representation at the structural level, are the most capable of achieving advanced chemical reasoning [15]. With this Archimedean point, phenomenological reasoning becomes a cognitive bridge facilitating learning towards mastery in structural reasoning [1].

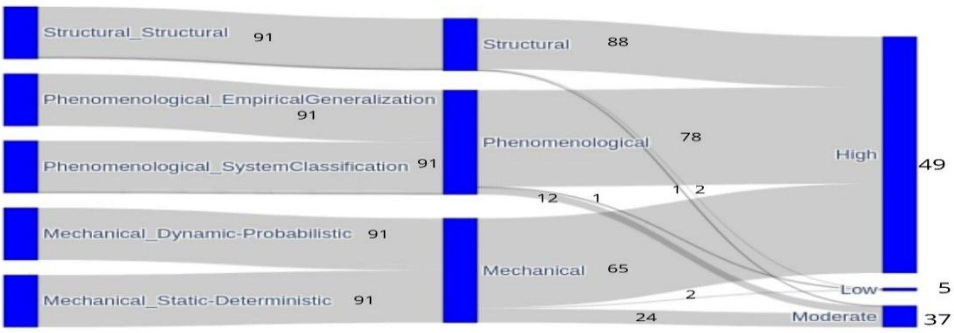

Figure 10. Sankey's Diagram of Reasoning Transition

The mechanical reasoning, static-deterministic, and dynamic-probabilistic paths undergo more diffuse shifts in contrast. There are 24 more students in the Mechanical category who enter the Moderate group, and the students of the Mechanical strength are 65 to enter the High Reasoning group. This suggests that it is not the type of procedural problem solving, but rather the purposeful inclusion of conceptual integration that promotes the development of higher order reasoning. Although minor, their flow into the Low Reasoning group comes principally from mechanical pathways with no conceptual base. It emphasizes to what extent procedural content can appear to become stagnant if not bound to construct frameworks (cf. [41]). For instance, only two students make a transition related to Mechanical Static-Deterministic and Mechanical Dynamic-Probabilistic reasoning, confirming then the notion that, absent the integration of dynamic or procedural reasoning and structural reasoning together, students could be experiencing considerable cognitive limitations [42].

These trends highlight the need for pedagogical strategies that combine higher-order empirical, procedural, and structural reasoning skills to help students progress above a moderate level of reasoning and develop a more complete view of chemistry. The challenge is not just to offer instruction in such a way (scaffolded learning environments) that students can bridge the gap from phenomenological observations through mechanistic reasons and symbolic forms, but to ensure that for all procedural knowledge mechanisms there is an equivalent domain of conceptual and structural reasoning [1], [42].

5. Dominant Misconception

Examining students' reasoning patterns (Figure 9) showed the presence of prevailing erroneous views on the ionization of ammonia. A large number of students who were placed into the moderate or low reasoning groups seemed to think that all bases had high pH, without sufficiently taking the ionizing power of the substance into account. This misunderstanding was particularly evident in students who demonstrated reasoning based more on procedural and mechanical reasoning, as indicated by the clustering and transitioning analyses [43].

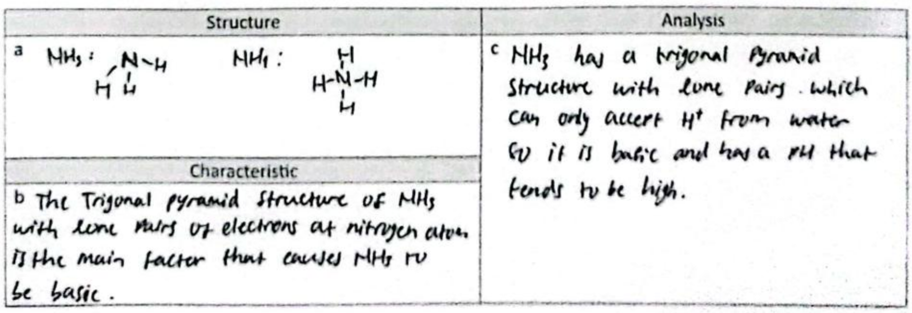

Figure 11. Example of a Student's Response Illustrating Misconceptions about Ammonia Ionization.

