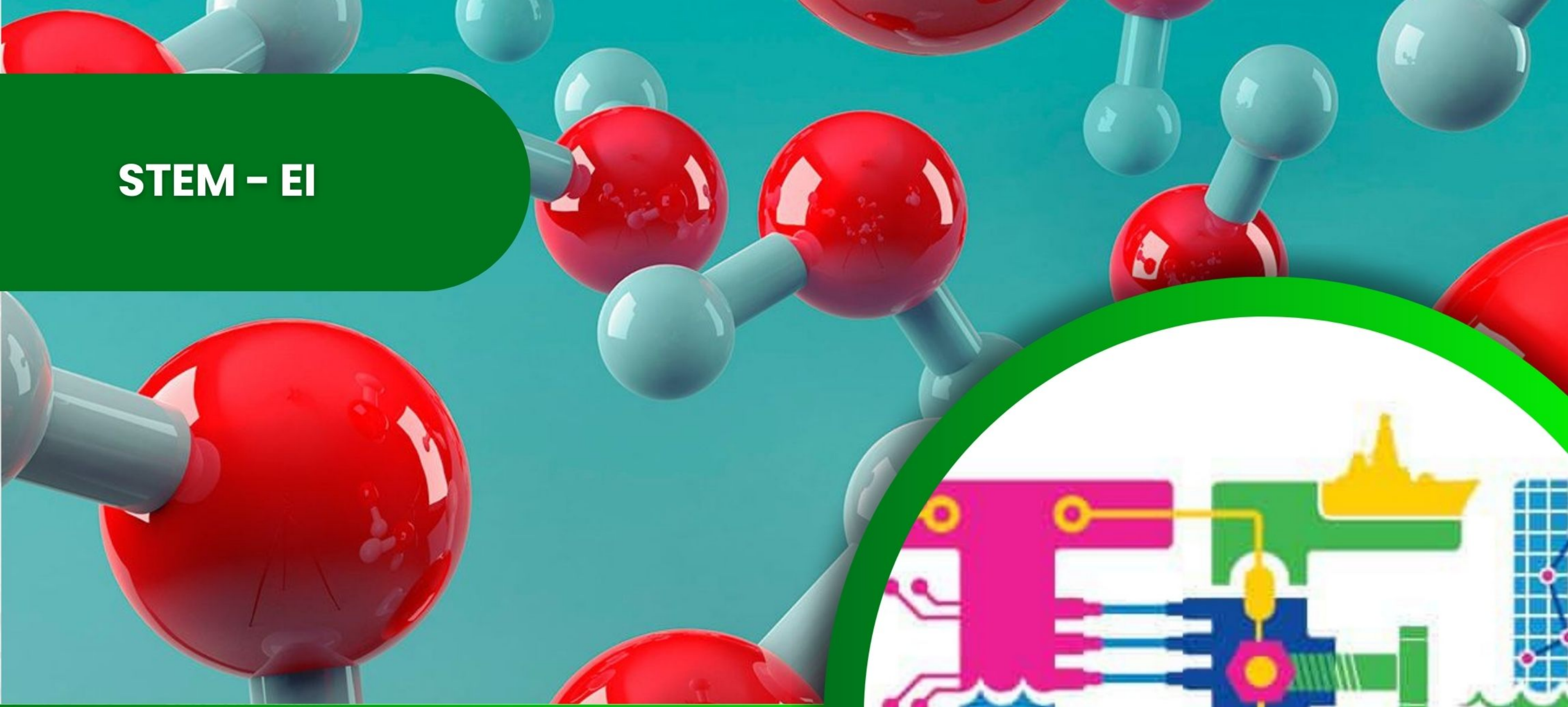

STEM - EI

STEM EDUCATION INTERNATIONAL

Vol. 1 No. 2, Dec 2025

Dream it, design it, do it—
STEM makes it possible

Published by
PT Ahan Group Publisher

Preface

It is with great enthusiasm that we present **Volume 1, Number 2** of *STEM Education International*. Following the successful debut of our first issue, this second instalment continues the journal's mission to disseminate high-quality, peer-reviewed research that advances understanding and practice in **STEM (Science, Technology, Engineering, and Mathematics) education** across diverse learning contexts.

The articles in this issue focus on two prominent themes within contemporary STEM education: **instructional strategies for fostering higher-order thinking** and **the development and validation of diagnostic assessment instruments.** These contributions collectively underscore the importance of both **designing meaningful learning experiences** and **accurately measuring student understanding** to inform effective teaching and curriculum development.

The opening study, **"Implementation of Problem-Based STAD Learning to Improve Students' Critical Thinking Skills on Ecosystems,"** investigates the integration of *problem-based learning* with *Student Teams-Achievement Divisions (STAD)* to enhance students' ability to analyse and reason about complex environmental systems. This work highlights the potential of cooperative, inquiry-oriented pedagogies in promoting critical thinking—an essential 21st-century competency.

Complementing this, **"Comparison of Discovery Learning and Problem-Based Learning Models on Concept Understanding (C1-C4) of Junior High School Students"** examines how two distinct learner-centred approaches influence students' conceptual mastery across Bloom's cognitive levels. Through comparative analysis, this study provides insights into instructional design choices that can deepen students' conceptual frameworks in science education.

The next set of articles emphasises **assessment innovation**, presenting rigorous efforts to construct and validate four-tier diagnostic instruments. **"Development of a Four-Tier Diagnostic Instrument to Identify High School Students' Understanding of Salt Hydrolysis"** and **"Development of Four-Tier Diagnostic Instruments to Identify Students' Understanding of Electrolyte and Non-Electrolyte Solutions"** both contribute practical tools that enable educators to identify student conceptions and misconceptions with precision. Such diagnostic instruments are critical for designing targeted interventions and supporting meaningful learning progressions in chemistry and related STEM domains.

Finally, **"A Four-Tier Diagnostic Instrument in Acid-Base Properties of Salt Solution: Development Procedure"** provides a detailed account of the systematic process of instrument construction. By documenting development procedures and validation practices, this article offers a valuable model for future researchers seeking to create robust assessment tools that accurately capture nuanced student understanding.

Taken together, the articles in this issue represent a rich blend of **pedagogical innovation** and **assessment research** that contribute to the ongoing transformation of STEM education. By advancing theory, methodology, and practice, these studies exemplify the journal's commitment to fostering educational excellence and supporting learners' preparedness for the challenges of a rapidly evolving world.

We extend our sincere gratitude to the authors, reviewers, and editorial team whose dedication has made this issue possible. We look forward to continued engagement from the STEM education community as we work together to improve teaching, learning, and assessment in STEM across all levels of education.

Editorial Board

List of Content

Preface i
List of Content iii

Implementation of Problem-Based STAD Learning to Improve Students' Critical Thinking Skills on Ecosystems 48
Alfina Nuril Lailia

Comparison of Discovery Learning and Problem-Based Learning Models on Concept Understanding (C1-C4) of Junior High School Students 62
Wahyu Sofi Martalinda

Development of a Four-Tier Diagnostic Instrument to Identify High School Students' Understanding of Salt Hydrolysis 74
Anne Nailul Aziz, Darsono Sigit

Development of Four-Tier Diagnostic Instruments to Identify Students' Understanding of Electrolyte and Non-Electrolyte Solutions 80
Septiana Ayuningrum Nofinadya, Rima Nuraini, Isnanik Juni Fitriyah

A Four-Tier Diagnostic Instrument in Acid-Base Properties of Salt Solution: Development Procedure 87
Rima Nuraini

STEM Education International,
Vol. 1, No. 2, December 2025, Page 48-61

p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx

STEM-EI
STEM Education International

Implementation of Problem-Based STAD Learning to Improve Students' Critical Thinking Skills on Ecosystems

Alfina Nuril Lailia

Science Education Department, State University of Malang, Malang, East Java, Indonesia

**Corresponding author: alfina.nuril.2103516@students.um.ac.id*

Abstract: This study investigates whether the Problem-Based Student Teams Achievement Division (STAD) learning model can enhance students' critical thinking skills to a greater extent than conventional lecture-based instruction on the topic of Human Impact on Ecosystems. A quasi-experimental method was employed, using test items as the research instrument. A total of 31 seventh-grade students from class VII I served as the control group, while 30 students from class VII G constituted the experimental group. Descriptive analysis was used to describe the post-test results for each indicator of critical thinking skills. An independent t-test was then conducted to examine the differences in essential thinking abilities between the experimental and control groups. Additionally, an N-gain test was conducted to assess the magnitude of improvement in students' critical thinking skills in the experimental group relative to the control group. The findings indicate that the Problem-Based STAD learning model improved students' critical thinking skills in the experimental group at a moderate level, with an N-gain score of 0.43. In contrast, the conventional learning approach led to a low level of improvement, with an N-gain score of 0.24. This study emphasises the implementation of a problem-based STAD model derived from students' real-life issues, making the learning process more contextual. The topic of Human Impact on Ecosystems was chosen for its strong relevance to students' everyday experiences.

Keywords: STAD Learning Model, Problem-Based Learning, Critical Thinking, 21st Century Skills

INTRODUCTION

Science and technology, in their development, require a strategic role from the educational sector. Moreover, education emphasises different areas of focus (Musahrain et al., 2024). In the 21st century, students are faced with various educational challenges, including: (1) collaborating with others, (2) presenting information clearly, (3) thinking critically and solving problems, and (4) creativity and innovation (Nuraina & Nestiadi, 2025). Therefore, critical thinking skills have become essential competencies to develop. To solve personal and social problems effectively, students need to acquire a range of 21st-century skills (Yulianti et al., 2022).

Critical thinking skills are essential in students' lives. Students need critical thinking as a tool to solve real-world problems (Suradika et al., 2023). These skills are crucial for fostering deep learning, as they enable students to utilise evidence, evaluate presented materials, and reflect on ideas (Chen et al., 2024). Moreover, critical thinking can help students minimise errors when making decisions (Ariadila et al., 2023). Therefore, critical thinking enables students to assess and evaluate events, leading to more thoughtful and informed decision-making.

According to Ennis, critical thinking is the process of logically and deliberately processing information to make informed decisions. It is a skill that involves utilising various objects and concepts to solve problems (Muhibbuddin et al., 2023). Norris and Ennis (1989) identified five components of critical thinking skills, which include: (1) providing a simple explanation, (2) building basic skills, (3) drawing conclusions, (4) offering further clarification, and (5) managing strategies and tactics (Chusni et al., 2020).

Lestari and Annizar (2020) stated that critical thinking skills can be assessed using PISA items, as they require reasoning and strategic problem-solving abilities—skills that define critical thinking. According to the 2022 PISA results, Indonesia ranked 68th out of 81 participating countries (OECD, 2023). The low level of critical thinking skills among Indonesian junior high school students was also highlighted by Nuraina and Nestiadi (2025), who reported that students cannot identify problems and propose ideas or solutions in response to case studies presented during learning activities. These findings indicate that Indonesian students' critical thinking skills remain relatively low compared to those of students in other countries.

The low level of students' critical thinking skills can be attributed to several factors, including the limited use of effective teaching methods and learning models in the classroom. In practice, classroom instruction is still predominantly delivered through teacher-centred lectures, which limits students' active engagement in the learning process. On the other hand, students' understanding is highly influenced by the instructional materials and learning models employed (Nuraina & Nestiadi, 2025). Therefore, to address this issue, alternative solutions are needed—one of which is implementing innovative learning models that foster critical thinking skills. Various efforts have been made to enhance students' critical thinking abilities, including the use of electronic student worksheets (E-LKPD) (Rofik et al., 2025), STEAM-based Project-Based Learning (PjBL)(Rofik et al., 2025, Sari et al., 2025), and E-modules (Amalia et al., 2024).

Roger and Johnson (1994) stated that cooperative learning can promote higher-order thinking skills in students by encouraging the exchange of ideas and collaborative problem-solving with peers through group discussions (Qismullah Yusuf et al., 2015). The Student Teams Achievement Divisions (STAD) model is a cooperative learning approach that provides students with opportunities to engage in the learning process actively. STAD organises students into small groups of four to five members with heterogeneous composition in terms of academic ability, gender, and ethnicity (Yusuf et al., 2015). According to Sinaga et al (2022), STAD is considered an effective model for fostering creativity, coordination skills, critical thinking, and the ability to support others.

In addition, students can develop their critical thinking skills as they are increasingly exposed to complex problems. Such problems train students to solve them by making maximum efforts to analyse and express opinions, categorise information, provide justification, reflect, interpret the meaning of judgments, and draw conclusions from problem-solving processes (Afifah & Nurfalah, 2019). To effectively solve a problem, individuals need critical thinking skills, as these skills encourage them not to settle for a single answer but to explore alternative solutions based on analysis and information derived from the problem itself (Sholihah & Lastariwati, 2020). Therefore, problem-based learning models can be considered a viable approach to fostering students' analytical problem-solving abilities.

The topic Human Impact on Ecosystems was selected because it reflects real-world problems encountered in students' daily lives. For example, activities such as dumping waste into rivers or clearing land can lead to declines in biodiversity within ecosystems. This content is closely related to critical thinking skills, as it allows students to observe environmental issues in their surroundings directly and stimulates their curiosity (Nuraina & Nestiadi, 2025). By studying the material contextually, students are expected to develop their critical thinking skills by investigating the causes of environmental problems and exploring potential solutions.

The STAD learning model can be integrated with a problem-based learning approach, resulting in the Problem-Based STAD model. This instructional model emphasises student collaboration within groups to understand and solve real-world problems, thereby achieving learning objectives and constructing understanding based on students' cognitive processes (Rianti et al., 2021). The integration of Problem-Based Learning (PBL) and STAD provides students with opportunities to examine, respond to, and collaborate critically. Students are also trained to confront problems, think critically to find solutions, and are challenged to resolve both academic and real-life issues (Andriyati & Noviani, 2023). The syntax of the Problem-Based STAD model, as implemented by Dwi Anjani et al. (2023), illustrates the integration of PBL stages into cooperative STAD learning, as follows: (1) the phase of presenting objectives, motivation, and information aligns with orienting students to the problem; (2) the group formation phase aligns with organizing students for learning; (3) the discussion phase corresponds with guiding individual or group investigations; (4) the phase involving presentations, quizzes, and rewards aligns with presenting students' work; and (5) the evaluation and conclusion phase corresponds with analyzing and evaluating the problem-solving process.

In the Problem-Based STAD learning model, during the phase of presenting objectives, motivation, and information, students listen to the teacher's presentation on environmental issues in their surroundings. In the second phase, group formation, students are organised into groups consisting of 4–5 members. The third phase, discussion, involves students engaging in dialogue about the causes and impacts of environmental problems and exploring possible solutions and preventive measures. During the presentation phase, students present the results of their group work and receive feedback from other groups, which further promotes peer discussion. This phase is followed by a quiz and the distribution of rewards, aimed at developing students' individual competencies. In the final phase, evaluation and conclusion, students summarise what they have learned throughout the lesson. Based on these student activities within the Problem-Based STAD learning syntax, students are trained to provide explanations and arguments to their peers, engage in deeper thinking during discussions, and develop strategic approaches to address environmental problems. Therefore, this learning model is expected to foster and enhance students' critical thinking skills effectively.

A study conducted by Sulistyani and Pratama (2024) demonstrated that the integration of Problem-Based Learning and STAD effectively enhances students' critical and mathematical thinking skills. Similar findings were reported by Karma et al. (2023) (Karma et al., 2023), who found a positive impact on students' critical thinking abilities following the implementation of Problem-Based Learning within an STAD framework. Based on these findings, it is expected that students' critical thinking skills can also be improved through the application of the Problem-Based STAD model compared to conventional lecture-based instruction, particularly on the topic of Human Impact on Ecosystems.

Based on the aforementioned explanation, problem-based learning can cultivate students' critical-thinking skills, while the STAD learning model can facilitate students' ability to work collaboratively in teams. These models are appropriate for efforts to improve students' critical-thinking skills in Indonesia, which remain relatively low. This study examines the statistical differences in students' critical-thinking abilities before and after learning through the Problem-Based STAD model. The purpose of this study is to determine the improvement in students' critical-thinking skills regarding human impacts on ecosystems following the implementation of the Problem-Based STAD learning model.

METHOD

This study employed a quasi-experimental approach using a pre-test post-test non-equivalent control group design (Cohen, 2007). It involved two class groups, designated as the experimental and control groups. The experimental group received instruction through a Problem-Based STAD learning model, hereafter referred to as the PB-STAD class (Problem-Based Student Teams Achievement Divisions). In contrast, the control group did not receive this instructional model. Instead, the control group was taught using a lecture-based method and is hereafter referred to as the lecture class.

One of the junior high schools in Nganjuk Regency served as the research site for this study, involving 30 seventh-grade students from class VII G—16, male and 14 female students, comprising the PB-STAD group. Additionally, 31 students from class VII I—16 male and 15 female—participated as the lecture group. In this study, the implementation of the Problem-Based STAD model functioned as the independent variable. At the same time, students' critical thinking skills on the topic of Human Impact on Ecosystems served as the dependent variable.

