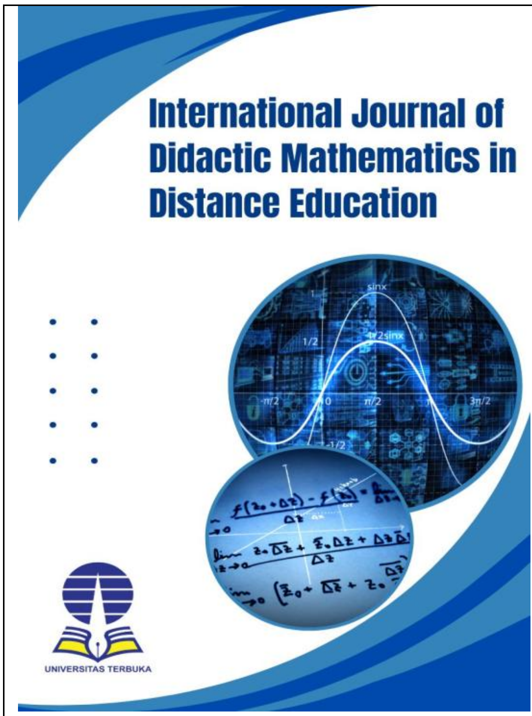
International Journal of Didactic Mathematics in Distance Education

Journal homepage: https://jurnal.ut.ac.id/index.php/ijdmde



Effectiveness of ChatGPT-integrated discovery learning on mathematical literacy in three-variable linear equation systems: A quasi-experimental study

Rismun Sufia Monike[1], Sudirman Sudirman[2], Thesa Kandaga[3], Camilo Andrés Rodríguez-Nieto[4]

[1]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, sufiamonike@gmail.com

[2]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, sudirman.official@ecampus.ut.ac.id

[3]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, thesa.official@ecampus.ut.ac.id

[4]Department of Exact Sciences, Universidad de la Costa, Barranquilla, Colombia, crodrigu79@cuc.edu.co

To cite this article:
Monike, R.S., Sudirman, S., Kandaga, T., & Rodríguez-Nieto, C.A. (2026). Effectiveness of ChatGPT-integrated discovery learning on mathematical literacy in three-variable linear equation systems: A quasi-experimental study. *International Journal of Didactic Mathematics in Distance Education*, *3*(1), 64-82.
To link to this article:
https://jurnal.ut.ac.id/index.php/ijdmde
Published by:
Universitas Terbuka
Jl. Pd. Cabe Raya, Pd. Cabe Udik, Kec. Pamulang, Kota Tangerang Selatan, Banten 15437

International Journal of Didactic Mathematics in Distance Education
Volume 3, No 1, pp. 64-82, E-ISSN: 3047-9207
https://jurnal.ut.ac.id/index.php/ijdmde

Effectiveness of ChatGPT-integrated discovery learning on mathematical literacy in three-variable linear equation systems: A quasi-experimental study

Rismun Sufia Monike[1*], Sudirman Sudirman[2], Thesa Kandaga[3], Camilo Andrés Rodríguez-Nieto[4]
[1]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, sufiamonike@gmail.com
[2]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, sudirman.official@ecampus.ut.ac.id
[3]Department of Mathematics Education, Universitas Terbuka, Jakarta, Indonesia, thesa.official@ecampus.ut.ac.id
[4]Department of Exact Sciences, Universidad de la Costa, Barranquilla, Colombia, crodrigu79@cuc.edu.co

*Coresponding Author: sufiamonike@gmail.com

Abstract
This quasi-experimental study investigated the effectiveness of ChatGPT-integrated discovery learning on high school students' mathematical literacy in three-variable linear equation systems, while examining the moderating role of self-efficacy. Eighty tenth-grade students from a public high school in Indonesia were assigned to either an experimental group receiving ChatGPT-supported discovery learning instruction or a control group receiving conventional instruction, with mathematical literacy assessed through problem-solving tests and self-efficacy measured using validated questionnaires. Statistical analyses using non-parametric tests, including Wilcoxon Signed Rank Test, Mann-Whitney U Test, and Rank-Based ANCOVA, revealed that the experimental group achieved significantly higher post-test scores ($M=79.25$, $SD=8.42$) compared to the control group ($M=65.30$, $SD=9.18$), yielding a large effect size ($d=1.24$, $p<.001$) that substantially exceeded typical outcomes reported in meta-analyses of computer-assisted mathematics instruction. Contrary to established theoretical frameworks positing strong relationships between self-efficacy and mathematics achievement, this study found that self-efficacy neither significantly predicted mathematical literacy ($F=3.171$, $p=.083$, $\eta^2=.077$) nor interacted with the learning model, suggesting that ChatGPT's pedagogical benefits operated uniformly across students with varying confidence levels. These findings suggest that ChatGPT functioned as adaptive conversational scaffolding that provided immediate formative feedback, facilitated metacognitive engagement through dialogic prompting, and generated multiple mathematical representations to support conceptual understanding. The null self-efficacy effect may indicate that robust AI scaffolding serves as an external compensatory mechanism that equalizes learning opportunities by reducing students' dependence on internal confidence judgments, thereby contributing novel theoretical insights into the conditional nature of motivational constructs in AI-mediated learning environments and offering practical implications for implementing conversational AI in mathematics education.