Certainly, students in the lower reasoning cluster seemed to recall the basic character of ammonia mostly around the pH, not offering sometimes an explanation of the incomplete ionization typical of weak bases. Their responses often revealed a naive level of presentation of the acid-base concept, which focused mainly on macromechanical behaviour, but did not consider submicroscopic and symbolic representations. This trend is consistent with previous studies that reported identical misconceptions in the acid-base chemistry [43], [44].

Furthermore, the Sankey diagram (Figure 10) interpretation revealed that a large proportion of students with poor structural reasoning skills could not move up towards higher levels of reasoning, and the naive idea of universal high pH values for bases might have affected their progress towards a more embedded, integrated understanding of chemical phenomena [35].

A student's response in Figure 11 provides an illustrative case of this misunderstanding. He correctly describes the molecular shape of ammonia (trigonal pyramidal) because of the presence of the lone pair on nitrogen. Still, he wrongly infers that from the shape he can predict the base strength and relate it to a high pH of 10 (aq). This is a conceptual error on the degree of ionization since ammonia is a weak base and ionizes partially in water. These misconceptions were not rare: 5.5% of students in our sample also displayed reasoning errors, such as considering that all bases had high pH, independently of their basicity. This misguided generalization is due to premature access of strong bases such as NaOH, limited treatment of its equilibrium context and inappropriate models of acquisition which do not consider representations and integration of molecular concepts [45].

These results confirm the necessity of explicit instruction, which forces students to differentiate between strong and poor bases and incorporate macroscopic, sub-microscopic, and symbolic levels, so that students can relate structure, behavior of ionization, and the observed chemical properties in a consistent and scientifically accurate way [46].

6. Predicting Reasoning Categories

A predictive analytics approach was undertaken using the XGBoost model to group students into high, moderate, and low reasoning categories for the phenomenological, mechanistic, and structural reasoning scores [16]. Students were categorised as high, moderate, or low reasoning students based on broad phenomenon, mechanistic, and structural reasoning classifications using a predictive analysis approach with an XGBoost model. As can be observed from Figure 12 and 13, the accuracy of the model is extremely good with 95% as overall accuracy and excellent accuracy to discriminate students within High Reasoning group. It achieved precision, recall and F1-score of 0.95, 1.00 and 0.97 for this group, indicating that the model correctly classified all high-reasoning students without errors—a good show of its diagnostic capability in identifying well-developed reasoning profiles [47].

But the model performed zero precision, recall, and F1-score for that category for students of Moderate Reasoning when they were classified. This suggests that all Moderate category students were incorrectly classified as either High or Low, which is a core problem for machine classification of transitional cognitive profiles. The macro-average F1-score (0.49) accounts for this imbalance but the weighted average (F1 = 0.92) is biased by the large number of correct high-level classifications [47], [48].

This performance difference indicates that moderate reasoning is, by nature, more ambiguous and structurally domain-adjacent. It also raises the questions of class imbalance and collinearity between some features, preventing the model to learn "pure" decision boundaries. These constraints suggest the necessity for higher-quality feature designing, potential reduction of dimensionality, or even some adopting sampling strategies (e.g., SMOTE) to achieve better classification credibility of all groups [5], [47].