The PB-STAD class participated in lessons and completed student worksheets (LKPD) aligned with the Problem-Based STAD model's syntax. This alignment enabled observation of students' adherence to the PB-STAD learning steps through their responses on the prepared worksheets. The cultivation of students' critical thinking skills was not limited to the PB-STAD class; it was also implemented in the lecture class. In the lecture group, students' critical thinking skills were fostered through teacher-led prompts and completion of student worksheets.

This study employed a critical thinking assessment instrument comprising five open-ended questions, administered as both pre- and post-tests. The pre-test was administered before the instructional intervention to assess students' initial essential thinking skills. The post-test was administered after the instructional intervention had been delivered. Both the experimental and control groups completed these assessments to identify differences in students' critical thinking performance following the implementation of different instructional models. Before its implementation, the student worksheet (LKPD) instrument underwent validation, including validity and reliability testing using the percentage-of-agreement method.

Table 1. Percentage of Similarity in LKPD Validation Using the Percentage of Agreement Method

Aspect		Average	Criteria
Learning and Materials	**Readability and Visual Design**		
100%	94%	97%	Almost Perfect

In addition, the test instrument also underwent the same validation procedures, with the results presented in Table 2. The validation results showed that the LKPD achieved a percentage of agreement of 97%, while the test items reached 98.4%. These results indicate that both the test instrument and the LKPD are valid and reliable for use with students, as supported by Borich (1994) in Widiawati et al. (2022), who stated that an instrument is considered valid when it achieves a percentage of agreement (PA) of 75% or higher (≥75%).

Table 2. Percentage of Similarity in Pre-test and Post-test Validation Using the Percentage of Agreement Method

Aspect			Average	Criteria
Item Feasibility	**Language Appropriateness**	**Item Appropriateness**		
99,3%	98,6%	97,3%	98,4%	Almost Perfect

The data analysis in this study employed a descriptive approach by outlining students' post-test results based on the critical thinking skill indicators proposed by Norris and Ennis (1989), which include: (1) providing simple explanations, (2) building basic skills, (3) drawing conclusions, (4) offering further clarification, and (5) managing strategies and tactics. Before analysis, the data were subjected to normality testing (Shapiro-Wilk) and homogeneity testing (Levene's Test). An independent t-test was used to compare students' critical thinking skills between the PB-STAD class and the lecture class. Additionally, an N-Gain analysis was conducted to determine the magnitude of improvement in students' critical thinking skills in both groups. The level of improvement is classified as high if the gain score exceeds 0.7 ($g > 0.7$), moderate if it falls between 0.3 and 0.7 ($0.3 < g \leq 0.7$), and low if it is less than or equal to 0.3 ($g \leq 0.3$), as suggested by (Meltzer, 2002). Furthermore, students' critical thinking skills were categorised into levels as defined by Ramdani et al. (2020), using the criteria in Table 3.

Table 3. Criteria for Critical Thinking Skills

Gain Score	Category
81,25 < x ≤ 100	Very High
71,50 < x ≤ 81,25	High
62,50 < x ≤ 71,50	Moderate
43,75 < x ≤ 62,5	Low
0 < x ≤ 43,75	Very Low

RESULTS AND DISCUSSION

The collected pretest and posttest data were initially subjected to prerequisite analyses, namely the Shapiro–Wilk normality test and Levene's homogeneity test. The results showed that the data were normally distributed and homogeneous, as indicated by significance values greater than 0.05, thereby allowing the analysis to proceed to the subsequent tests. The results of the normality and homogeneity tests are presented in the table below:

Table 4. Normality Test Results

Class		Shapiro-Wilk Statistic	p-value	Note
PB-STAD	Pretest	0,969	0,512	Normal
	Posttest	0,936	0,069	Normal
Lecture-based	Pretest	0,956	0,233	Normal
	Posttest	0,979	0,782	Normal

The analysis of students' pre-test responses indicated that the initial critical thinking abilities of students in both the PB-STAD class and the lecture class were relatively equivalent, with average pre-test scores of 37.8 and 39.1, respectively. The equivalence of initial critical thinking skills between the two groups was confirmed through an Independent T-test, which showed no significant difference. This result supports the comparability of the two classes. The pre-test data for both the PB-STAD and the lecture classes are presented in Table 6.

Table 5. Homogeneity Test Results

Test	Levene's Statistic	p-value	Note
Pretest	0,629	0,413	Homogeneous
Posttest	0,145	0,705	Homogeneous

The PB-STAD class engaged in learning activities designed to develop critical thinking skills through the Problem-Based STAD model, guided by structured student worksheets (LKPD). Students began by identifying ecosystem-related problems presented in the LKPD, with guidance from the teacher. They then formed pre-assigned groups to discuss the issues and complete the additional questions in the worksheet. Each group presented the results of their discussion, followed by individual quizzes to evaluate students' understanding. In contrast, the lecture class focused on teacher-centred instruction. The teacher explained the topic of human impact on ecosystems while intermittently posing questions to stimulate students' critical thinking. This was followed by students working on pre-prepared worksheets designed to reinforce the material.

Table 6. Average Pre-test Scores of the PB-STAD Class and the Lecture Class

Class	Mean Pretest Score	Category
PB-STAD	37,8	Very Low
Lecture-based	39,1	Very Low

Following the implementation of the Problem-Based STAD learning model, post-test scores increased in the PB-STAD class. The improvement in scores, based on the comparison of pre-test and post-test results, is presented in the table below:

Table 7. Comparison of Students' Critical Thinking Skills in the PB-STAD Class and the Lecture Class

Class	Pretest Results		Posttest Results		Score Improvement
	Score	Category	Score	Category	
PB-STAD	37,8	Very low	64,2	Moderate	26,4
Lecture-based	39,1	Very low	54,0	Low	14,9

Following the implementation of the Problem-Based STAD learning model, differences in students' critical thinking levels between the PB-STAD class and the lecture class were observed. This can be seen in the percentage distribution of students' essential levels of thinking based on their post-test scores, which shows that a higher proportion of students in the PB-STAD class fell into the "very high" and "high" categories than those in the lecture class. The detailed data are presented in Table 8.

Table 8. Distribution of Students' Critical Thinking Skills in the PB-STAD Class and the Lecture Class.

Criteria for Critical Thinking Skills	PB-STAD Class	Lecture Class
Very high	10%	3,2%
High	26,7%	6,5%
Moderate	23,3%	25,8%
Low	26,7%	41,9%
Very low	13,3%	22,6%

For the first critical thinking indicator, providing a simple explanation, 67% of students in the PB-STAD class achieved high scores, compared to only 16% in the lecture class. This finding indicates that the PB-STAD class had more students demonstrating high-level critical thinking skills on this indicator than the lecture class. An example of a high-scoring student response from the PB-STAD class on a question related to this indicator—regarding the impact of burning waste—is as follows "(1) Air pollution occurs because the smoke from burning waste releases ash into the air, leading to pollution; (2) When plastic waste is burned, the chemical substances from the plastic can seep into the soil, causing soil contamination; (3) It can cause respiratory and skin diseases; the smoke, which contains ash, may lead to breathing difficulties when inhaled and can cause skin irritation when it comes into contact with the skin".

The second critical thinking indicator, building basic skills, included a question about activities that can lead to river pollution. The data show that 30% of students in the PB-STAD class achieved high scores, compared with 26% in the lecture class. This finding suggests that a greater proportion of students in the PB-STAD class demonstrated high-level critical thinking skills for this indicator than those in the lecture class. An example of a high-scoring student response from the PB-STAD class regarding activities that may cause river pollution is as follows "(1) Disposing of waste and sewage: carelessly throwing waste into rivers can lead to accumulation of garbage, potentially causing floods due to blocked water flow; (2) Dumping chemicals into rivers: this can harm aquatic life, as the chemicals may be toxic to organisms living in the river; (3) Fishing using explosives: the use of bombs can kill aquatic animals due to the force of the explosion".

The third critical thinking indicator, concluding, showed that more students in the PB-STAD class demonstrated high-level critical thinking skills than those in the lecture class. Specifically, 47% of students in the PB-STAD class achieved high scores on this indicator, compared with only 22% in the control group. The question for this indicator asked students to conclude an article discussing the causes and impacts of marine waste issues. An example of a high-scoring response from the experimental class is as follows: "The causes include the continued use of plastic by many industries in their products, as well as the public's reliance on single-use plastics in daily life. The impacts include soil, water, and air pollution. For example, marine animals may ingest plastic waste, mistaking it for food, leading to their death. Additionally, harmful chemicals from plastics can seep into the soil and contaminate sources of clean water".

For the fourth critical thinking indicator, providing further clarification, the data similarly showed that more students in the PB-STAD class demonstrated higher levels of critical thinking than those in the lecture class. Specifically, 43% of students in the PB-STAD class scored in the high category for this indicator, while 38% of students in the lecture class did so. The question for this indicator involved analysing the potential consequences for deer populations if mining activities in the Kalimantan forests continue to expand. One example of a high-scoring response from a student in the PB-STAD class is as follows: "The deer population could face extinction by 2024 due to the continued expansion of mining areas and the resulting reduction of forest habitats. As their habitats become increasingly restricted, both deer and tigers lose their natural habitats and are forced to migrate to other areas. In these smaller areas, the likelihood of encounters between deer and tigers increases significantly. Since tigers prey on deer, this could lead to a sharp decline in the deer population. Moreover, with limited habitat, deer may struggle to find food, potentially leading to starvation and, ultimately, extinction".

Meanwhile, the percentage of students who achieved high scores on the fifth critical thinking indicator, managing strategies and tactics, was relatively similar between the PB-STAD and lecture classes. A total of 83% of students in the PB-STAD class and 84% in the lecture class attained high scores on this indicator. This finding reflects that students in both groups demonstrated comparable proficiency in handling strategy and tactics. However, students in the PB-STAD class received more structured training in critical thinking related to this indicator. In PB-STAD learning, students were presented with two alternative solutions to address a problem and were asked to select or design the most appropriate strategy to resolve the issue. This approach encouraged students to make thoughtful decisions by considering various perspectives. Moreover, classroom observations revealed that the PB-STAD class was more actively engaged, as students were required to express their opinions on the most suitable strategies and tactics for addressing environmental problems. This suggests that the PB-STAD model is more effective in enhancing student engagement.

Table 9. N-Gain of Pretest and Posttest

Class	N-Gain	Note
PB-STAD	0,43	Moderate
Lecture-based	0,24	Low

Students in the PB-STAD class generally demonstrated higher levels of critical thinking achievement compared to those in the lecture class. This suggests that implementing the Problem-Based STAD learning model contributes more to enhancing students' critical thinking skills than the conventional lecture-based method. This indication is supported by the post-test data, which show an average score of 64.0 in the PB-STAD class and 54.3 in the lecture class. The results of the Independent T-test further support this finding, revealing a Sig. (2-tailed) value of 0.015. Since the Sig. (2-tailed) The value 0.015 is less than 0.050; therefore, there is a statistically significant difference between the PB-STAD and lecture classes. An N-Gain test was subsequently conducted to determine the extent of the impact of the Problem-Based STAD model. The results of the N-Gain analysis for both the PB-STAD and lecture classes are presented in Table 9.

Table 10. Comparison of N-Gain Scores for Each Indicator

Indicator	N-Gain of PB-STAD Class	Category	N-Gain of Lecture Class	Category
Provide a simple explanation	0,44	Moderate	0,15	Low
Developing Basic Skills	0,23	Low	0,20	Low
Drawing a Conclusion	0,31	Moderate	0,02	Low
Provide further clarification	0,49	Moderate	0,34	Moderate
Managing Strategies and Tactics	0,74	High	0,63	Moderate

Based on the table above, the overall N-Gain score for the PB-STAD class was 0.43, indicating a moderate level of improvement. In contrast, the lecture class had an N-

Gain score of 0.24, indicating a low level of improvement. It was also observed that the N-Gain for each indicator in the PB-STAD class was consistently higher than that of the lecture class. The N-Gain score of 0.43 in the PB-STAD class indicates that the Problem-Based STAD learning model effectively improved students' critical thinking skills at a moderate level, in contrast to the conventional lecture-based approach, which achieved a lower N-Gain score of 0.24, indicating a low level of improvement. Furthermore, the research data revealed that students' critical thinking skills in the PB-STAD class improved to a moderate category across four essential indicators of thinking: providing simple explanations, concluding, offering further clarification, and managing strategies and tactics. The percentage data on students' critical thinking levels in both the PB-STAD and lecture classes also showed that more students in the PB-STAD class fell into the "very high" and "high" categories than in the lecture class. These findings support the notion that students' critical thinking skills can be developed through the implementation of the Problem-Based STAD learning model. This indication is consistent with the findings of Karma et al. (2023), who demonstrated that integrating the cooperative STAD learning model with a problem-based learning approach can enhance students' critical thinking skills. Similarly, the study by Sulistyani and Pratama (2024) reported an increase in students' critical thinking abilities following the application of problem-based STAD instruction. These outcomes can be attributed to the numerous benefits offered by both the STAD cooperative learning approach and problem-based learning.

The STAD learning model, which emphasises collaborative group work, provides ample opportunities for students to interact, share ideas, and exchange feedback. Through this type of learning, students can evaluate their ideas, make informed decisions, and draw thoughtful conclusions. Moreover, it also allows students to communicate their learning outcomes within the classroom (Ermin & Marsaoly, 2021). The ability to organise and analyse problems effectively, as well as to think mathematically, can further contribute to the development of critical thinking skills (Anderson & Krathwohl, 2001).

Problem-based learning (PBL) also offers a range of advantages. This approach enables students to connect their existing knowledge with real-world situations, making the learning process more contextual and meaningful (Lusmianingtyas & Sriyanto, 2022). In PBL, students take a central role in the learning process, which helps them develop critical thinking and problem-solving skills (Hidayatussakinah et al., 2021). Additionally, this model equips students with the ability to solve problems, present logical arguments, identify issues from multiple perspectives, and propose appropriate solutions to the problems they encounter (Adilah & Rosyida, 2024).

Although students' critical thinking skills have improved, as shown in the table above, the overall essential levels of thinking in both the PB-STAD and lecture classes are not yet considered high. Students in the PB-STAD class still demonstrated moderate levels of critical thinking, while students in the lecture class remained in the low category (Ramdani et al., 2020). This may be attributed to students' lack of motivation to engage in critical thinking. As noted by Chusni et al. (2020) Students are often not motivated to learn or practice critical thinking because they perceive the learning process as ineffective. Moreover, the development of critical thinking skills requires consistent and sufficient practice over time (Ermin & Marsaoly, 2021). The limited duration of this study may have hindered the full development of students' critical thinking abilities, resulting in outcomes that still fall within the moderate category. In addition, students' critical thinking abilities varied significantly, as reflected in the diverse range of post-test scores. This variation may be due to differences in intellectual development levels among students, which can be influenced by age. As Purwanto (1998) explains, individuals' capacity for mature thinking tends to increase as they grow older.

CONCLUSIONS

The results of this study indicate that implementing the Problem-Based STAD learning model on the topic of human impact on ecosystems contributed moderately to improvements in students' critical thinking skills. This is evidenced by post-test results following the application of different instructional models in the PB-STAD and lecture classes, as well as by N-Gain scores, which showed a 0.43 score in the PB-STAD class, indicating a moderate increase, and a 0.24 score in the lecture class, indicating a low increase. The Problem-Based STAD learning model may be considered a viable alternative instructional strategy for fostering students' critical thinking skills. This study was limited to two classes within a single school, and therefore the data may not fully represent the general condition of seventh-grade students' critical thinking abilities. Future studies are expected to implement Problem-Based STAD learning on a broader scale, involving multiple classes and schools, so that the resulting data can more accurately reflect students' critical thinking abilities after implementation.

ACKNOWLEDGEMENT

We extend our deepest gratitude to all parties involved in this study, including the school that granted permission for the research and the students who participated.

REFERENCES

Adilah, G. P., & Rosyida, F. (2024). Peningkatan Kemampuan Berpikir Kritis Geografi: Model Pembelajaran Berbasis Masalah Berbantuan Microlearning di MAN 1 Malang. *Al Qalam: Jurnal Ilmiah Keagamaan Dan Kemasyarakatan*, *18*(1), 466. https://doi.org/10.35931/aq.v18i1.2759

Afifah, Y., & Nurfalah, E. (2019). Kemampuan Berpikir Kritis Siswa SMP Negeri 1 Jenu Berdasarkan Langkah Facion Pada Pokok Bahasan Jajaran Genjang dan Trapesium. *Jurnal Riset Pembelajaran Matematika*, *1*(1). http://journal.unirow.ac.id/index.php/jrpm

Amalia, B., Mulyati, Y., Nugraheni, D., Habiddin, H., Annisa, N., Hasan, S., & Wicaksono, A. (2024). The Potential of Self-Organised Learning Environment E-Module to Enhance Students' Critical Thinking Skills and Cognitive Learning Outcomes. *Indonesian Journal of Biology Education*, *7*(2), 12–20.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.