Article History
Received:
24 July 2025
Revised:
24 October 2025
Accepted:
28 October 2025
Published Online:
03 December 2025

Keywords:
ChatGPT; Discovery learning; Mathematical literacy; Quasi-experimental study; Three-variable linear equations

1. Introduction

Mathematical literacy represents a crucial competency that students must possess in the 21st century, encompassing not only computational abilities but also the capacity to reason, interpret, and apply mathematical concepts in everyday life. The Programme for International Student Assessment (PISA) measures mathematical literacy as the ability of 15-year-old students to formulate, employ, and interpret mathematics in various contexts to describe, predict, and explain phenomena (OECD, 2024). However, Indonesian students' mathematical literacy achievement remains below the OECD average, with nearly 70% of Indonesian students scoring below level 2 across all assessed topics (Stacey, 2011; Gustiningsi et al., 2024), demonstrating the urgent need for improved teaching methods that emphasize not only formula memorization but also the development of conceptual understanding and higher-order thinking skills. The topic of three-variable linear equation systems (SPLTV) presents particular challenges as it requires advanced analytical, representational, and problem-solving abilities, necessitating instructional approaches that facilitate active knowledge construction.

The discovery learning model has proven effective in mathematics education by positioning students as active learners who explore and construct their own knowledge. Research by Ramadhani et al. (2023) demonstrates that discovery learning provides opportunities for students to discover and explore various methods of solving specific mathematical problems, while also increasing students' interest in learning mathematics. Siregar et al. (2020) found that implementing a discovery learning module in geometry effectively enhanced secondary school students' mathematical reasoning skills, communication, and self-confidence. A study by Milenković et al. (2025) also confirmed that discovery-based learning using physical manipulatives had a significantly positive impact on students' geometry knowledge with a medium effect size on knowledge retention. However, despite various studies confirming the effectiveness of discovery learning in traditional educational contexts, the integration of artificial intelligence, particularly ChatGPT, within this model remains severely underexplored empirically. Most AI research in education has focused on language learning or adaptive instructional systems (Almarashdi et al., 2024), while its potential as a learning assistant capable of providing real-time feedback and guiding the discovery process in mathematics has rarely been examined theoretically or empirically.

ChatGPT as a large language model possesses conversational capabilities that can facilitate pedagogical dialogue and provide cognitive scaffolding. A scoping survey by Pepin et al. (2025) indicates that ChatGPT shows potential in enhancing self-regulated learning, providing real-time feedback, and supporting critical thinking in mathematics education A systematic review by Almarashdi et al. (2024) found that teachers use ChatGPT to generate examples, assess difficulty levels, provide explanations, support problem-solving, and prepare tests, with results showing that ChatGPT is promising in enhancing student engagement and understanding. Gouia-Zarrad and Gunn (2024) in their study also confirmed that ChatGPT can enhance students' learning experiences in mathematics classrooms. However, a systematic review of Scopus-indexed articles from 2023-2025 identified significant research gaps: the impact of ChatGPT on students' critical thinking development remains unclear, there is a lack of comparative research on ChatGPT's effectiveness across different educational levels, and limited practical guidelines exist for integrating ChatGPT into existing mathematics curricula (Farrokhnia et al., 2024). The question of how ChatGPT can specifically optimize the stages of discovery learning to enhance mathematical literacy on the SPLTV topic remains an unanswered area.