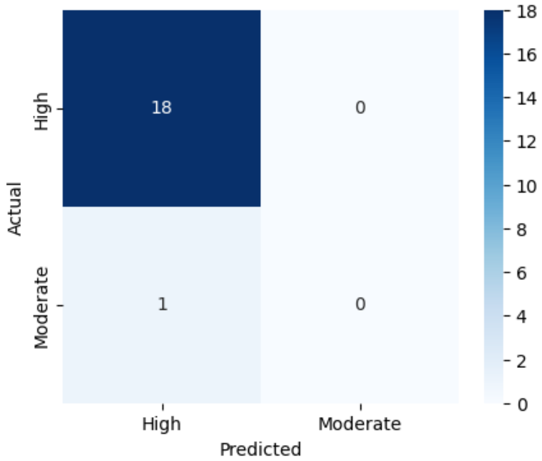

Figure 12. Confusion Matrix-XGBoost

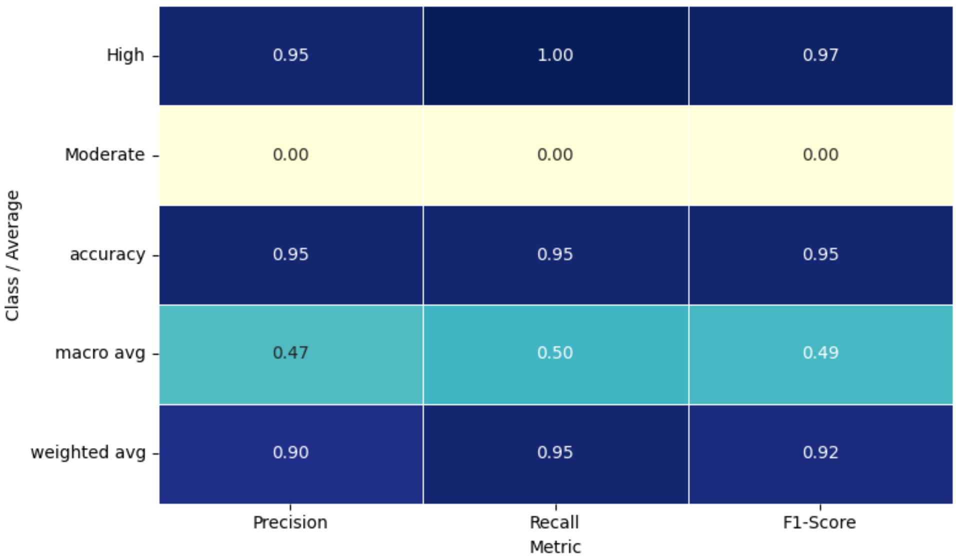

Figure 13. XGBoost Classification Metrics

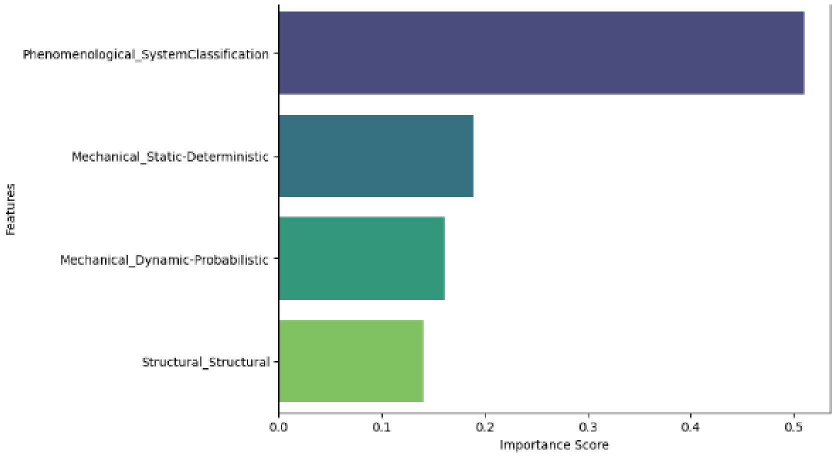

Figure 14. Future Importance Analysis

Additional inspection of Figure 14 demonstrates that the Phenomenological System Classification is the strongest feature in the prediction of reasoning categories, with such factor contributing to over 50% of the model's predictive efficacy. Mechanical Static-Deterministic and Mechanical Dynamic-Probabilistic reasoning and corresponding centrality scores (about 19% and 16% in size) followed, whereas Structural Reasoning makes the smallest contribution (about 14%). These findings reiterate that sorting phenomena is a fundamental skill from which higher-order reasoning skills grow [1]. The relatively lower weight of structural thinking as a predictor indicates that it acts more as a distillate of the other components rather than a precursor. This requires a phase-based structure of instruction structure that initially nourishes phenomenological reasoning, and then slowly merges mechanistic and structural reasoning [49]. Further, the model's challenge in predicting moderate reasoning also emphasizes the need for the use of adaptive formative assessment, and an inquiry-based teaching approach that can aid students in overcoming cognitive plateaus to develop stable, transferable chemical reasoning abilities [49], [50].