Andriyati, R., & Noviani, L. (2023). Kolaborasi Model PBL dan Model STAD Dalam Pembelajaran Ekonomi. *Jurnal Pendidikan Ekonomi (JUPE)*, *11*(2). https://doi.org/10.26740/jupe.v11n2.p9

Ariadila, S. V., Silalahi, Y. F. N., Fadiyah, F. H., Jamaludin, U., & Setiawan, S. (2023). Analisis Pentingnya Keterampilan Berpikir Kritis Terhadap Pembelajaran Bagi Siswa. *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 664–669. https://doi.org/10.5281/zenodo.8436970

Chen, X., Zhao, H., Jin, H., & Li, Y. (2024). Exploring college students' depth and processing patterns of critical thinking skills and their perception in argument map(AM)-supported online group debate activities. *Thinking Skills and Creativity*, *51*. https://doi.org/10.1016/j.tsc.2024.101467

Chusni, M. M., Saputro, S., Suranto, & Rahardjo, S. B. (2020). Review of critical thinking skill in Indonesia: Preparation of the 21st century learner. In *Journal of Critical Reviews* (Vol. 7, Issue 9, pp. 1230–1235). Innovare Academics Sciences Pvt. Ltd. https://doi.org/10.31838/jcr.07.09.223

Cohen, Louis. M. lawrence. M. K. (2007). *Research Methods in Education* (6th ed.). Routledge.

Dwi Anjani, A., Senjayawati, E., & Suciati, O. (2023). Penerapan Model Problem-Based Learning Dengan Pembelajaran Kooperatif Tipe STAD Untuk Meningkatkan Motivasi Belajar Siswa Kelas VII. *Jurnal Pembelajaran Matematika Inovatif (JPMI)*, *6*(2), 879–888. https://doi.org/10.22460/jpmi.v6i2.17523

Ermin, & Marsaoly, N. (2021). Pengaruh Model Pembelajaran STAD terhadap Kemampuan Berpikir Kritis Siswa Multietnis di SMP Negeri Kota Ternate. *Jurnal Ilmiah Wahana Pendidikan*, *7*(8). https://doi.org/10.5281/zenodo.5781366

Hidayatussakinah, Marzuki, I., & Ulfa, N. A. (2021). Peningkatan Kemampuan Berpikir Kritis Melalui Model Pembelajaran Berbasis Masalah. *Biolearning Journal*, *8*(1), 2406–8241. http://www.criticalthinking.org/aboutCT

Karma, I. W., Suma, K., & Astawan, I. G. (2023). Pengaruh Model Pembelajaran Berbasis Masalah Berseting Kooperatif Tipe STAD Terhadap Hasil Belajar IPA Ditinjau dari Kemampuan Berpikir Kritis Siswa. *Jurnal Imiah Pendidikan Dan Pembelajaran*, *7*(2), 356–364. https://doi.org/10.23887/jipp.v7i2.61027

Lestari, A. C., & Annizar, A. M. (2020). Proses Berpikir Kritis Siswa dalam Menyelesaikan Masalah PISA Ditinjau dari Kemampuan Berpikir Komputasi. *Jurnal Kiprah*, *8*(1), 46–55. https://doi.org/10.31629/kiprah.v8i1.2063

Lusmianingtyas, I., & Sriyanto, S. (2022). Model Pembelajaran Berbasis Masalah untuk Meningkatkan Kemampuan Berpikir Kritis Siswa pada Pembelajaran IPS. *Proceedings Series on Social Sciences & Humanities*, *3*, 520–525. https://doi.org/10.30595/pssh.v3i.322

Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores. *American Journal of Physics*, *70*(12), 1259–1268.

Muhibbuddin, M., Artika, W., & Nurmaliah, C. (2023). Improving Critical Thinking Skills Through Higher Order Thinking Skills (HOTS)-Based Science. *International Journal of Instruction*, *16*(4), 283–296. https://doi.org/10.29333/iji.2023.16417a

Musahrain, Ainurrahmi, Ferniawan, & Sabrina, A. (2024). Analisis Kemampuan Berpikir Kritis Pada Mata Pelajaran IPA SMP Kelas IX Di Kabupaten Sumbawa. *Jurnal PIPA: Pendidikan Ilmu Pengetahuan Alam*, *5*(2). https://doi.org/10.56842/jp-ipa

Nuraina, N., & Nestiadi, A. (2025). Pengembangan Lkpd Berbasis Problem Based Learning Materi Pengaruh Manusia Terhadap Ekosistem Untuk Melatih Kemampuan Berpikir Kritis Siswa SMP. *Eduproxima: Jurnal Ilmiah Pendidikan IPA*, *7*(2), 819–829. http://jurnal.stkippgritulungagung.ac.id/index.php/eduproxima

OECD. (2023). *PISA 2022 Results Factsheets Indonesia PUBE*. https://oecdch.art/a40de1dbaf/C108.

Purwanto, N. (1998). *Psikologi Pendidikan Bandung: Remaja Rosdakarya Offset*.

Qismullah Yusuf, Y., Natsir, Y. M., & Hanum BEd, L. (2015). A Teacher's Experience in Teaching with Student Teams-Achievement Division (STAD) Technique. In *International Journal* (Vol. 8, Issue 2). www.e-iji.net

Ramdani, A., Jufri, A. W., Jamaluddin, J., & Setiadi, D. (2020). Kemampuan Berpikir Kritis dan Penguasaan Konsep Dasar IPA Peserta Didik. *Jurnal Penelitian Pendidikan IPA*, *6*(1), 119–124. https://doi.org/10.29303/jppipa.v6i1.388

Rianti, W., Ayu, C., & Asilestari, P. (2021). Designing Problem Based STAD Learning Models to Improve Students' Writing Skill. *AL-ISHLAH: Jurnal Pendidikan*, *13*(2), 1194–1205. https://doi.org/10.35445/alishlah.v13i2.756

Rofik, F. S., Habiddin, H., & Fajar Marsuki, M. (2025). Student Worksheet of Light and Optics for STEM-Based Problem-Based Learning: Critical Thinking Skills Perspective. In *STEM Education International* (Vol. 1, Issue 1). https://doi.org/10.71289/61kdc115

Sari, S. P., Nugraheni, D., Hamimi, E., Pratiwi, N., & Habiddin, H. (2025). Development of Project Based Learning STEAM based Water Purifier Educational Kit on Water Pollution to Train Critical Thinking Skills. *Jurnal Pendidikan Sains Indonesia*, *13*(1), 233–253.

Sholihah, T. M., & Lastariwati, B. (2020). Problem based learning to increase competence of critical thinking and problem solving. *Journal of Education and Learning (EduLearn)*, *14*(1), 148–154. https://doi.org/10.11591/edulearn.v14i1.13772

Sinaga, R., Purba, M., Sirait, W. S., Sihombing, I., & Siahaan, F. B. (2022). Efektivitas Pendekatan Open-Ended Dengan Model STAD (Student Teams Achievement Divisions) Terhadap Kemampuan Berpikir Kritis Matematis Siswa Pada Materi Sistem Persamaan Linear Dua Variabel (SPLDV) kelas VIII SMP Gajah Mada Medan. *SEPREN: Journal of Mathematics Education and Applied*, *NICoMSE*. https://doi.org/10.36655/sepren.v3i2

Sulistyani, E., & Pratama, L. D. (2024). Pengaruh model pembelajaran STAD berbasis pbl untuk meningkatkan sikap matematis dan berpikir kritis siswa SMP Ihyauddiniyah. *NOTASI: Jurnal Pendidikan Matematika*, *2*(2), 60–68. https://doi.org/10.70115/notasi.v2i2.170

Suradika, A., Dewi, H. I., & Nasution, M. I. (2023). Project-Based Learning and Problem-Based Learning Models in Critical and Creative Students. *Jurnal Pendidikan IPA Indonesia*, *12*(1), 153–167. https://doi.org/10.15294/jpii.v12i1.39713

Widiawati, R., Hikmawati, H., & 'Ardhuha, J. (2022). Pengembangan Perangkat Pembelajaran Berbasis Model Problem Based Learning untuk Meningkatkan Kemampuan Pemecahan Masalah Fisika Peserta Didik pada Materi Fluida Dinamis. *Jurnal Ilmiah Profesi Pendidikan*, *7*(3c), 1803–1810. https://doi.org/10.29303/jipp.v7i3c.857

Yulianti, Y., Lestari, H., Rahmawati, I., Agama, I., & Sahid, I. (2022). Penerapan Model Pembelajaran RADEC Terhadap Peningkatan Kemampuan Berpikir Kritis Siswa. *Jurnal Cakrawala Pendas*, *8*(1). https://doi.org/10.31949/jcp.v6i1.3350

STEM Education International, ***p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx***
Vol. 1, No. 2, December 2025, Page 62-73

STEM-EI
STEM Education International

Comparison of Discovery Learning and Problem-Based Learning Models on Concept Understanding (C1-C4) of Junior High School Students

Wahyu Sofi Martalinda

Department of Science Education, Faculty of Mathematics and Science Education, Universitas Negeri Malang, Indonesia

Corresponding author: Email: wahyushopie1212@gmail.com

Abstract: A teacher must be able to choose a learning model that has the potential to enhance students' understanding of a concept. The purpose of this research is to determine the effects of Discovery Learning (DL) and Problem-Based Learning (PBL) models on the conceptual understanding of junior high school students regarding the material on simple machines. Additionally, this study aims to compare students' understanding when taught using these two models. This quasi-experimental research employs a nonequivalent pretest-posttest control group design, involving two groups of eighth-grade students from a junior high school in Malang Regency for the 2024-2025 academic year. The sampling method used is cluster random sampling. The instruments applied in this study were multiple-choice tests. Several statistical procedures, including normality tests, homogeneity tests, independent sample t-tests, paired samples t-tests, and N-Gain tests, were utilised to analyse the data. The results indicate that both Problem-Based Learning and Discovery Learning positively affect students' conceptual understanding. However, learning through PBL resulted in a higher N-gain (%) in understanding (70.22%) than in the Discovery Learning group (55%). Based on these findings, it can be concluded that the use of DL and PBL can enhance students' understanding of concepts and encourage them to participate more actively in the learning process.

Keywords: Learning Model, Problem-Based Learning, Discovery Learning, Concept Understanding

INTRODUCTION

Competitive and superior human resources are obtained through education. In junior high school education, a deep understanding of concepts is crucial for building a strong foundation for students to face more complex subject matter at higher levels. However, students often struggle to grasp basic concepts, especially in science subjects like physics (Paudi, 2020). One of the main aspects of education is the quality and outcomes of learning (Paramitha et al., 2023). Wabula et al (2020) explain that an educator must have the ability to choose an appropriate teaching model, considering the conditions of the students, the material, and the learning resources. This is done to ensure that the learning model has the potential to be effectively used to enhance students' conceptual understanding.

Low teaching outcomes among students indicate an inability to deliver the educational process effectively to achieve learning objectives (Paramitha et al., 2023). Learning generally centers around the teacher rather than focusing on the students, resulting in monotonous teaching that leads to ignorance and a lack of understanding among students. The education provided in schools often does not meet expectations. The classroom education process is very passive, as it only involves listening, completing tasks, and focusing on textbooks (Ariyani & Kristin, 2021). Students absorb and remember information only during exams (Paputungan et al 2022). As highlighted in a case study by Ardianto et al (2019), many students do not meet the Minimum Completeness Criteria (KKM) during the learning process. Furthermore, there is little classroom learning activity, and conventional teaching models are still used, making it difficult for students to understand the subject matter. Most students ignore the teacher, merely retain information, are reluctant to ask questions, chat with friends, and only receive practice problems from the teacher, which they then complete.

A case study by Sulastry et al. (2023) indicates that, because students are dominated by receiving material directly from the teacher, educators cannot focus on students, resulting in students' inability to develop new understandings from existing theories. Consequently, students find it challenging to grasp concepts. Another case study by Irda et al (2023) reveals that students at SMPN 1 Lawa have a low understanding of simple machines, with only 5.8% demonstrating comprehension. Students' limited knowledge of simple machines is evident. It is estimated that a lack of student understanding leads to disinterest in learning, making it difficult to grasp concepts and identify misconceptions.

The effectiveness of the learning model is essential for fostering students' desire to learn and sparking their curiosity. A practical and suitable teaching model can enhance students' motivation to learn and stimulate their curiosity about the topics studied, leading to a longer retention of the concepts learned (Sari et al., 2021). The effectiveness of the learning model significantly influences student engagement in the learning process. Using the proper methods, students become not only motivated to learn but also more actively participate in discussions and exploration of the material, deepening their understanding.

The recommended learning models in the independent curriculum include Blended Learning, Discovery Learning, Inquiry Learning, Problem-Based Learning (Rohmah et al., 2025), and Project-Based Learning (Bawadi et al., 2023). Each model allows for the development of attitudes, knowledge, and skills. The PBL and DL models are among several solutions in this study. This is because both scientific approach models are based on a problem and emphasise the importance of students taking an active role during the learning process. While PBL focuses on solving real-life issues related to students' daily lives, the DL model encourages students to explore concepts through experience. Both have their advantages and challenges, and choosing the right model can significantly affect students' understanding.

Discovery Learning is a method where students acquire scientific skills and are guided to discover ideas. DL learning tends to be carried out independently by students, but it is not without teacher supervision (Adinata et al, 2022). DL focuses on discovering principles and understanding previously unknown ideas. In this case, the teacher intentionally designs problems for students to solve (Afnan & Syamsudin, 2022). Students can gain knowledge through experiences and interactions with their environment when they actively participate during the learning process. Curiosity is an essential component of effective learning. Therefore, students will remember what they learn for a long time, and the results are not easily forgotten. An efficient method for enhancing students' knowledge through direct experience, exploration, and investigation is through DL learning (Sayangan et al., 2024). This process provides students with opportunities to learn actively and find solutions to existing problems.

The DL learning model can help improve understanding of concepts taught in VIII grade science lessons (Saputri, 2023). By using the DL learning model, students will more actively understand concepts through observation or experimentation. Additionally, this model may broaden and enrich students' understanding of science. The syntax of DL learning includes stimulus (providing stimulation), identifying problems, processing data, verifying data, and drawing conclusions (Bruner, 1961).

Problem-Based Learning is a teaching model that enables students to develop problem-solving and critical thinking skills in the context of real-world problems. This approach also allows students to gain conceptual understanding and essential knowledge from the subject matter (Wibowo, 2018). PBL is an active learning model that emphasises the importance of acquiring knowledge through problem-solving activities, independent study, and small group discussions. Teachers play a crucial role in PBL, as they not only help students learn to solve problems but also to collaborate, study independently, and discover their intrinsic motivation (Shimomura & Utsumi, 2025). PBL is a learning process that enables students to understand concepts through problems (Silvi et al., 2020).

PBL involves students seeking solutions to problems, activating them, and encouraging them to be creative in their search (Sintya Devi & Wira Bayu, 2020). This allows students to be more active in learning. As a result, their creative and critical thinking skills are strengthened, as is their ability to solve problems in real-world situations (Hidayati et al., 2024). By utilising the PBL learning model, students' understanding can be enhanced because they are motivated to find their own solutions to the problems presented (Syarifah et al, 2020).

PBL is largely self-directed. In groups, students collaborate to identify sub-problems, analyse them, and find the facts and information needed to create solutions and address learning problems. Ultimately, students are responsible for solving problems using the knowledge they have acquired (Karttunen et al., 2025). Therefore, the knowledge students create for themselves will remain embedded in their memory for a long time. If students have a strong understanding of the topics studied, they will have better learning capabilities (Supiandi & Julung, 2016). Students will feel interested and motivated to solve these problems because PBL presents real-world issues that require solutions (Silalahi et al., 2023). Orienting students toward problems, managing/organising students, guiding them during investigations, developing and demonstrating results, analysing, and evaluating problem-solving are all parts of the PBL learning process (Lestari et al., 2023).

Previous research has compared the two innovative models of Discovery Learning (DL) and PBL. A study by Gani et al. (2021) reported that in the 2019/2020 academic year, fourth-grade students at SDN Bojongrangkas 01 in Bogor district showed significant improvement in learning about biodiversity in the environment when using the DL and PBL models. The results indicate that students using the Discovery Learning model performed better than those using the PBL model. Furthermore, Paramitha et al. (2023) reported that both DL and PBL models impact conceptual understanding, with students in class XI IPS at a high school in Bojonegoro district during the 2021-2022 academic year showing differences between the two models. Classes using PBL achieved higher grades than those using the DL model. Wabula et al. (2020) support the finding that the PBL model is superior to the DL model in enhancing conceptual understanding. This was conducted in class X at SMA Negeri 1 Ambon. In the PBL class, the average cognitive, psychomotor, and affective scores were higher than in classes using the Discovery Learning model. Another study by Chodijah et al. (2019) indicated that in class XI with science subjects, the Problem-Based Learning (PBL) model is more efficient than the DL model. Although many studies have compared these two learning models across educational levels, such as elementary and high school, no study has specifically examined the effectiveness of DL and PBL in science learning at the junior high school level, particularly for physics material. Previous research, such as that conducted by Gani et al. (2021), Paramitha et al. (2023), Wabula et al. (2020), and Chodijah et al. (2019), shows varied results depending on the level and material studied.