7. Limitations

Although it offers insight into students' chemical reasoning perspective, clearly there remains to be a significant amount of empirical work that needs to be done to better understand such patterns of reasoning. First, reliance on the paper scaffolding assessment, which effectively structures student responses, may not adequately capture the dynamic and process-oriented trajectory of conceptual development that might be elicited in interviews, think-aloud protocols, or live digital environments. Second, the sample in this study was only 91 students from a high school in Semarang, so that the findings cannot certainly be generalized into broader and more diverse educational settings. Third, the predictive model, XGBoost, while being very accurate in identifying students in the High Reasoning category, did not classify any students in the Moderate group – demonstrating the challenges of feature sensitivity and class imbalance. It is worth mentioning that cross-validation methods were not in use here, and in the future models robust validation and rebalancing methods should be integrated to improve the fairness and accuracy of the predictions.

Further, the rationale test is expert validated, however, it might not be completely representative of the entire range of chemical reasoning, particularly for less-scaffolded or inter-disciplinary environments. The study also did not investigate the influence of demographic variables (e.g., gender, academic track) on reasoning performance and learning trajectories. Finally, analysis of misconceptions, although high on inter-rater agreement, was based on manually coded student responses and may contain subjective biases despite the well-structured rubric. These limitations should be redressed in future research with mixed-methods approaches, larger/more diverse samples, and the use of automated or interceding coding techniques to enhance the reliability and validity of the identification of cognitive patterns.

CONCLUSION

In the present study, we outline students' patterns of chemical reasoning about ammonia and its characterization as a weak base. Results indicate that phenomenological reasoning plays an important role as an intermediate level of reasoning in students' phasing out of structural reasoning, and that good classification- and generalization-skilled students work with a higher degree of cognitive integration. Mechanical reasoning, in contrast, develops as if in silos, with less and less chance to emerge fully during the advancement of everyday thinking without teaching intervention. Moderate reasoning profiles were predominant among students, highlighting the importance of well-designed scaffolding strategies.

Dialogic teaching methods that progressively bridge empirical events, phenomenological conjectures, and symbolic representations are necessary for fostering higher-order development of chemical reasoning. Greater investigation is needed of these transition processes between levels of reasoning, to address the ongoing levels of misconceptions around the ionizing strength of weak bases, the incorporation of scaffolding platforms that allow for enriched time-lapsed assessment, and the fine-tuning of predictive models to improve the accuracy in classification, especially in the moderate profiles of reasoning categories.

ACKNOWLEDGMENT

This research was funded by the *Penelitian Terapan Kepakaran* program through the DPA LPPM UNNES Year 2025, under Grant Number 748.14.3/UN37/PPK.11/2025.

REFERENCES

[1] M. Weinrich and V. Talanquer, "Mapping Students' Modes of Reasoning When Thinking About Chemical Reactions Used to Make a Desired Product," *Chem. Educ. Res. Pract.*, vol. 17, no. 2, pp. 394–406, 2016, doi:10.1039/c5rp00208g.

[2] T. Nkosi and L. Mnguni, "The Impact of Physical Molecular Models on

Students' Visuo-Semiotic Reasoning Skills Related to the Lewis Structure and Ball-and-Stick Model of Ammonia," *J. Balt. Sci. Educ.*, vol. 19, no. 4, pp. 594–604, 2020, doi:10.33225/jbse/20.19.594.

[3] M. M. Cooper, H. Kouyoumdjian, and S. M. Underwood, "Investigating Students' Reasoning About Acid–Base Reactions," *J. Chem. Educ.*, vol. 93, no. 10, pp. 1703–1712, 2016, doi:10.1021/acs.jchemed.6b00417.

[4] G. Bhattacharyya, "From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism," *J. Chem. Educ.*, vol. 90, no. 10, pp. 1282–1289, 2013, doi:10.1021/ed300765k.

[5] V. Talanquer, "The Complexity of Reasoning About and With Chemical Representations," *JACS Au*, vol. 2, no. 12, pp. 2658–2669, 2022, doi:10.1021/jacsau.2c00498.

[6] N. J. Kim, B. R. Belland, and A. Walker, "Effectiveness of Computer-Based Scaffolding in the Context of Problem-Based Learning for STEM Education: Bayesian Meta-Analysis," *Educ. Psychol. Rev.*, vol. 30, no. 2, pp. 397–429, 2017, doi:10.1007/s10648-017-9419-1.