Additionally, previous research at the junior high school level by Junaid et al., (2021) and Siahaan & Sihotang (2023) only investigated one learning model without comparing both teaching models, thus not determining which learning model is most suitable and adequate for enhancing students' conceptual understanding. Junaid et al., (2021) conducted research on the effect of PBL in seventh-grade students at SMPN 17 Tebo on their knowledge of science concepts in the 2020/2021 academic year, finding a significant impact of implementing problem-based learning on seventh-grade students' understanding of physics concepts at SMPN 17 Tebo. Meanwhile, Siahaan & Sihotang (2023) investigated the effect of the discovery learning model on students at SMP Satrya Budi Perdagangan, enhancing their understanding of science concepts. The DL model improved students' knowledge of science, with the final results indicating an average score of 80%.

It is hoped that this study will determine which learning model is superior and serve as a reference for educators in selecting a more effective model to enhance junior high school students' conceptual understanding. Based on these issues, the researcher continues the study with junior high school students, aiming to determine the impact of the Discovery Learning and Problem-Based Learning models on students' conceptual understanding of simple machines and to compare students' knowledge between those taught with both models.

METHOD

This research is a quasi-experiment comparing the Discovery Learning (DL) and Problem-Based Learning (PBL) models in junior high school students' understanding of concepts related to simple machines. The nonequivalent pretest-posttest control group design is used because it allows the researcher to compare two groups that are not randomly assigned, namely the group receiving the DL model treatment and the group receiving the PBL model treatment. Students in the experimental group 1 received the DL treatment, while experimental group 2 received the PBL treatment, studied using this design or concept (Karmila et al., 2020). This design involves measuring before the treatment (pretest) and after the treatment (posttest) for both groups.

The research population comprises all eighth-grade students in junior high schools. The study was conducted in two classes over two weeks. Meetings consisted of two face-to-face sessions in the classroom to deliver the learning models to both experimental classes 1 and 2. The learning duration was 5×35 minutes. The samples in this study are students from classes VIII B and VIII C. The DL model was used to teach class VIII B, while the PBL model was used for class VIII C. The samples were collected using cluster random sampling, ensuring that the class characteristics align with the research objectives. The use of cluster random sampling in this study was chosen because the research population consists of all eighth-grade junior high school students divided into clear groups (clusters), namely eighth-grade classes. Classes VIII B and VIII C were randomly selected as clusters, where each class (VIII B and VIII C) is considered one cluster (Trochim & Donnelly, 2006).

Table 1. *Nonequivalent pretest-posttest control group design*

Group/Class	Pretest	Treatment	Posttest
Experiment 1	O_1	x	O_2
Experiment 2	O_1	y	O_2

Before being used to test conceptual understanding, the research instrument underwent validation, validity, and reliability tests. The reliability of the validation results for the teaching modules was calculated using inter-rater reliability with a similarity percentage technique among validators.

Table 2. Percentage of Similarity in Validation of Discovery Learning and Problem-Based Learning Teaching Modules

	Model	Aspect			Avarage	Criteria
% Similarity	DL	88%	99%	100%	96%	Almost Perfect
	PBL	91%	99%	100%	97%	Almost Perfect

Based on the aspects of general information, core components, and appendices, a result of 96% was obtained, indicating an almost perfect level of similarity. The discovery learning-based teaching module was used for learning in experimental class 1 and PBL in experimental class 2. In this study, multiple-choice questions totalling 20 items have been validated and empirically tested.

Table 3. Empirical Test of Discovery Learning and Problem-Based Learning Questions

Question DL				Question PBL			
Question	r_{xy}	r_{tabel}	Reliability	Question	r_{xy}	r_{tabel}	Reliability
1	0.523	0.455	0,877	1	0.755	0.468	0,697
2	0.546	0.455		2	0.553	0.468	
3	0.468	0.455		3	0.553	0.468	
4	0.684	0.455		4	0.534	0.468	
5	0.805	0.455		5	0.603	0.468	
6	0.843	0.455		6	0.529	0.468	
7	0.843	0.455		7	0.670	0.468	
8	0.843	0.455		8	0.659	0.468	
9	0.609	0.455		9	0.924	0.468	
10	0.808	0.455		10	0.572	0.468	

Based on the calculation where rxy > rtabel and the reliability test results indicating a Cronbach's Alpha value > 0.60 (Alfajri et al., 2019), it can be concluded that the 10 tested questions have met the validity and reliability criteria. Therefore, this test instrument can be considered valid and reliable for use in this research. After these questions met the validity and reliability criteria, they can be used to assess students' conceptual understanding in both experimental classes through pretest and posttest. The instrument for measuring students' knowledge of concepts is multiple-choice questions. The questions were created based on indicators of conceptual understanding.

Data were analysed using t-tests (Paired samples t-test and Independent Sample t-test) after conducting normality tests (Shapiro-Wilk) and homogeneity tests (Levene's). The paired-samples t-test was used to determine whether there was an effect of both the DL and PBL models on understanding a concept, while the independent-samples t-test compared the two models. The N-Gain test was used to measure the improvement in conceptual understanding based on pretest and posttest results. The N-gain results were then categorised according to Hake's (1998) criteria, as shown in Table 4.

Table 4. N Gain Score Categories (Hake, 1998)

N Gain Score	Category	N-Gain Score	Category
Score < 0.3	Low Improvement	< 40%	Ineffective
0.3 ≤ score < 0.7	Moderate Improvement	40-50%	Less Effective
Score ≥ 0.7	High Improvement	56-75%	Quite Effective
		≥ 76%	Effective

RESULTS AND DISCUSSION

The data collected in this study consist of quantitative data, including test scores from both classes, and were analysed using an Independent Samples t-test. The purpose of this analysis is to identify differences in average conceptual understanding between students in experimental classes 1 and 2 by measuring participants' initial abilities through pretests and changes in their conceptual knowledge scores. However, before conducting the independent samples t-test, the data must first be tested for normality and homogeneity. The results are presented in Figure 1.

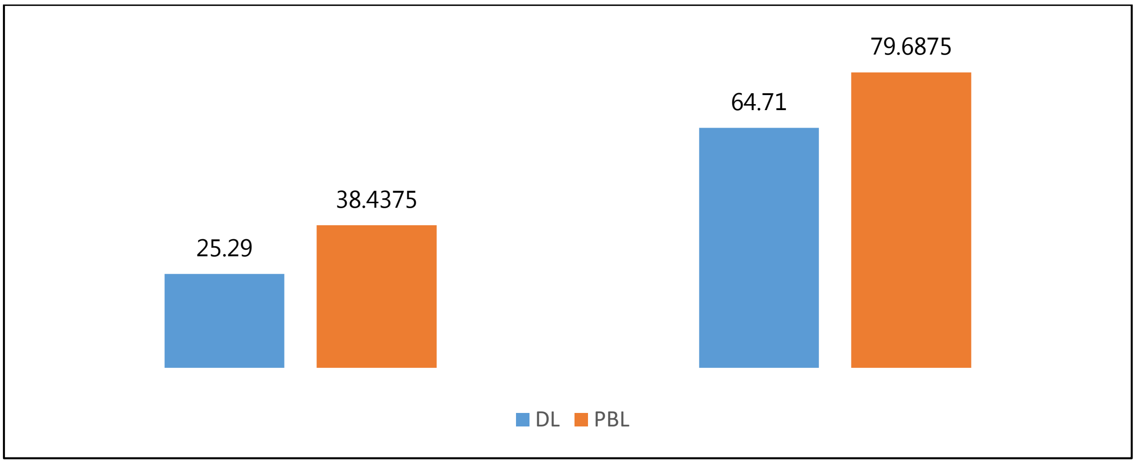

Figure 1. Average Pretest-Posttest Results for PBL & DL

Figure 1 shows that the average pretest score for experimental class 1 was 25.29 before teaching using the Discovery Learning (DL) model. After applying the DL model, the average posttest score for group 1 increased to 64.71. In experimental class 2, with the Problem-Based Learning (PBL) approach, the average pretest score before treatment was 38.44. After using the PBL model, the average posttest score for group 2 also increased to 79.69. Once the pretest data from experimental class 1 and the conceptual understanding results from class 2 were collected, the next step was to conduct normality testing (Shapiro-Wilk test) and homogeneity testing using SPSS version 23, as shown in Tables 5 and 6.

Table 5. Normality and Homogeneity Tests

Test	Class	Sig (Normality)	df_1	df_2	Sig (Homogeneity)
Pretest	Experiment 1	0,69	1	64	0,364
	Experiment 2	0,50			
Posttest	Experiment 1	0,83	1	64	0,744
	Experiment 2	0,90			

Based on the results in Table 5, the normality test indicated that the significance values for both pretest and posttest data in each class were greater than 0.05. Similarly, the homogeneity test also showed significance values greater than 0.05. This means that the data can be considered normal and homogeneous (Qurnia Sari et al., 2017). After confirming that the data were regular and homogeneous, the next step was to conduct the t-test.

The Impact of the Discovery Learning Model on Students' Conceptual Understanding

To analyse the effect of the DL model on students' conceptual understanding, a Paired Samples t-test was used. The results of this test showed a significance value (sig. 2-tailed) of 0.000. Since 0.000 < 0.05 (α = 0.05) (Fauzi et al., 2021). This indicates a significant difference between the pretest and posttest scores. After implementing the

DL model, students demonstrated a different level of conceptual understanding compared to before. The increase in students' conceptual understanding is also evident from the average pretest score of 25.88 and the average posttest score of 64.71 ($\bar{X}_{Post} > \bar{X}_{Pre}$). Thus, the application of the Discovery Learning model positively affects students' understanding of concepts related to simple machines, meaning that H_0 is rejected and H_a is accepted. Students who received this treatment showed a significant improvement in their knowledge after participating in the lessons. This aligns with previous research by Saputri (2023), which indicated that the discovery learning model can help improve the conceptual understanding of eighth-grade students in science at SMP Negeri 13 Makassar.

The syntax used in the discovery learning model also contributes to better conceptual understanding. The stimuli provided through videos and simple experiments successfully boosted students' curiosity within this model. During the problem identification stage, students exhibited strong questioning skills, prompting further research. They were very interested in collecting data and actively sought references to validate their understanding, which contributed to better scores in the post-test. Students demonstrated the ability to analyse data and apply the physics concepts they had learned during the data processing phase. Moreover, they showed meticulousness in verifying the accuracy of their answers during the verification phase.

Despite some challenges, such as a lack of confidence hindering active participation, this model is deemed less effective in conceptual understanding, according to Table 7. Overall results indicate that the DL model can enhance students' conceptual understanding, as noted by Suryaningrum (2023), who explained that the DL model helps students become active in grasping what they learn. Students deepen their understanding of concepts through observation and experiments using the DL model. This model also has the potential to enhance knowledge and make science concepts more enjoyable.

The Impact of the Problem-Based Learning Model on Students' Conceptual Understanding

A Paired Samples t-test was employed to analyse the impact of the PBL model on students' conceptual understanding. The analysis results revealed a significance value (sig. 2-tailed) of 0.000 for the comparison of pretest and posttest scores. Since the significance value of 0.000 is less than α (0.000 < 0.05), it can be concluded that there is a significant difference between the pretest and posttest scores. After implementing the PBL model, students exhibited a noticeable improvement in their conceptual understanding compared to before.

This improvement is reflected in the average pretest score of 38.53 and the average posttest score of 78.53, thus indicating that ($\bar{X}_{Post} > \bar{X}_{Pre}$). This shows that the PBL model has a positive impact on students' understanding, meaning that the null hypothesis (H_0) s rejected and the alternative hypothesis (H_a) is accepted. In other words, the application of the PBL model successfully enhances students' conceptual understanding during the learning process in experimental group 2. This provides empirical evidence that the PBL model is effective in supporting deeper and more interactive learning, which can enhance students' conceptual understanding. This is consistent with previous research by Junaid et al (2021) which found a significant effect of PBL implementation on seventh-grade students' understanding of physics concepts at SMPN 17 Tebo. Wibowo (2018) also stated that PBL is a learning approach that enables students to develop problem-solving and critical thinking skills in the context of real-world problems. This approach allows students to gain conceptual understanding and essential knowledge from the subject matter.

PBL significantly enhances students' understanding, starting with the presentation of videos about real problems that stimulate curiosity and encourage students to ask questions. Students show interest in conducting simple experiments and actively asking questions, especially when using tools like simple devices and Phet practical links. Working on physics problems increases students' desire to seek information, while the use of crossword puzzles reinforces critical thinking and collaboration. In the development and presentation stage, students are able to convey information effectively, strengthening their understanding. Finally, discussions and feedback during problem analysis enhance students' critical thinking skills.

The research findings indicate that PBL not only boosts student interest and participation but also their analytical abilities, in line with findings from other studies by Junaid et al (2021) which emphasises the effectiveness of this method in learning. This model is considered quite effective and shows an impact on understanding, as indicated in Table 7. This aligns with Syarifah etal. (2020), who stated that utilizing the PBL model can enhance students' understanding because they are motivated to find solutions independently to the problems presented. PBL is a learning process that enables students to understand concepts through problems (Silvi et al., 2020).

Comparison Between the Discovery Learning and Problem-Based Learning Models

The study compared the effectiveness of two learning models, namely DL and PBL, in improving students' conceptual understanding. To assess differences, an independent samples t-test was conducted, which is the appropriate statistical method for comparing two groups. The results of this test will indicate the extent to which differences in conceptual understanding between students taught using both models are statistically significant. Below are the results of the independent samples t-test that was conducted:

Table 6. Independent samples t-test

Test	Class		Sig.(2-tailed)
Pretest	Experiment 1 Experiment 2	Equal variances assumed	0,014
Posttest	Experiment 1 Experiment 2	Equal variances assumed	0,008

According to Table 6, the results of the independent samples t-test show a Sig. (2-tailed) value for the pretest of 0.014 < α (0.05). This indicates a difference in students' conceptual understanding between the DL and PBL model classes, as measured by pretest scores–similarly, the Sig. (2-tailed) value for the posttest is 0.008 < α (0.05). This indicates a notable difference in posttest scores between the DL and PBL model classes.

To compare the results of conceptual understanding between the two learning models, an n-gain test was conducted. The primary purpose of this test is to measure each model's effectiveness in enhancing students' understanding by comparing pretest and posttest scores. The results of the n-gain analysis will provide a clear picture of the improvement students achieved after participating in the learning process for each model. The following n-gain test results show the average results for experimental classes 1 and 2.

Table 7. Average Results of Comparison Between Experimental Groups 1 and 2

Model	Average		Improvement	N-gain	Conclusion
	pretest	posttest			
DL	25,29	64,71	39,42	53%	Less effective with moderate improvement
PBL	38,44	79,69	41,25	70,22%	Quite effective with high improvement

Table 7 shows that the average pretest for the DL group is 25.29, while the average posttest is 64.71, resulting in an improvement of 39.42. Despite this improvement, the n-gain for the DL group reaches only 53%, indicating that this model is considered less effective. Conversely, the PBL group shows an average pretest score of 38.44 and a posttest score of 79.69, representing an improvement of 41.25. The n-gain for the PBL group reaches 70.22%, indicating that this model is quite effective in enhancing students' understanding.

Furthermore, Table 7 provides a more detailed overview of the average test improvements between experimental classes 1 and 2. The DL group shows a moderate improvement, indicated by an n-gain of 0.5300. On the other hand, the PBL group demonstrates a more significant progress, with an n-gain of 0.7022, indicating a high level of improvement. These results suggest that the PBL learning model is more efficient at enhancing students' conceptual understanding than the DL model.

This research shows that the PBL model is more effective than the discovery learning (DL) model in improving students' conceptual understanding, which means that H_0 is rejected and H_a is accepted. The average posttest score for PBL is 79.69, with an n-gain of 70.22%, whereas DL achieves 64.71 with an n-gain of 53%. This also aligns with previous research comparing these two innovative models, where Wabula et al., (2020) Confirmed that the PBL model is more effective than DL in enhancing students' understanding. Research conducted in class X at SMA Negeri 1 Ambon showed that the PBL group outperformed the discovery learning group in cognitive, psychomotor, and affective skills. Another study by Paramitha et al (2023) also stated that students taught using the PBL model had higher conceptual understanding than those taught using the DL model in social studies for eleventh-grade students. Similarly, another study by Chodijah etal. (2019) found that the Problem-Based Learning (PBL) model is more effective than the discovery learning model in science subjects for eleventh-grade students.