[7] B. R. Belland, A. Walker, N. J. Kim, and M. Lefler, "Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education," *Rev. Educ. Res.*, vol. 87, no. 2, pp. 309–344, 2016, doi:10.3102/0034654316670999.

[8] Z. D. Kırbulut and Ö. Geban, "Using Three-Tier Diagnostic Test to Assess Students' Misconceptions of States of Matter," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 10, no. 5, 2014, doi:10.12973/eurasia.2014.1128a.

[9] P. Zhang and F. Zhang, "Enhancing Education and Teaching Management Through Data Mining and Support Vector Machine Algorithm," *Int. J. e-Collaboration*, vol. 20, pp. 1–16, Oct. 2024, doi:10.4018/IJeC.357998.

[10] R. A. Triwardani, Y. Fuad, and M. Masriyah, "Students' Reasoning Behavior on Generalization of Figural Pattern," 2018, doi:10.2991/miseic-18.2018.62.

[11] E. Lintunen et al., "Cognitive abilities predict performance in everyday computer tasks," *Int. J. Hum. Comput. Stud.*, vol. 192, p. 103354, 2024, doi:10.1016/j.ijhcs.2024.103354.

[12] N. Zaini and N. Rahmat, "Exploring the Relationship between Learning Strategies Used in Language Learning," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 13, pp. 2556–2572, May 2023, doi:10.6007/IJARBSS/v13-i5/17030.

[13] Y. Yuberti, Y. Suryani, and I. Kurniawati, "Four-Tier Diagnostic Test With Certainty of Response Index to Identify Misconception in Physics," *Indones. J. Sci. Math. Educ.*, vol. 3, no. 2, pp. 245–253, 2020, doi:10.24042/ijsme.v3i2.6061.

[14] N. Malik, N. Sangwan, N. Bohra, A. Kumari, D. Sheoran, and M. Dabas, "HyperSegRec: enhanced hypergraph-based recommendation system with user segmentation and item similarity learning," *Cluster Comput.*, vol. 27, pp. 11727–11745, Jun. 2024, doi:10.1007/s10586-024-04560-x.

[15] T. Gong, Y. Jiang, L. Saldivia, and C. Agard, "Using Sankey diagrams to visualize drag and drop action sequences in technology-enhanced items," *Behav. Res. Methods*, vol. 54, Jun. 2021, doi:10.3758/s13428-021-01615-4.

[16] A. Asselman, M. Khaldi, and A. Souhaib, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interact. Learn. Environ.*, vol. 31, pp. 1–20, May 2021, doi:10.1080/10494820.2021.1928235.

[17] F. V. Ryzhkov, Y. E. Ryzhkova, and M. N. Elinson, "Python in Chemistry: Physicochemical Tools," *Processes*, vol. 11, no. 10, p. 2897, 2023, doi:10.3390/pr11102897.

[18] J. Lönnbro, L. D. Holmqvist, E. Persson, P. Thysell, N. D. Åberg, and S. M. Wallerstedt, "Inter-rater Reliability of Assessments Regarding the Quality of Drug Treatment, and Drug-related Hospital Admissions," *Br. J. Clin. Pharmacol.*, vol. 87, no. 10, pp. 3825–3834, 2021, doi:10.1111/bcp.14790.

[19] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159–174, Apr. 1977, doi:10.2307/2529310.

[20] I. N. Suastika and A. A. I. D. A. Utami, "Test the Validity of the Multicultural Education Course at the E-Learning (Moodel Learning Management System) Universitas Pendidikan Ganesha," 2023, doi:10.4108/eai.28-10-2022.2326354.

[21] T. C. Postma and J. G. White, "Developing Integrated Clinical Reasoning Competencies In Dental Students Using Scaffolded Case-Based Learning – Empirical Evidence," *Eur. J. Dent. Educ.*, vol. 20, no. 3, pp. 180–188, 2015, doi:10.1111/eje.12159.

[22] M. M. Cooper, L. M. Corley, and S. M. Underwood, "An Investigation of College Chemistry Students' Understanding of Structure–Property Relationships," *J. Res. Sci. Teach.*, vol. 50, no. 6, pp. 699–721, 2013, doi:10.1002/tea.21093.