This is also in line with their learning experiences in experimental class 2, where PBL effectively encourages collaboration and discussion among students, enhancing their engagement in the learning process. This approach creates a more interactive learning experience in which students are directly involved in solving real problems. In contrast, DL tends to show a lack of student enthusiasm despite the stimuli provided. Although simple experiments and material exploration through media can enhance students' active attitudes, some students still struggle to understand the concepts being taught. Thus, PBL is superior to DL because it not only increases active student participation but also helps students understand the material more deeply through observation and direct experimentation to solve real problems. Adinata et al (2022) stated that discovery learning tends to be conducted independently and requires intensive teacher guidance. Additionally, activities in PBL involve experiments and practical work, which aligns with Saraswati et al (2018), who noted that in the learning pyramid, material is more easily absorbed when students conduct observations and experiments directly. Therefore, it can be concluded that PBL creates a more collaborative and interactive learning environment, while DL faces challenges in maintaining student motivation and understanding.

CONCLUSIONS

The results of this study indicate that the use of the Discovery Learning (DL) and Problem-Based Learning (PBL) models has a positive impact on understanding simple machines. Although the findings show that the PBL model is more efficient than the DL model, this does not mean teachers should recommend using the PBL model exclusively for teaching this material. Both models have proven effective and can be utilised in teaching simple machines. However, there is initial evidence that PBL tends to be more effective. Nonetheless, firm conclusions about this comparison need to be verified through further research involving a larger and more diverse student population.

REFERENCES

Adinata, R. E., Septiana, N., & Devi, H. R. P. (2022). Bagaimana meningkatkan pemahaman kegiatan ekonomi melalui pembelajaran discovery learning? (Studi pada sekolah dasar Asy Syafi'iyah Surabaya). *Edunusa: Journal of Economics and Business Education*, *2*(1), 11–17.

Alfajri, A. R., Maizora, S., & Agustinsa, R. (2019). Kepraktisan Soal-Soal Higher Order Thinking Untuk Menghasilkan Soal Yang Praktis Untuk Siswa Kelas Xi Man 1 Kota Bengkulu. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *3*(2), 205–217. https://doi.org/10.33369/jp2ms.3.2.205-217

Chodijah, R. S., Rais, M., & Hadi, N. (2019). Perbedaan Problem Based Learning Dan Discovery Learning Terhadap Pemahaman Sistem Reproduksi Tumbuhan Dan Hewan. *Quagga: Jurnal Pendidikan Dan Biologi*, *11*(2), 55. https://doi.org/10.25134/quagga.v11i2.1863

Fauzi, F., Erna, M., & Linda, R. (2021). The Effectiveness of Collaborative Learning Throughtechniques on Group Investigation and Think Pair Share Students' Critical Thinking Ability on Chemical Equilibrium Material. *Journal of Educational Sciences*, *5*(1), 198. https://doi.org/10.31258/jes.5.1.p.198-208

Hake, R. R. (1998). Interactive-engagement versus traditional methods. *American Journal of Physics*, *66*(1), 64–74.

Junaid, M., Salahudin, S., & Anggraini, R. (2021). Pengaruh Model Pembelajaran Problem Based Learning Terhadap Pemahaman Konsep Ipa Siswa Di Smpn 17 Tebo. *Physics and Science Education Journal (PSEJ)*, *1*(April), 16. https://doi.org/10.30631/psej.v1i1.709

Karmila, H., Amilda, A., & Jayanti, E. (2020). Perbandingan Hasil Belajar Menggunakan Model

Problem Based Learning Dan Think Pair Share Pada Materi Sistem Koloid. *Orbital: Jurnal Pendidikan Kimia*, *4*(1), 75–87. https://doi.org/10.19109/ojpk.v4i1.5506

Paramitha, A. P., Istiqomah, N., & Mastura, S. (2023). The influence of problem-based learning and discovery learning models on learning outcomes. *Jurnal Penelitian Ilmu Pendidikan*, *16*(1), 22–34. https://doi.org/10.21831/jpipfip.v16i1.52423

Qurnia Sari, A., Sukestiyarno, Y., & Agoestanto, A. (2017). Batasan Prasyarat Uji Normalitas dan Uji Homogenitas pada Model Regresi Linear. *Unnes Journal of Mathematics*, *6*(2), 168–177.

Rohmah, A., Fardhani, I., Nugraheni, D., & Habiddin, H. (2025). Profile of Junior High School Students' Problem-Solving Ability on The Topic of Human Relationship to Ecosystems Through The Role-playing Method. *STEM Education International*, *1*(1 SE-Research Articles), 16–28. https://doi.org/10.71289/zrhqax71

Saraswati, R. D., Kusumawardani, R., & Usman. (2018). Analisis Keterampilan Proses Sains Siswa Sma Yang Diajar Menggunakan Model Pembelajaran Pogil Dengan Metode Praktikum Pada Pokok Bahasan Koloid. *Pros.Semanas KPK*, *1*.

Silvi, F., Witarsa, R., & Ananda, R. (2020). Kajian Literatur tentang Kemampuan Pemecahan Masalah Matematika dengan Model Problem Based Learning pada Siswa Sekolah Dasar. *Jurnal Pendidikan Tambusai*, *4*(3), 3360–3368.

Suryaningrum, G. D. (2023). Efektivitas Model Pembelajaran Discovery Learning Ditinjau Dari Kemampuan Berpikir Kritis Siswa Kelas 5 SD. *Jurnal Pendidikan Dan Kebudayaan*, *13*(3), 222–230.

Syarifah, N., Indrawati, V., Sulandjari, S., & Purwidiani, N. (2020). Penerapan Model Pembelajaran Berbasis Masalah Untuk Meningkatkan Hasil Belajar Siswa Pada Kompetensi Dasar Sandwich. *E-Jurnal Education*, *9*(2), 745–754.

Trochim, W., & Donnelly, J. P. (2006). *The Research Methods Knowledge Base*.

Wabula, M., Papilaya, P. M., & Rumahlatu, D. (2020). Pengaruh model pembelajaran discovery learning berbantuan video dan problem based learning terhadap motivasi dan hasil belajar siswa. *Edubiotik : Jurnal Pendidikan, Biologi Dan Terapan*, *5*(01), 29–41. https://doi.org/10.33503/ebio.v5i01.657

Wibowo, H. (2018). *Model dan Teknik Pembelajaran Bahasa Indonesia (Sesuai Kurikulum 2013 Revisi dan Kecakapan Abad XXI*. Puri Cetak Media.

STEM Education International,
Vol. 1, No. 2, December 2025, Page 74-79
p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx

STEM-EI
STEM Education International

Development of a Four-Tier Diagnostic Instrument to Identify High School Students' Understanding of Salt Hydrolysis

Anne Nailul Aziz, Darsono Sigit

Department of Chemistry, Universitas Negeri Malang, Indonesia 65145

**Corresponding author: annelula9@gmail.com*

Abstract: This research aims to (1) produce a four-tier instrument in salt hydrolysis, (2) determine the validity and reliability of the instrument to facilitate students' understanding of salt hydrolysis. The development of the instrument followed the platform developed by Habiddin & Page (2019), which was adapted from the Treagust (1988), which consisted of six steps. The results of instrument validation were 81.74% with a very feasible category, yielding 23 valid questions, with a reliability of 0.7985.

Keywords: understanding, four-tier diagnostic instrument, salt hydrolysis

INTRODUCTION

Chemistry is developed through experiments to answer questions about what, why, and how natural phenomena, particularly those related to the composition, structure, properties, transformations, dynamics, and energetics of matter. Salt hydrolysis is a chemistry topic studied by 11th-grade science students in high schools, particularly in Indonesia (Amala & Habiddin, 2022; Habiddin et al., 2022). Students must not only acquire knowledge but also engage in critical and creative thinking (Nafiah et al., 2025). Therefore, efforts to uncover students' deep understanding help inform the design of proper chemistry teaching. The process of identifying misconceptions can be done using diagnostic tests. Diagnostic tests are used to determine the cause of students' learning failures. A diagnostic test is a test used to identify weaknesses (misconceptions) in specific topics and to provide feedback on students' responses to improve their performance. A four-tier format has been used in many chemistry studies for this purpose, including chemical equilibrium (Tyson et al., 1999), chemical bonding (Amalia & Habiddin, 2024; Peterson et al., 1989; Tan & Treagust, 1999), qualitative analysis (Tan et al., 2002), acid-base properties of salt solutions (Habiddin et al., 2021), thermodynamics (Sreenivasulu & Subramaniam, 2013), metal transition (Sreenivasulu & Subramaniam, 2014), chemical kinetics (Habiddin & Page, 2023; Yan & Subramaniam, 2018) and other topics. The four-tier diagnostic test is an extension of the three-tier multiple-choice diagnostic test, adding a confidence level for each answer and reason (Caleon & Subramaniam, 2010). Adding a confidence rating to each answer and reason can measure differences in students' knowledge levels and help detect the extent of their misconceptions. The four-tier diagnostic test was developed to determine how well students have mastered concepts by measuring their confidence in answering questions.

The first tier of the four-tier diagnostic test consists of multiple-choice questions with three distractors and one correct answer that students must select. The second tier is the students' confidence level in determining their answers. The third tier is the reason students answered the question, consisting of three pre-defined reason options and one open-ended reason. The fourth tier is the students' confidence level in selecting the reason (Habiddin & Nofinadya, 2021; Habiddin & Page, 2019).

METHOD

The instrument development in this study adapted the 6-stage procedures, including (1) concept mapping, (2) testing and interviewing, (3) defining students' unscientific ideas, (4) developing the four-tier prototype, (5) validating the four-tier prototype, and (6) refining the final four-tier instrument (Habiddin & Nofinadya, 2021; Habiddin & Page, 2019). The instrument was evaluated by 2 validators: one lecturer from the Chemistry Department and one chemistry teacher from a public secondary school in Tulungagung, East Java, Indonesia. The initial stage employed open-ended multiple-choice questions and involved 103 students from a public secondary school in Tulungagung. From this mapping, a set of 30 questions was developed and tested with another group of 69 students who had studied salt hydrolysis. The empirical data obtained from students' answers were analysed to determine the validity, reliability, difficulty level, item discrimination, and distractor effectiveness of the four-tier instrument of salt hydrolysis.

RESULTS AND DISCUSSION

A total of 23 test items were identified as valid and reliable. Students choose one answer and one reason that they believe is correct, and they also select their level of confidence in answering the question and providing the reason. The validation was conducted to test the feasibility of the developed four-tier diagnostic instrument and to assess the suitability of the questions, question indicators, and key concepts in the salt hydrolysis material. The content validation results showed that the developed instrument had an average percentage of 81.74%, which falls into the very feasible category (Arikunto, 2021). In empirical validation, a validity level analysis is performed. Based on calculations, the 23 developed questions were declared valid, with r-calculated > r-table at a significance level of 0.05. The results of the item validity level analysis are presented in Table 1.

Table 1. Validity of Items

Soal		1	2	3	4	5	6	7	8	9	10	11	12
A Tier	r_{xy}	0.32	0.62	0.51	0.76	0.75	0.50	0.78	0.36	0.79	0.48	0.87	0.24
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
R Tier	r_{xy}	0.46	0.61	0.23	0.27	0.46	0.49	0.60	0.33	0.64	0.24	0.27	0.28
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
B Tier	r_{xy}	0.39	0.61	0.37	0.51	0.60	0.49	0.69	0.34	0.72	0.36	0.58	0.26
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
Soal		**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	
A Tier	r_{xy}	0.49	0.38	0.55	0.30	0.48	0.40	0.48	0.43	0.53	0.46	0.28	
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	
R Tier	r_{xy}	0.37	0.31	0.29	0.47	0.34	0.40	0.59	0.25	0.26	0.52	0.24	
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	
B Tier	r_{xy}	0.43	0.35	0.42	0.38	0.41	0.40	0.54	0.34	0.39	0.49	0.26	
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	

The reliability of the 23 items tested was 0.883 for the A tier (answers tier), 0.714 for the R tier (reason tier), and 0.7985 for the B tier. In the item discrimination power analysis, in the answer tier (A), there is 1 item categorised as poor, 13 items categorised as fair, 5 items categorised as good, and 4 items categorised as very good. In the reason tier (R), there are 7 items categorised as poor, 9 as fair, and 7 as good. In the both tier (B)/in both tiers, there is 1 item categorised as poor, 15 items categorised as fair, and 7 items categorised as good. The item difficulty level analysis showed that 9 items were categorised as easy, 14 as moderate, and 1 as difficult in the answer tier (A). The test results for difficulty level are presented in Figure 1.

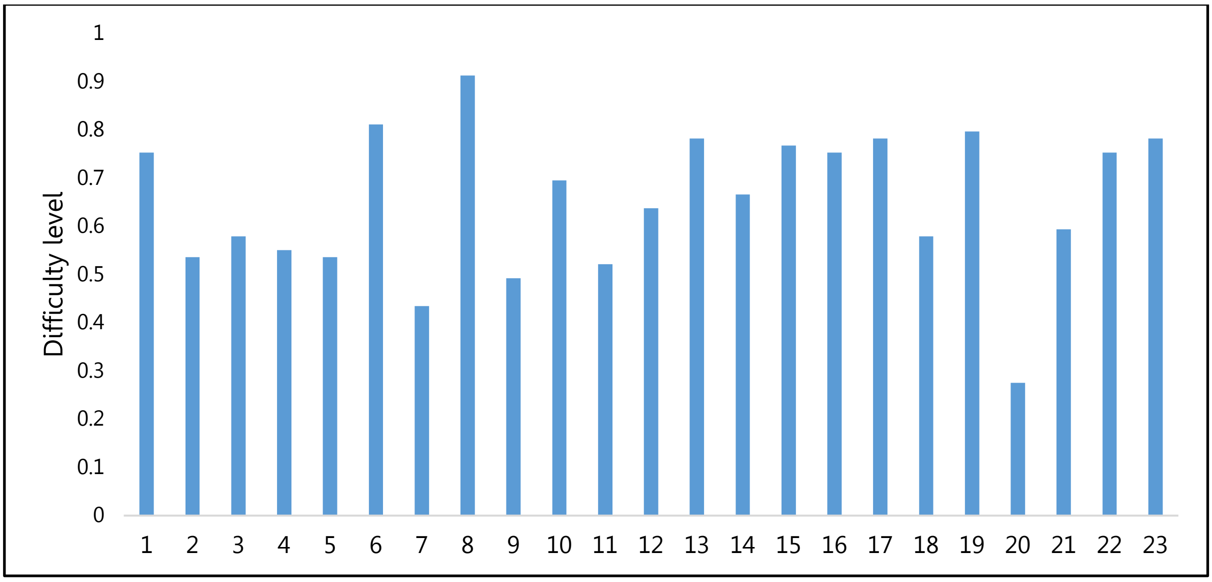

Figure 1. Difficulty indices of Answer Tier (A)

At the reasoning tier (R), 8 questions were considered easy and 15 moderate. The difficulty level test results are presented in Figure 2.

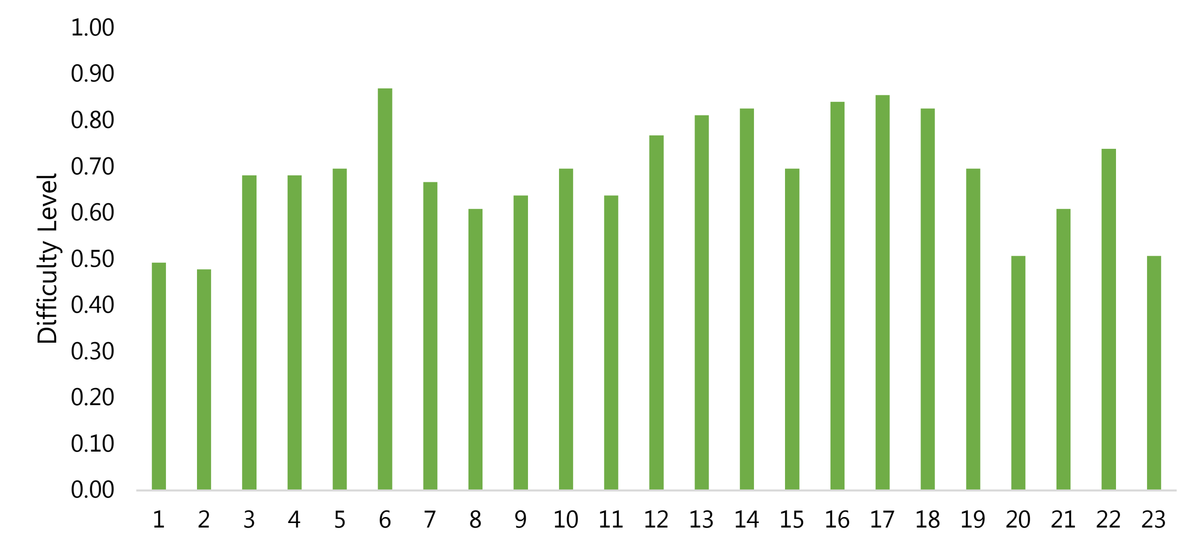

Figure 2. Difficulty indices of Reason Tier (R)

At both tiers (B), there are 9 easy questions, 13 medium questions, and 1 difficult question. The difficulty level test results are presented in Figure 3. The quality of the distractors (their effectiveness) is measured for each multiple-choice question. The criteria for determining whether a distractor is functioning well are met if it is selected by at least 5% of test-takers (Arikunto, 2021). The results of the analysis for each indicator are presented in a table showing the percentage level of each indicator, which represents the analysis of misconceptions occurring for each indicator. The results of the distractor effectiveness calculation are shown in Table 2.

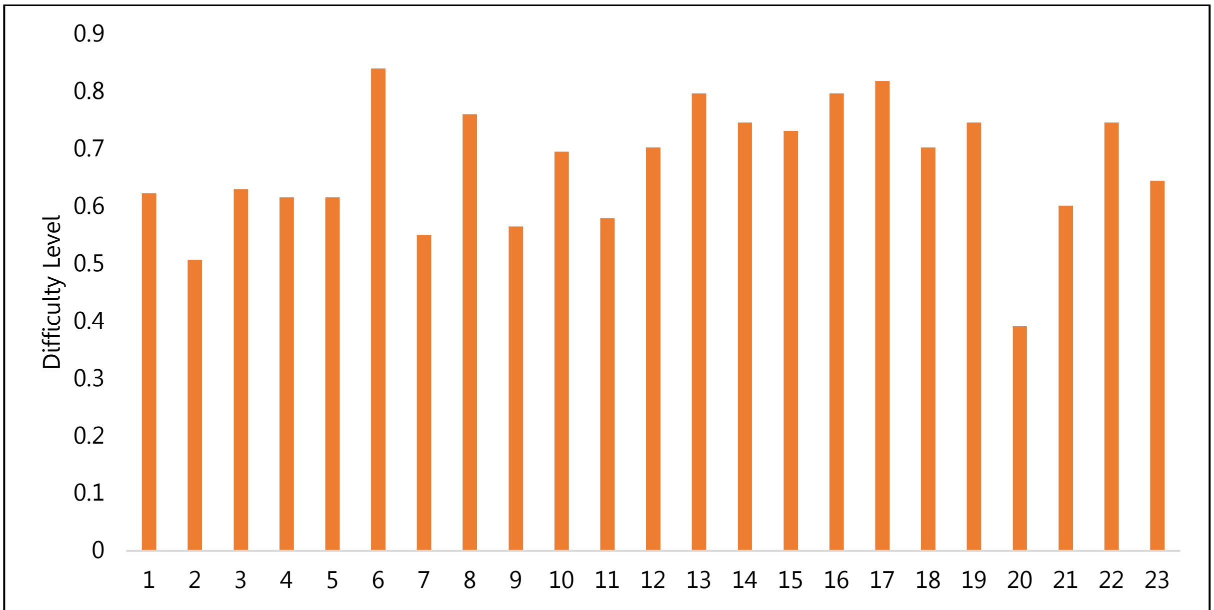

Figure 3. Difficulty indices of Both Tier (B)

Table 2. Distractor Effectiveness of Items

Option	1		2		3		4		5		6		7		8	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	8.70	20.29	53.62	1.45	57.97	8.70	31.88	68.12	31.88	69.57	15.94	8.70	27.54	17.39	0.00	5.80
B	75.36	30.43	24.64	39.13	27.54	68.12	8.70	7.25	14.49	0.00	2.90	2.90	43.48	8.70	2.90	60.87
C	11.59	49.28	2.90	11.59	2.90	15.94	4.35	4.35	0.00	13.04	81.16	86.96	2.90	66.67	5.80	21.74
D	4.35	0.00	18.84	47.83	11.59	7.25	55.07	18.84	53.62	17.39	0.00	1.45	26.09	7.25	91.30	11.59

Option	9		10		11		12		13		14		15		16	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	20.29	13.04	1.45	69.57	52.17	63.77	31.88	0.00	78.26	81.16	66.67	5.80	1.45	69.57	7.25	84.06
B	49.28	63.77	27.54	13.04	15.94	4.35	23.19	13.04	7.25	14.49	11.59	4.35	76.81	13.04	13.04	4.35
C	1.45	5.80	69.57	10.14	7.25	18.84	18.84	10.14	10.14	2.90	2.90	7.25	21.74	4.35	4.35	8.70
D	28.99	17.39	1.45	7.25	24.64	13.04	26.09	76.81	4.35	1.45	18.84	82.61	0.00	13.04	75.36	2.90

Option	17		18		19		20		21		22		23	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	10.14	1.45	8.70	82.61	5.80	69.57	28.99	10.14	59.42	5.80	75.36	73.91	4.35	2.90
B	78.26	1.45	57.97	4.35	79.71	5.80	27.54	18.84	24.64	13.04	4.35	8.70	78.26	13.04
C	4.35	85.51	1.45	2.90	11.59	4.35	30.43	50.72	10.14	60.87	7.25	5.80	7.25	33.33
D	7.25	11.59	31.88	10.14	2.90	20.29	13.04	20.29	5.80	20.29	13.04	11.59	10.14	50.72

The students' response when completing this instrument was that they had never done diagnostic four-tier model questions before. Therefore, some students still felt confused at the beginning of the test, complaining that the questions were multi-page, which made them less enthusiastic. Before the pilot test is conducted, the researcher must also explain in detail the steps for answering the questions. Additionally, students are unfamiliar with microscopic image questions, leading them to provide less severe answers and prompting them to guess when responding. Some students, when answering the level of confidence in choosing answers and reasons, answered carelessly, either guessing everything or answering with complete certainty. This indicates that some students are not taking the given questions seriously.

CONCLUSIONS

The study developed 23 questions on salt hydrolysis in a four-tier format, with a reliability of 0.79, which falls within the acceptable category. The set of 23 questions was derived from the 30 initial items after applying the validation procedures. All the items were also found to be valid and suitable to identify secondary school students' understanding of salt hydrolysis. The confidence level attached to the reason tier for the instrument uses a 5-point scale (1 = guessing, 2 = unsure, 3 = moderate, 4 = confident, 5 = very confident) as proposed in the previous study (Habiddin & Nofinadya, 2021).

REFERENCES

Amala, F., & Habiddin, H. (2022). Pemahaman konsep dalam topik sifat asam basa larutan garam: studi pada siswa SMA di Blitar. *Jurnal Zarah*, *10*(2), 91–100. https://doi.org/10.31629/ZARAH.V10I2.4321

Amalia, E., & Habiddin, H. (2024). Four-tier chemical bonding instrument: Its development and validation. *Edu Sains: Jurnal Pendidikan Sains Dan Matematika*, *12*(1), 21–35. https://doi.org/ttps://doi.org/10.23971/eds.v12i1.5048

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi Aksara.

Caleon, I., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337.

Habiddin, H., Akbar, D. F. K., Husniah, I., & Luna, P. (2022). Uncovering Students'

Understanding: Evidence For The Teaching of Acid-Base Properties of Salt Solution. *Educación Química*, *33*(1), 64–76. https://doi.org/10.22201/FQ.18708404E.2022.1.79488

Habiddin, H., Atikah, A., Husniah, I., Haetami, A., & Maysara, M. (2021). Building scientific explanation: A study of acid-base properties of salt solution. *AIP Conference Proceedings*, *2330*(1), 20047. https://doi.org/10.1063/5.0043215

Habiddin, H., & Nofinadya, S. A. (2021). *The Multi-Tier Instrument in the Area of Chemistry and Science* (M. Bouezzeddine (ed.); p. Ch. 1). IntechOpen. https://doi.org/10.5772/intechopen.100098

Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Habiddin, H., & Page, E. M. (2023). Uncovering Students' Genuine Misconceptions: Evidence to Inform the Teaching of Chemical Kinetics. *Acta Chimica Slovenica*, *70*(2), 184–195. https://doi.org/https://doi.org/10.17344/acsi.2022.7880

Nafiah, N. I., Damayanti, A., Winarno, A. N. P., Akmalia, D. R., Nurdaningrum, F., Azizah, F. A., Juditha, I., Widya, I. S., Afnia, M. N., & Martiningrum, R. (2025). Learning Approach for Enhancing Students' Creativity. *STEM Education International*, *1*(1), 29–35. https://doi.org/10.71289/wrycf010

Peterson, R. F., Treagust, D. F., & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate grade-11 and -12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, *26*(4), 301–314.

Sreenivasulu, B., & Subramaniam, R. (2013). University Students' Understanding of Chemical Thermodynamics. *International Journal of Science Education*, *35*(4), 601–635.

Sreenivasulu, B., & Subramaniam, R. (2014). Exploring Undergraduates' Understanding of Transition Metals Chemistry with the use of Cognitive and Confidence Measures. *Research in Science Education*, *44*(6), 801–828.

Tan, K.-C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching*, *39*(4), 283–301.

Tan, K.-C. D., & Treagust, D. F. (1999). Evaluating Students' Understanding of Chemical Bonding. *School Science Review*, *81*(294), 75–83.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169.

Tyson, L., Treagust, D. F., & Bucat, R. B. (1999). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education*, *76*(4), 554–558.

Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chemistry Education Research and Practice*, *19*(1), 213–226.

STEM Education International, *p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx*
Vol. 1, No. 2, December 2025, Page 80-86

STEM-EI
STEM Education International

Development of Four-Tier Diagnostic Instruments to Identify Students' Understanding of Electrolyte and Non-Electrolyte Solutions

Septiana Ayuningrum Nofinadya*[1], Rima Nuraini[1], Isnanik Juni Fitriyah[2]

[1]Department of Chemistry, Universitas Negeri Malang, Indonesia 65145
[2]Department of Science Education, Universitas Negeri Malang, Indonesia 65145

**Corresponding author: nadyasepti13@gmail.com*

Abstract: This paper outlines the level of validity and reliability of a four-tier diagnostic instrument to identify the level of understanding of students on electrolyte and non-electrolyte solutions. The instrument development follows the stages proposed by Habiddin & Page (2019), namely mapping concepts, testing and interviewing, defining students' unscientific ideas, developing a four-tier prototype, validating the four-tier prototype, and refining the final four-tier instrument. The product eligibility percentage is 89.08%, which falls within the very feasible category.

Keywords: development, four-tier, electrolyte and non-electrolyte solutions, students' conception, unscientific understanding

INTRODUCTION

Difficulties and misconceptions among students regarding electrolyte and non-electrolyte solutions have been reported in the literature. A study in Padang found that 22.5% students from a public secondary school experienced misconceptions in this topic (Fany & Ulianas, 2021). Another study in Kalimantan also found that many students experienced misconceptions in the colligative properties of solutions (Pratiwi et al., 2023). While in another study, 64.70% of students experienced misconceptions about ionic compounds in electrolyte and non-electrolyte solutions (Siswaningsih et al., 2015). Misconceptions can be interpreted as an understanding of concepts that are incompatible with the concept accepted by the scientific community (Habiddin & Nofinadya, 2021). Students cannot explain the theoretical basis of electrolyte and non-electrolyte solutions (Rahmadhany et al., 2023). Or, in other words, students only memorise, remember, and guess to answer questions about related concepts.

Students' misconceptions can affect the learning process. In general, students still maintain misconceptions even after experiencing a different learning process from the previous one (Kirik & Boz, 2012). To identify these misconceptions, an appropriate instrument is required. Multiple-choice and description questions, which are generally used as learning evaluation tools, can be used to measure students' level of understanding. However, both types of instruments have drawbacks. The weakness of the description questions is that there is a subjective tendency when analysing students' answers, and the questions contained in the description questions cannot cover all the material that has been conveyed in learning. Multiple-choice questions, although considered more efficient for identifying students' understanding, also have weaknesses. Multiple-choice questions provide a greater opportunity for students to guess the correct answer than do description questions (Simkin & Kuechler, 2005). Another weakness is that this type of instrument can only evaluate students' content knowledge without considering the reasons behind their choice of answers (Chandrasegaran et al., 2007).

Given these limitations, other instruments are used to assess students' understanding of concepts. The two-tier multiple-choice instrument was among the first to be developed. The instrument consists of two levels: the first is multiple-choice questions, and the second is the reason for choosing an answer at the first level. The weakness of this type of instrument is that it can only identify students' mistakes when they experience misconceptions (Pesman & Eryilmaz, 2010). Then, the form is developed by adding confidence to each item or by using a three-tier multiple-choice instrument. This instrument also has a weakness: students are allowed to choose only a single confidence level. The level of confidence of these learners applies to the answers and reasons for each item. Thus, it cannot be distinguished between students who are sure of the answers and the reasons they choose, and students who are only sure of the answers and not of the reasons, or vice versa. This results in difficulties in assessing and analysing students' answers (Arslan et al., 2012). Based on this description, the researcher wants to develop a four-tier multiple-choice diagnostic instrument. This results in difficulties in assessing and analysing students' answers. Based on this description, the researcher wants to develop a four-tier multiple-choice diagnostic instrument. This results in challenges in determining and analysing students' answers (Arslan et al., 2012). Based on this description, the researcher wants to develop a four-tier multiple-choice diagnostic instrument.

The four-tier diagnostic test is a four-level test, with each question accompanied by a question about the student's level of confidence in each answer and the reason they chose that answer. This instrument allows students to select different levels of confidence in their answers and reasons, so that researchers can clearly determine each student's level of understanding (Habiddin & Page, 2019). This diagnostic test has several advantages, among others. Through a diagnostic test, the teacher can gain a deeper understanding of students' misconceptions. The teacher can also emphasise certain parts of the material that require this when they are explained, and then plan the steps, better learning steps to reduce student misconceptions. Several studies have employed this type of instrument to identify students' understanding and misunderstanding in chemistry (Ardina & Habiddin, 2023; Habiddin et al., 2020, 2022; Husniah et al., 2019).

METHOD

This research and development uses a procedure developed by Habiddin & Page (2019) with six main steps, namely (1) mapping concept, (2) testing and interviewing, (3) defining students' unscientific ideas, (4) developing the prototype of a four-tier, (5) validating the prototype four-tier, (6) refining the final four-tier. At the mapping concept, concept identification, question grid arrangement, and open-reason multiple-choice instruments were prepared. The concept identification stage aims to identify, describe and compile concepts that will identify possible misconceptions. The lecturers and chemistry teachers validated the open-ended multiple-choice instrument developed in the previous step. The feasibility assessment is based on nine criteria/components. The suggestions and comments given are used as the basis for improving or revising the open-reason multiple-choice instrument.

The testing and interviewing stage is a preliminary test conducted using a validated and revised multiple-choice instrument for open-ended responses. The initial data collection was conducted with students at the SMA/MA level who had taken the subject of electrolyte and non-electrolyte solutions. This initial data collection aims to identify and collect students' non-scientific understanding, so that this understanding can be used as an alternative concept in the development of a four-tier instrument.

Based on the data obtained from the preliminary test in the previous stage, an analysis of the answers and reasons was conducted during the stage of defining students' unscientific ideas. All collected reasons were classified into four categories, namely true concepts, false concepts, random or guesswork errors, and no reason.

The development of the prototype four-tier stage begins with preparing and developing a four-tier diagnostic instrument. Then analyse the items used in the initial data collection. This is done to know and correct invalid items. The four-tier diagnostic instrument has been validated to determine its feasibility. The validators referred to in this study were chemistry lecturers of FMIPA UM and chemistry teachers. The validation sheet at this stage differs from the one for multiple-choice open-ended questions because it consists of 10 assessment criteria. The validation results are used as a reference to improve the four-tier diagnostic instrument. The results of the validation, namely the validator's suggestions and comments as qualitative data, while quantitative data is a score given by the validator for each item. After the prototype of the four-tier diagnostic instrument has been validated by the validator, the next step is to validate the empirical (validating the prototype). The parameters used in the empirical validation are reliability and validity, Item difficulty level, item difference power and distractor effectiveness. Based on the results of empirical validation, improvements/revisions were made to the four-tier diagnostic instrument to identify students' level of understanding of electrolyte and non-electrolyte solutions, or the final four-tier refinement stage.