[23] A. Kondinski, J. Bai, S. Mosbach, J. Akroyd, and M. Kraft, "Knowledge Engineering in Chemistry: From Expert Systems to Agents of Creation," *Acc. Chem. Res.*, vol. 56, no. 2, pp. 128–139, 2022, doi:10.1021/acs.accounts.2c00617.

[24] J. Pelamonia and A. D. Corebima, "Cognitive Basis and Semantic Structure of Phenomenological Reasoning on Science Among Lower Secondary School Students: A Case of Indonesia," *J. Balt. Sci. Educ.*, vol. 14, no. 4, pp. 474–486, 2015, doi:10.33225/jbse/15.14.474.

[25] I. Medhi, M. Lakshmanan, K. Toyama, and E. Cutrell, "Some Evidence for the Impact of Limited Education on Hierarchical User Interface Navigation," 2013, doi:10.1145/2470654.2481390.

[26] J. Olander, C. Stenberg, S. Stenlund, and M. Andrée, "Didactic Reasoning About Using Chemicals in Teaching Upper Secondary Chemistry," *J. Chem. Educ.*, vol. 100, no. 1, pp. 45–52, 2022, doi:10.1021/acs.jchemed.2c00511.

[27] K. Heisterkamp and V. Talanquer, "Interpreting Data: The Hybrid Mind," *J. Chem. Educ.*, vol. 92, no. 12, pp. 1988–1995, 2015, doi:10.1021/acs.jchemed.5b00589.

[28] K. S. Taber, "Revisiting the Chemistry Triplet: Drawing Upon the Nature of Chemical Knowledge and the Psychology of Learning to Inform Chemistry Education," *Chem. Educ. Res. Pract.*, vol. 14, no. 2, pp. 156–168, 2013, doi:10.1039/c3rp00012e.

[29] G. Chittleborough and D. F. Treagust, "The Modelling Ability of Non-Major

Chemistry Students and Their Understanding of the Sub-Microscopic Level," *Chem. Educ. Res. Pract.*, vol. 8, no. 3, pp. 274–292, 2007, doi:10.1039/b6rp90035f.

[30] A. J. Dood and F. M. Watts, "Students' Strategies, Struggles, and Successes With Mechanism Problem Solving in Organic Chemistry: A Scoping Review of the Research Literature," *J. Chem. Educ.*, vol. 100, no. 1, pp. 53–68, 2022, doi:10.1021/acs.jchemed.2c00572.

[31] M. B. Permatasari, S. Rahayu, and I. W. Dasna, "Chemistry Learning Using Multiple Representations: A Systematic Literature Review," *J. Sci. Learn.*, vol. 5, no. 2, pp. 334–341, 2022, doi:10.17509/jsl.v5i2.42656.

[32] T. R. Besold, A. S. d. Garcez, K. Stenning, L. v. d. Torre, and M. v. Lambalgen, "Reasoning in Non-Probabilistic Uncertainty: Logic Programming and Neural-Symbolic Computing as Examples," *Minds Mach.*, vol. 27, no. 1, pp. 37–77, 2017, doi:10.1007/s11023-017-9428-3.

[33] K. Kamaluddin, N. K. Sani, I. W. Darmadi, and N. Nurgan, "Analysis of Students Scientific Reasoning Ability and the Correlation to Students Cognitive Ability in Physics Learning," *J. Penelit. Pendidik. Ipa*, vol. 9, no. 11, pp. 10173–10179, 2023, doi:10.29303/jppipa.v9i11.5657.

[34] D. B. Haas, F. M. Watts, A. J. Dood, and G. V. Shultz, "Analysis of Organic Chemistry Students' Developing Reasoning Elicited by a Scaffolded Case Comparison Activity," *Chem. Educ. Res. Pract.*, vol. 25, no. 3, pp. 742–759, 2024, doi:10.1039/d4rp00021h.

[35] L. Keiner and N. Graulich, "Transitions Between Representational Levels: Characterization of Organic Chemistry Students' Mechanistic Features When Reasoning About Laboratory Work-Up Procedures," *Chem. Educ. Res. Pract.*, vol. 21, no. 1, pp. 469–482, 2020, doi:10.1039/c9rp00241c.