RESULTS AND DISCUSSION

Research and Development Results

The development of a four-tier diagnostic instrument began with an open-ended multiple-choice instrument. The development of an open-reason multiple-choice instrument began with mapping the concepts of electrolyte and non-electrolyte solutions, which served as the basis for formulating 12 indicators of competency achievement (GPA). The competency attainment indicators developed do not measure algorithmic aspects but focus more on conceptual elements. Each predetermined competency achievement indicator is then developed into question indicators. Thus, the number of questions for the multiple-choice open-ended instrument being developed also amounted to 22 items. The open-reason multiple-choice instrument contains a question and four answer choices, but there are also items with three or two answer choices. This is because in these questions, it is complicated to make four homogeneous answer choices. The multiple-choice instrument for the reasons was then validated by two validators: a lecturer in the Department of Chemistry, Mathematics and Natural Sciences at UM and a High School Chemistry Teacher. Meanwhile, the classification results from the initial data collection are used to select reasons or tier reasons for the prototype of the four-tier diagnostic instrument that will be developed. The choice of reasons consisted of one genuine concept and three false concepts, with considerations: concepts that contained misconceptions, logically incorrect concepts, concepts with high frequency, and concepts with a confidence rating. Then, the questions on the multiple-choice instrument that allowed open-ended responses were analysed to determine which items were feasible to develop into a prototype four-tier diagnostic instrument. The results of the analysis of the items in the multiple-choice instrument for open-ended responses were based on several parameters, namely validity, reliability, difficulty level, and item differences.

Table 1. Example of Development Results for a Four-tier Diagnostic Instrument

Indicator of Competence: Identify the type of solution based on its conductivity.		
Indicator	**Question**	**Answer**
Students can identify non-electrolyte solutions from the brief information given.	*FIRST TIER* Sucrose, with the molecular formula $C_{12}H_{22}O_{11}$, is dissolved in water. If viewed from the electrical conductivity, the solution is ... A. Strong electrolyte solution B. Weak electrolyte solution C. Non-electrolyte solution *SECOND TIER* The level of confidence in the selected answer: 1) Just guessing 2) Not sure 3) Moderate 4) Sure 5) Very sure	C

The product of this research and development is a four-tier diagnostic instrument to assess students' conceptual understanding of electrolyte and non-electrolyte solutions, consisting of 13 items. The resulting product specifications are (1) a four-tier diagnostic instrument developed based on KD. 3.8; (2) the developed four-tier diagnostic instrument is equipped with instructions on how to work on questions and a grid of questions consisting of basic competencies, the intended material, indicators of competency achievement, indicators of questions, items; (3) the developed four-tier diagnostic instrument consists of four levels, the first tier is in the form of questions with several answer choices, the second tier is the level of students' confidence in the answers they choose on a scale of 1-5 (1 = only guessing, 2 = not sure, 3 = moderate, 4 = sure and 5 = very sure), the third tier is a choice of four reasons for the first tier, and the fourth tier is the level of confidence of students for reasons on a scale of 1-5. An example of the results of developing a four-tier diagnostic instrument prototype is shown in the Table 2.

Table 2. Example of Development Results for a Four-tier Diagnostic Instrument

Indicators of Competence: Identify the type of solution based on its conductivity	
Question	**Answer**
THIRD TIER Which is the correct reason for the answer chosen? A. $C_{12}H_{22}O_{11}$ cannot be ionised in water, so it cannot conduct electricity. B. $C_{12}H_{22}O_{11}$ has a small number of ions in the solution. C. $C_{12}H_{22}O_{11}$ has ions that can move freely in the solution. D. $C_{12}H_{22}O_{11}$ ionises completely in water so that it can conduct electricity.	A
FOURTH TIER The level of confidence in the selected answer: 1) Just guessing 2) Not sure 3) Moderate 4) Sure 5) Very sure	

Only 13 of the previous 22 items were developed. This occurs because the reasons derived from respondents'/students' answers are insufficient to serve as the third tier of a four-tier diagnostic instrument prototype. The four-tier diagnostic instrument that had been developed was then corrected based on the content validation assessment by two validators: a lecturer in Chemistry at FMIPA UM and a high school chemistry teacher. This was done with the aim that the developed four-tier diagnostic instrument was in accordance with the realms of the material, the realm of construction, and the realm of language. Several items need to be corrected in response to the validator's comments and suggestions.

The revised four-tier diagnostic instrument was subsequently empirically validated, involving 62 respondents from two Class X classes at one of the public high schools in Ponorogo. Before analysing the items based on the empirical validation results, students' scores were calculated for each item. The data were divided into three: the scoring results for the selection of answers (tier answer), the scoring results for the selection of reasons (tier reason), and the scoring results for the second selection (both tiers). The guidelines for scoring the answers and reasons are presented in Table 3.

Table 3. Scoring Guidelines (Habiddin & Page, 2019)

Tier Answer **(A tier)**	*Tier Reason* **(R tier)**	**A combination of A tier and R tier**
Yes (1)	Yes (1)	1
Yes (1)	False (0)	0
False (0)	Yes (1)	0
False (0)	False (0)	0

Data collection was conducted online using Google Forms due to conditions that prevented in-person collection in classrooms because of the spread of the COVID-19 virus. The four-tier diagnostic instrument for identifying students' conceptual understanding of the developed electrolyte and non-electrolyte solution has advantages and disadvantages. The benefits of these four-tier diagnostic instruments include (1) four-tier diagnostic instruments are still not widely developed in chemistry education, especially in electrolyte and non-electrolyte solutions, (2) four-tier diagnostic instruments can be used as tools to identify students' conceptual understanding of electrolyte and non-electrolyte solution material. The weaknesses of the developed four-tier diagnostic instrument include the number of answer choices and the reason for choosing four tiers rather than five, such as the number of answer choices at the high school level.

Results of Content Validation and Empirical Validation

Before developing a four-tier diagnostic instrument as a final product, an open-ended multiple-choice instrument was developed, involving 153 respondents from two X classes at SMAN 3 Malang and three X classes at SMAN 8 Malang, with the criterion that they had taken electrolyte and non-solution material. Electrolyte. The instrument's reliability was in the high category, and all items were valid. In the difficulty level test, there were 5 questions in the easy category, 16 in the medium category, and 1 in the difficult category. Meanwhile, the difference calculation showed that 20 items were in the good category and 2 were in the enough category.

For the developed four-tier diagnostic instrument, both content and empirical validation were carried out. The content validation results showed that 13 items were valid, including those in the very feasible and feasible categories. Meanwhile, the empirical validation of the four-tier diagnostic instrument involved 62 respondents from two X classes at SMAN 2 Ponorogo who had studied electrolyte and non-electrolyte solution materials. The calculation of the parameters used in the empirical validation was divided into scoring results for the selection of answers (tier answer), for the selection of reasons (tier reason), and for the second selection (both tiers). The reliability test of the 13 items showed that the three tiers fell within the high category. The validity of the items on all three tiers was categorised as valid. In the analysis of the item difficulty level across the three tiers, questions were categorised as easy, medium, or difficult. Meanwhile, for the test item difference in the three tiers, it was categorised as sufficient, good and very good. The results of the calculation of the effectiveness of the distractor for scoring the choice of answers (tier answer) and the selection of reasons (tier reason) fell within the same percentage range, namely 6.45% - 72.58%.

CONCLUSIONS

The four-tier diagnostic instrument developed consisted of 13 items with electrolyte and non-electrolyte solutions. Among other advantages, this four-tier diagnostic instrument can serve as an evaluation tool to identify students' conceptual understanding of electrolyte and non-electrolyte solutions. The weakness of the developed four-tier diagnostic instrument, among others, is that the number of answer choices and reasons for the four-tier diagnostic instrument is less than five, such as the number of answer choices at the high school level, so that it can be used as a tool for evaluating learning outcomes based on the school curriculum; adjustments need to be made. Further research is required to disseminate this four-tier diagnostic instrument more widely. This was done to obtain a large number of respondents to assess the product's effectiveness in identifying students' understanding of concepts in electrolyte and non-electrolyte solutions.

REFERENCES

Ardina, D., & Habiddin, H. (2023). Acid-base properties of salt solution: Study at a secondary school in Banyuwangi. *AIP Conference Proceedings*, *2569*(1), 30018. https://doi.org/10.1063/5.0112074

Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A Three-Tier Diagnostic Test to Assess

Pre-Service Teachers' Misconceptions about Global Warming, Greenhouse Effect, Ozone Layer Depletion, and Acid Rain. *International Journal of Science Education*, *34*(11), 1667–1686.

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307.

Fany, L. A., & Ulianas, A. (2021). Analysis of Students Misconception using Two-Tier Multiple Choice Diagnostic Test on Electrolyte and Nonelectrolyte Topic in SMAN 2 Padang. *International Journal of Progressive Sciences and Technologies; Vol 29, No 1 (2021)*. https://doi.org/10.52155/ijpsat.v29.1.3403

Habiddin, H., Akbar, D. F. K., Husniah, I., & Luna, P. (2022). Uncovering Students' Understanding: Evidence for the Teaching of Acid-Base Properties of Salt Solution. *Educacion Quimica*, *33*(1), 64–76. https://doi.org/10.22201/fq.18708404e.2022.1.79488

Habiddin, H., Ameliana, D. N., & Su'aidy, M. (2020). Development of a Four-Tier Instrument of Acid-Base Properties of Salt Solution. *JCER (Journal of Chemistry Education Research)*, *4*(1), 51–57. https://doi.org/10.26740/jcer.v4n1.p51-57

Habiddin, H., & Nofinadya, S. A. (2021). The Multi-Tier Instrument in the Area of Chemistry and Science. In *Insights Into Global Engineering Education After the Birth of Industry 5.0* (pp. 1–17). IntechOpen. https://doi.org/10.5772/INTECHOPEN.100098

Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Husniah, I., Habiddin, H., Sua'idy, M., & Nuryono, N. (2019). Validating an instrument to investigate students' conception of Salt hydrolysis. *Journal of Disruptive Learning Innovation (JODLI)*, *1*(1), 1–6.

Kirik, O. T., & Boz, Y. (2012). Cooperative learning instruction for conceptual change in the concepts of chemical kinetics. *Chemistry Education Research and Practice*, *13*(3), 221–236.

Pesman, H., & Eryilmaz, A. (2010). Development of a Three-Tier Test to Assess Misconceptions About Simple Electric Circuits. *Journal of Educational Research*, *103*(3), 208–222.

Pratiwi, A. N., Erlina, E., Lestari, I., Masriani, M., & Rasmawan, R. (2023). Identification of Students' Misconceptions Using a Four-Tier Multiple Choice Diagnostic Test on Colligative Properties of Solutions. *Jurnal Penelitian Pendidikan IPA*, *9*(11 SE-Research Articles), 10033–10042. https://doi.org/10.29303/jppipa.v9i11.4018

Rahmadhany, D. N., Putri, H. Y., Raudah, N. A., & Arianisa, S. (2023). Analisis Pemahaman Siswa SMA Pada Materi Larutan Elektrolit Dan Non Elektrolit. *Jurnal Sadewa : Publikasi Ilmu Pendidikan, Pembelajaran Dan Ilmu Sosial*, *2*(1 SE-Articles), 209–216. https://doi.org/10.61132/sadewa.v2i1.497

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education*, *3*(1), 73–98. https://doi.org/https://doi.org/10.1111/j.1540-4609.2005.00053.x

Siswaningsih, W., Firman, H., & Rofifah, R. (2015). Pengembangan Tes Diagnostik Two-Tier Berbasis Piktorial Untuk Mengidentifikasi Miskonsepsi Siswa Pada Materi Larutan Elektrolit dan Nonelektrolit. *Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam (JPMIPA)*, *20*(2), 144–149. https://doi.org/https://doi.org/10.18269/jpmipa.v20i2.36236

STEM Education International, *p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx*
Vol. 1, No. 2, December 2025, Page 87-94

STEM-EI
STEM Education International

A Four-Tier Diagnostic Instrument in Acid-Base Properties of Salt Solution: Development Procedure

Rima Nuraini

Department of Chemistry, Universitas Negeri Malang, Indonesia 65145

**Corresponding author: rimaanuraini@gmail.com*

Abstract: This study aims to determine the feasibility of a four-tier diagnostic instrument on salt hydrolysis. The development uses a four-tier diagnostic instrument development procedure by Habiddin & Page (2019) with 6 stages: concept identification, initial test and interview, identification of unscientific student concepts, development of a four-tier diagnostic instrument prototype, prototype validation, and final prototype improvement. The four-tier diagnostic instrument was developed from a multiple-choice instrument open to reasons for capturing student concepts. At last, the finding from this research and development obtained the final product in the form of a four-tier diagnostic instrument with 27 questions that have four levels (tier), the first tier is in the form of questions and answers, second tier is in the level of confidence in the answer chosen, third tier is in the form of selecting the first tier, and the fourth tier is the level of confidence in the reasons chosen. The level of confidence is measured on a scale of 1-5. The instrument developed has an average content validity of 89.45%, with a very decent category and very high reliability (0.858). This shows that the developed four-tier diagnostic instrument is highly feasible for identifying students' misconceptions about salt hydrolysis material.

Keywords: four-tier diagnostic instruments, misconceptions, salt hydrolysis, students' understanding

INTRODUCTION

Salt hydrolysis is one of the chemistry topics taught in 11^{th}-grade high school, according to the 2013 Curriculum. The complex nature of this material lies in the interconnectedness of the concepts being studied with previous concepts. To understand salt hydrolysis well, students are required to understand reaction equilibrium, the dissociation process, and the acid-base properties of reactants and products (Orwat et al., 2017). Additionally, salt hydrolysis is one of the most essential topics in the field of acid-base reactions, yet it is often misunderstood (Secken, 2010). Misconceptions are widely held understandings that do not align with scientific experts' understanding (Pesman & Eryilmaz, 2010). These misconceptions are generally very difficult to change and can persist for a long time, especially if the teacher-designed classroom learning does not facilitate conceptual change (Demircioğlu et al., 2005). Misconceptions that occur in students during learning can hinder their complete understanding of the material.

The research results of Maratusholihah et al. (2017) state that 28.12% of students consider salt hydrolysis to be a reaction between water and salt cations or anions, producing H_3O^+ and OH- ions, because water breaks down the salt into its cations and anions. Additionally, 18.75% of students believe that salts derived from strong acids and weak bases are acidic because they undergo anion hydrolysis, producing H_3O^+ ions, thus increasing the concentration of H_3O^+ ions in water. Furthermore, Orwat et al. (2017) reported that 92% of students correctly stated that $ZnCl_2$ solutions are acidic, but the reaction equations they wrote were incorrect. Based on his research, 55% of students noted that the $MgCl_2$ solution was neutral, and 38% correctly stated that $MgCl_2$ was acidic. Most students who answered correctly wrote the hydrolysis reaction with $Mg(OH)_2$ as a product, with 70% of them using a one-way arrow (→). In comparison, 10% of the students who answered correctly wrote the hydrolysis reaction with $MgOH^+$ as a product. Based on the description, it can be concluded that students have not fully mastered the material on salt hydrolysis.

Students' misconceptions can be identified using several methods, including interviews (Osborne & Gilbert, 1980), concept maps (Novak, 1990), open-ended tests (Taber, 1999), multiple-choice tests (Beichner, 1994), short answer (Billah et al., 2024), Multi-tier instrument (Amala & Habiddin, 2022; Ardina & Habiddin, 2023; Gurel et al., 2015, 2017; Habiddin & Page, 2023; Laliyo et al., 2021) and others. Each instrument used to identify these misconceptions has its own advantages and disadvantages. Among the various methods for identifying misconceptions, the four-tier diagnostic instrument is effective. This test component consists of the first level, which is questions and answers with distractors; the second level is the confidence level of the answers at the first level; the third level is the reason for the answers at the first level; and the fourth level is the confidence level for the chosen reason (Gurel et al., 2017). This four-tier diagnostic instrument allows students to express their different levels of confidence in their answers and reasons, so that students' understanding level can be accurately determined (Habiddin & Page, 2019).

Research on misconceptions regarding salt hydrolysis material has been conducted by Orwat et al. (2017) using questions with four competency tasks, Amelia et al. (2014) using the CRI technique, (Tuysuz, 2009; Ulfah et al., 2024) using a two-tier diagnostic instrument, and Seçken (2010) using multiple-choice and open-ended tests. Based on the literature, no prior research has examined the identification of students' misconceptions about salt hydrolysis using a four-tier diagnostic instrument. Given the advantages of the four-tier diagnostic instrument as described, it is hoped that it will be easier to identify students' understanding of salt hydrolysis.

METHOD

The development of the four-tier diagnostic instrument in this study adapts the procedure developed by Habiddin & Page (2019) based on the two-tier diagnostic instrument development procedure by Treagust (1988), with modifications to suit. There are six stages involved in developing a four-tier diagnostic instrument: (1)

Concept mapping, (2) Initial testing and interviewing, (3) Identifying students' unscientific concepts, (4) Developing a prototype four-tier diagnostic instrument, (5) Validating the prototype, and (6) Refining the final prototype.

The research subjects for the initial test were students from class XI of SMAN 2 Pare, including classes XI IPA 1, XI IPA 3, and XI IPA 5, totalling 96 students. The research subjects for empirical validation were students from class XI of SMAN 2 Pare, including classes XI IPA 6 and XI IPA 7, totalling 71 students. Content validation was carried out by 1 chemistry lecturer and 2 high school chemistry teachers. The instrument used during the initial test was 30 open-ended multiple-choice questions. The instrument used during empirical validation was a 28-question four-tier diagnostic instrument.

The instrument used for content validation of the four-tier diagnostic instrument was a validation questionnaire with ten assessment indicators. Data analysis techniques include content validation, data analysis, and empirical validation. Empirical validation analysis includes test reliability analysis, item difficulty level, item discrimination power, distractor effectiveness, and item validation. An empirical validation analysis was conducted for each tier: A tier (Answer), R tier (Reason), and B tier (Both).

RESULTS AND DISCUSSION

Short Answer Question

Reliability

The test reliability was 0.863, indicating that the test items are highly reliable and can be used to develop a four-tier diagnostic instrument.

Validity

The validity test results show that 27 items are valid and 3 are not, namely items 5, 20, and 24. The invalid items are considered for revision.

Table 1. Validity of short answer questions

No	R	category	No	R	category	No	R	category
1	0.655	Valid	11	0.550	Valid	21	0.351	Valid
2	0.569	Valid	12	0.440	Valid	22	0.417	Valid
3	0.488	Valid	13	0.334	Valid	23	0.460	Valid
4	0.457	Valid	14	0.511	Valid	24	0.001	Invalid
5	0.191	Invalid	15	0.496	Valid	25	0.440	Valid
6	0.327	Valid	16	0.361	Valid	26	0.515	Valid
7	0.571	Valid	17	0.414	Valid	27	0.474	Valid
8	0.558	Valid	18	0.528	Valid	28	0.571	Valid
9	0.422	Valid	19	0.369	Valid	29	0.543	Valid
10	0.658	Valid	20	0.070	Invalid	30	0.663	Valid

Difficulty Level (P)

Table 2 shows that there are 6 easy questions, 22 moderate questions, and 2 difficult questions.

Table 2. Difficulty level of short answer questions

No	P	Category	No	P	Category	No	P	Category
1	0.58	moderate	11	0.30	difficult	21	0.52	moderate
2	0.71	easy	12	0.46	moderate	22	0.71	easy
3	0.41	moderate	13	0.88	easy	23	0.64	moderate
4	0.48	moderate	14	0.55	moderate	24	0.39	moderate
5	0.95	easy	15	0.52	moderate	25	0.60	moderate
6	0.30	difficult	16	0.63	moderate	26	0.68	moderate
7	0.66	moderate	17	0.58	moderate	27	0.66	moderate
8	0.72	easy	18	0.69	moderate	28	0.55	moderate
9	0.64	moderate	19	0.50	moderate	29	0.68	moderate
10	0.54	moderate	20	0.55	moderate	30	0.70	easy

Distractor effectiveness (D)

The results of the distractor effectiveness calculation show that 16 questions have ineffective distractors, as the students who chose those distractors did not constitute 5% of the total test takers. Based on the analysis of Tables 1, 2, and 3, it is concluded that questions 5 and 24 were not selected for development into a four-tier diagnostic instrument. The four-tier diagnostic instrument was developed based on 28 open-ended multiple-choice questions.

Table 3. The percentage of the distractor effectiveness of short answer questions

No / Opt	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	14.6	69.4	43.0	15.4	0.0	26.7	3,2	71.8	12.6	17.8	10.8	27.8	6.4	27.3	25.8
B	12.5	13.3	17.2	50.5	2.0	16.7	67,7	15.6	7.4	18.9	34.9	18.9	89.4	60.2	16.1
C	14.6	8.20	26.9	26.4	94.8	24.4	3,2	1.0	63,5	5.6	12.0	4.4	4.2	9.0	53.8
D	58.3	6.20	13.0	13.2	3.1	32.2	25,8	11.5	15.7	57.8	42.2	48.9	0.0	3.4	7.5
No / Opt	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
A	5.7	10.2	15.9	52.7	59.5	16.3	75.6	1.0	22.2	11.5	73.9	4.5	12.8	4.3	12.8
B	24.1	23.9	75.0	18.7	32.6	23.9	18.9	66.3	25.9	0.0	10.2	4.5	19.1	12.9	77.9
C	68.9	7.9	7.9	14.3	6.7	54.3	3.3	30.4	45.7	60.4	11.4	20.2	56.4	12.9	2.3
D	1.1	63.6	1.1	14.3	0.0	5.4	2.2	2.1	6.1	23.0	4.5	70.8	11.7	69.9	6.9

Four-tier instrument

Content Validity

The average percentage of instrument feasibility obtained based on content validation was 89.45%. According to Arikunto's (2015: 89) criteria for feasibility levels, the four-tier diagnostic instrument developed by the researcher met the very feasible criteria, so no significant revisions were needed. The four-tier diagnostic instrument was only partially revised in response to suggestions from the validators prior to testing.

Empirical validity

Reliability. Reliability for the B tier (0.858) is higher than for the A tier (0.864) and R tier (0.775). Based on the analysis, the reliability level for the A tier is very high, and for the R tier, it is high. Meanwhile, the reliability level for the entire test (B tier) is very high.

Difficulty Level. Based on the average, the developed instrument is moderately difficult. The average of the difficulty index for the A tier (0.53) is lower than that for the R tier (0.55), indicating that more students chose the correct option for the reason (R tier) than for the answer (A tier). This suggests that most students understand the concept well. Meanwhile, the difficulty index for the B tier (0.42) is lower than that for the A and R tiers because, to answer correctly, students must have a good understanding.

Table 4. Difficulty Level of A, R, and B tiers.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A *tier*	0.65	0.80	0.59	0.42	0.56	0.65	0.66	0.51	0.65	0.38	0.27	0.56	0.69	0.73
R *tier*	0.66	0.72	0.61	0.44	0.75	0.61	0.54	0.72	0.49	0.34	0.30	0.41	0.58	0.76
B *tier*	0.59	0.70	0.51	0.41	0.49	0.55	0.45	0.46	0.34	0.28	0.15	0.34	0.52	0.69
No	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
A *tier*	0.27	0.21	0.23	0.66	0.46	0.46	0.52	0.38	0.61	0.56	0.61	0.55	0.63	0.63
R *tier*	0.45	0.41	0.39	0.55	0.51	0.75	0.44	0.31	0.49	0.35	0.66	0.55	0.83	0.73
B *tier*	0.15	0.17	0.17	0.49	0.41	0.42	0.35	0.25	0.41	0.34	0.51	0.45	0.62	0.58

Discriminatory Indices. The analysis results show that the DI for the A, R, and B tiers ranged from poor to good, with no test items having a very good DI. Items 12 and 16 each had a negative DI value of -0.10 and -0.04, respectively. This indicates that the questions cannot distinguish between students with good conceptual understanding and those with low conceptual understanding, so the questions need to be revised. However, there are several considerations before making revisions. In some cases, items with low DI values can be retained because the primary purpose for developing the items was to identify students' conceptual understanding, not to differentiate between high-achieving and low-achieving students (Habiddin & Page, 2019).

Table 5. Discriminatory indices of A, R, and B tiers using Pearson Correlation

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A *tier*	0,49	0,40	0,38	0,44	0,43	0,38	0,57	0,49	0,21	0,36	0,25	0,38	0,52	0,43
R *tier*	0,52	0,46	0,41	0,30	0,51	0,18	0,27	0,29	0,30	0,27	0,08	-0,10	0,69	0,43
B *tier*	0,49	0,54	0,49	0,30	0,47	0,24	0,38	0,47	0,27	0,27	0,25	0,10	0,69	0,52
No	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
A *tier*	0,19	0,13	0,16	0,57	0,13	0,35	0,24	0,02	0,41	0,43	0,29	0,41	0,40	0,40
R *tier*	0,21	-0,04	0,33	0,52	0,15	0,35	0,35	0,10	0,47	0,07	0,35	0,41	0,12	0,20
B *tier*	0,25	0,05	0,16	0,58	0,13	0,33	0,36	0,16	0,52	0,16	0,44	0,49	0,43	0,29

Distractor Effectiveness. Based on the analysis results, most distractors are effective, as 84.5% were chosen by more than 5% of test participants.

Table 6. Distractor Effectiveness for each option

No	1		2		3		4		5		6		7	
Opt	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*
A	8.45	66.20	80.28	2.82	14.08	22.54	5.63	5.63	9.86	7.04	26.76	18.31	12.68	11.27
B	19.72	7.04	8.45	71.83	59.15	14.08	42.25	8.45	14.08	4.23	4.23	1831	11.27	19.72
C	7.04	15.49	5.63	11.27	12.68	60.56	4.23	43.66	19.72	74.65	4.23	2.82	66.20	53.52
D	64.79	11.27	5.63	14.08	14.08	2.82	47.89	42.25	56.34	14.08	64.79	60.56	9.86	15.49
No	**8**		**9**		**10**		**11**		**12**		**13**		**14**	
Opt	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*
A	50.70	11.27	19.72	30.99	38.03	7.04	21.13	29.58	25.35	33.80	14.08	14.08	73.24	11.27
B	15.49	4.23	7.04	49.30	42.25	33.80	33.80	14.08	56.34	40.85	12.68	14.08	16.90	76.06
C	19.72	12.68	64.79	12.68	8.45	16.90	18.31	14.08	12.68	18.31	69.01	14.08	4.23	1.41
D	14.08	71.83	8.45	7.04	11.27	42.25	26.76	42.25	5.63	7.04	4.23	57.75	5.63	11.27

No	15		16		17		18		19		20		21	
Opt	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*
A	26.76	45.07	32.39	25.35	18.31	14.08	19.72	53.52	11.27	49.30	7.04	74.65	15.49	16.90
B	2254	12.68	25.35	23.94	40.85	39.44	66.20	14.08	25.35	26.76	16.90	9.86	12.68	16.90
C	28.17	18.31	21.13	40.85	22.54	26.76	5.63	16.90	46.48	15.49	29.58	12.68	52.11	43.66
D	22.54	23.94	21.13	9.86	18.31	19.72	8.45	14.08	45.07	7.04	46.48	2.82	19.72	22.54
Soal	**22**		**23**		**24**		**25**		**26**		**27**		**28**	
Opsi	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*	A *tier*	R *tier*
A	53.52	49.30	60.56	21.13	4.23	35.21	60.56	12.68	14.08	54.93	2.82	4.23	63.38	9.86
B	38.03	30.99	11.27	18.31	23.94	39.44	19.72	66.20	16.90	16.90	28.17	83.10	9.86	11.27
C	5.63	14.08	18.31	11.27	56.34	15.49	9.86	14.08	14.08	12.68	63.38	8.45	22.54	73.24
D	2.82	5.63	9.86	49.30	15.49	9.86	9.86	7.04	54.93	15.49	5.63	4.23	4.23	5.63

Validity. The analysis results show that most of the developed questions are valid, but some items are not. A total of 3 questions were invalid at the A tier, 5 questions were invalid at the R tier, and 4 questions were invalid at the B tier. These invalid questions need to be considered for revision based on other parameters, namely difficulty level, discrimination index, and distractor effectiveness. Based on empirical validation analysis, item 16 was discarded, items 9, 10, 11, 12, and 19 were retained with revisions, and items 15 and 22 did not require revision.

Table 7. Validity of A, R, and B tiers

No		1	2	3	4	5	6	7	8	9	10	11	12	13	14
A *tier*	r_{xy}	0.483	0,.01	0.399	0.456	0.497	0.488	0.621	0.544	0.367	0.388	0.417	0.429	0.615	0.628
R *tier*	r_{xy}	0.624	0.611	0.577	0.329	0.608	0.308	0.377	0.373	0.229	0.389	0.119	-0.140	0.660	0.452
B *tier*	r_{xy}	0.582	0.603	0.628	0.417	0.613	0.407	0.547	0.556	0.181	0.421	0.344	0.121	0.728	0.645
No		**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
A *tier*	r_{xy}	0.286	0.179	0.400	0.641	0.216	0.424	0.295	0.106	0.511	0.560	0.447	0.476	0.484	0.469
R *tier*	r_{xy}	0.224	0.044	0.484	0.574	0.233	0.329	0.483	0.231	0.517	0.283	0.481	0.512	0.232	0.262
B *tier*	r_{xy}	0.442	0.196	0.371	0.646	0.274	0.331	0.534	0.201	0.561	0.320	0.576	0.509	0.475	0.407

CONCLUSIONS

The resulting product development is a four-tier diagnostic instrument to identify misconceptions of 11th-grade science students regarding salt hydrolysis material. The developed instrument consists of 27 questions. The specifications of the resulting product are: (1) The developed four-tier diagnostic instrument consists of four tiers, with the first tier being questions and answers with four answer options, the second tier representing students' confidence level in choosing the first tier on a scale of 1-5 (1=guessing only; 2=not sure; 3=moderate; 4=sure; 5=very sure), the third tier being the reason for the first tier, and the third tier being the confidence level for the reason on a scale of 1-5 (1=guessing only; 2=not sure; 3=moderate; 4=sure; 5=very sure); (2) The reason choices used are based on students' reasons in the open-ended multiple-choice initial test and relevant literature; (3) The developed four-tier diagnostic instrument consists of at least one question per indicator; (4) The instructions for answering the questions in the developed instrument include general instructions for answering the presented questions.

REFERENCES

Amala, F., & Habiddin, H. (2022). Pemahaman konsep dalam topik sifat asam basa larutan garam: studi pada siswa SMA di Blitar. *Jurnal Zarah*, *10*(2), 91–100. https://doi.org/10.31629/Zarah.V10I2.4321

Amelia, D., Marheni, M., & Nurbaity, N. (2014). Analisis Miskonsepsi Siswa Pada Materi Hidrolisis Garam Menggunakan Teknik CRI (Certainty Of Response Index) Termodifikasi. *JRPK: Jurnal Riset Pendidikan Kimia*. https://doi.org/10.21009/jrpk.041.05

Ardina, D., & Habiddin, H. (2023). Acid-base properties of salt solution: Study at a secondary school in Banyuwangi. *AIP Conference Proceedings*, *2569*(1), 30018. https://doi.org/10.1063/5.0112074

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, *62*(8), 750–762. https://doi.org/10.1119/1.17449

Billah, A. A., Octavia, D. A., Sholihah, F. H., Febriliyanto, M. R., Aisyah, N. N., Rohmah, N. H., Rahayu, S. M. S., Salsabela, T. A., Basimin, M. Q., & Habiddin, H. (2024). University Students' Understanding of Intermolecular Forces: A Comparison of First & Third-Year Students. *STEM Education International*, *1*(1 SE-Research Articles), 1–7. https://sciencesustain.com/index.php/stem/article/view/5

Demircioğlu, G., Ayas, A., & Demircioğlu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice*, *6*(1), 36–51. https://doi.org/10.1039/B4RP90003K

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science & Technological Education*, *35*(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094

Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Habiddin, H., & Page, E. M. (2023). Uncovering Students' Genuine Misconceptions: Evidence to Inform the Teaching of Chemical Kinetics. *Acta Chimica Slovenica*, 184–195.

Laliyo, L. A. R., Hamdi, S., Pikoli, M., Abdullah, R., & Panigoro, C. (2021). Implementation of four-tier multiple-choice instruments based on the partial credit model in evaluating students' learning progress. *European Journal of Educational Research*, *10*(2), 825–840. https://doi.org/10.12973/EU-JER.10.2.825

Maratusholihah, N. F., Sri, R., & Fauziatul, F. (2017). Analisis Miskonsepsi Siswa SMA pada Materi Hidrolisis Garam dan Larutan Penyangga. *Jurnal Pendidikan:Teori, Penelitian, Dan Pengembangan*, *2*(7), 919—926.

Novak, J. D. (1990). Concept Mapping - A Useful Tool For Science-Education. *Journal of*

Research in Science Teaching, *27*(10), 937–949.

Orwat, K., Bernard, P., & Mikuli, A. M. (2017). Alternative Conceptions of Common Salt Hydrolysis Among Upper-Secondary-School Students. *Journal of Baltic Science Education*, *16*(1), 64–76. https://doi.org/https://doi.org/10.33225/jbse/17.16.64

Osborne, R. ., & Gilbert, J. K. (1980). A Method for Investigating Concept Understanding in Science. *European Journal of Science Education*, *2*(3), 311–321.

Pesman, H., & Eryilmaz, A. (2010). Development of a Three-Tier Test to Assess Misconceptions About Simple Electric Circuits. *Journal of Educational Research*, *103*(3), 208–222.

Seçken, N. (2010). Identifying student's misconceptions about salt. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2010.03.004

Taber, K. S. (1999). Ideas About Ionisation Energy: A Diagnostic Instrument. *School Science Review*, *81*(295), 97–104.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169.

Tuysuz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essays*, *4*(6), 626–631.

Ulfah, M., Ulfah, M., Erlina, E., Pratiwiningrum, F. M., Wafiq, A. F., & Juahir, Y. B. (2024). Misconceptions of Intermolecular Forces in General Chemistry Courses. *Jurnal IPA & Pembelajaran IPA*, *8*(1), 39–51. https://doi.org/10.24815/jipi.v8i1.35641