[36] N. Becker, C. Stanford, M. H. Towns, and R. S. Cole, "Translating Across Macroscopic, Submicroscopic, and Symbolic Levels: The Role of Instructor Facilitation in an Inquiry-Oriented Physical Chemistry Class," *Chem. Educ. Res. Pract.*, vol. 16, no. 4, pp. 769–785, 2015, doi:10.1039/c5rp00064e.

[37] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conference Series: Materials Science and Engineering*, vol. 336, p. 012017, 2018, doi:10.1088/1757-899X/336/1/012017.

[38] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021, doi:10.1186/s13638-021-01910-w.

[39] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set," *Journal of Statistical Software*, vol. 61, no. 6, 2014, doi:10.18637/jss.v061.i06.

[40] C.-I. Lee, Y.-F. Yang, and S.-Y. Mai, "The Impact of a Scaffolded Assessment Intervention on Students' Academic Achievement in Web-Based Peer Assessment Activities," *International Journal of Distance Education Technologies*, vol. 14, no. 4, pp. 41–54, 2016, doi:10.4018/ijdet.2016100104.

[41] L. Hamerská, T. Matěcha, M. Tóthová, and M. Rusek, "Between Symbols and Particles: Investigating the Complexity of Learning Chemical Equations," *Education Sciences*, vol. 14, no. 6, p. 570, 2024, doi:10.3390/educsci14060570.

[42] A. Wiyarsi, A. R. Fachriyah, D. Supriadi, and M. I. B. M. Damanhuri, "A Test of Analytical Thinking and Chemical Representation Ability on 'Rate of Reaction' Topic," *Jurnal Cakrawala Pendidikan*, vol. 38, no. 2, pp. 228–242, 2019, doi:10.21831/cp.v38i2.23062.

[43] P. D. Natalia and A. Sudrajat, "Development of Three-Tier Diagnostic Test Instrument to Measure Misconceptions of Class XI Students on Reaction Rate Materials," *Jurnal Teknologi Pendidikan*, vol. 8, no. 1, p. 1, 2023, doi:10.33394/jtp.v8i1.6192.

[44] F. Mardika and A. Mahmudi, "An Analysis of Proportional Reasoning Ability of Junior High School Students," *Jurnal Riset Pendidikan Matematika*, vol. 8, no. 1, pp. 22–32, 2021, doi:10.21831/jrpm.v8i1.14995.

[45] G. D. Y. Kumlu and N. Yürük, "Investigation of Cognitive and Metacognitive Strategies Used by Preservice Science Teachers Exposed to Explicit and Peer Tutoring Reading Strategies Instruction," *International Journal of Contemporary Educational Research*, vol. 10, no. 1, pp. 1–24, 2023, doi:10.33200/ijcer.1082537.

[46] H. Herunata, I. M. Octavia, H. W. Wijaya, and P. Parlan, "Correlational Analysis of Conceptual Understanding, Chemical Representation, and Representational Competence on Acid-Base," 2023, doi:10.1063/5.0112177.

[47] S. B. S. Lai, N. H. N. B. M. Shahri, M. Mohamad, H. A. B. A. Rahman, and A. Rambli, "Comparing the Performance of AdaBoost, XGBoost, and Logistic Regression for Imbalanced Data," *Mathematics and Statistics*, vol. 9, no. 3, pp. 379–385, 2021, doi:10.13189/ms.2021.090320.

[48] P. Zhang, Y. Jia, and Y. Shang, "Research and Application of XGBoost in Imbalanced Data," *International Journal of Distributed Sensor Networks*, vol. 18, no. 6, p. 155013292211069, 2022, doi:10.1177/15501329221106935.

[49] A. Jufriadi, H. D. Ayu, and H. Y. Pratiw, "Developing E-Scaffolding Integrated With E-Assessment to Improve Students' Mastery of Concept," 2019, doi:10.2991/icesre-18.2019.37.

[50] M. Dobber, R. Zwart, M. Tanis, and B. van Oers, "Literature Review: The Role of the Teacher in Inquiry-Based Education," *Educational Research Review*, vol. 22, pp. 194–214, 2017, doi:10.1016/j.edurev.2017.09.002.