Beyond instructional models, students' self-efficacy, their belief in their own capabilities, also plays a crucial role in academic achievement. Zakariya (2022) in his systematic review found that self-efficacy predicts students' mathematics achievement, with all four sources of mathematics self-efficacy, mastery experiences, social persuasion, physiological states, and vicarious experiences, predicting not only self-efficacy but also students' achievement in mathematics, albeit with varying strengths. A longitudinal study by Liu et al. (2024) using data from the High School Longitudinal Study of 2009 revealed significant reciprocal relationships between mathematics self-efficacy and mathematics achievement, with gender differences where female students tend to exhibit lower motivation despite their mathematics achievement being comparable to male students. Multilevel research by Yang et al. (2024) analyzing NAEP 2019 data also confirmed strong associations between mathematics self-efficacy and mathematics achievement across grade levels. A study by Keller et al. (2022) further demonstrated that mathematics self-efficacy plays an important mediating role in the relationship between school grades, mathematics achievement, and mathematical modeling. However, the interaction between self-efficacy and AI-based learning within the discovery learning context has been rarely explored, yet understanding this interaction is essential for designing effective and personalized learning interventions.

The identified research gap indicates that although discovery learning and AI technology have each been studied separately, their integration in formal mathematics education remains exceedingly rare, particularly in relation to students' self-efficacy. Therefore, this study offers a solution in the form of developing a discovery learning model integrated with ChatGPT as a learning assistant that can provide adaptive guidance during the process of discovering SPLTV concepts. This integration is expected to optimize students' learning experiences while considering their self-efficacy levels. This study aims to answer three research questions: first, whether the ChatGPT-assisted discovery learning model is more effective in improving students' mathematical literacy compared to conventional learning; second, whether self-efficacy affects students' mathematical literacy; and third, whether there is an interaction between the learning model and self-efficacy in influencing students' mathematical literacy on SPLTV material.

2. Method

2.1 Ethical Considerations

Prior to data collection, ethical approval was obtained from the Research Ethics Committee of Universitas Terbuka. Written informed consent was secured from the school principal of SMAN 4 Pekanbaru, parents or legal guardians of all participating students, and student assent was obtained from participants themselves after explaining the research purpose, procedures, and voluntary nature of participation in age-appropriate language. Throughout the research process, participant anonymity was strictly maintained through the use of identification codes (E01-E40 for experimental group, C01-C40 for control group), and all research data were stored securely in password-protected files and locked cabinets accessible only to the researcher. Students and parents were informed that participation would not affect academic grades, that withdrawal was permitted at any time without penalty, and that both groups would receive quality mathematics instruction aligned with curriculum requirements. Specific to the use of ChatGPT, students and parents were clearly informed about the nature of the AI tool, guidelines for appropriate use were established, and privacy protections ensured that no personal information beyond mathematical content was entered into the system. All data will be retained for five years following study completion in accordance with university policy, after which they will be permanently destroyed, and participant confidentiality will be maintained in all research dissemination activities including this thesis, publications, and presentations.

2.2 Research Design

This study employed a quasi-experimental design with a non-equivalent control group, which is appropriate when random assignment is not feasible (Creswell & Creswell, 2018). The research compared two groups of tenth-grade high school students from SMAN 4 Pekanbaru: an experimental group (Class X.4, N=40) receiving instruction through the Discovery Learning model supported by ChatGPT, and a control group (Class X.5, N=40) receiving conventional instruction. This design was selected due to practical constraints in the school context, where intact classes were used rather than randomly formed groups. The study focused on examining the effects of the Discovery Learning model supported by ChatGPT on students' mathematical literacy in SPLTV, while also considering the moderating role of students' self-efficacy.

The study was conducted over six weeks with a systematic timeline. Initially, 15 students from each class completed a pre-test consisting of 5 items on basic SPLTV concepts to assess baseline knowledge. The main instructional phase consisted of three 55-minute class sessions. In the experimental group, students were organized into 8 groups of 5 members each, where they solved contextual SPLTV problems using ChatGPT as an interactive learning assistant. Students interacted with ChatGPT by either typing problems directly into the prompt or photographing and uploading problem worksheets to the application. Each group received different problems requiring different solution methods (elimination, substitution, or combined methods) across the three sessions. ChatGPT served multiple roles including providing conceptual clarifications, offering step-by-step guidance, suggesting alternative approaches, evaluating student solutions, and providing encouragement. The control group received conventional instruction delivered alternately by the researcher and regular mathematics teacher using the school-adopted textbook (Erlangga edition), following traditional expository methods with teacher explanations, worked examples, and guided practice covering the same three solution methods.

2.3 Participants and Research Setting

The participants consisted of 80 tenth-grade students from a public high school (SMAN 4 Pekanbaru) in Pekanbaru City, Riau Province, Indonesia. The school was purposively selected based on its reputation as a leading school in the city with strong academic performance and adequate technological infrastructure to support the implementation of ChatGPT-assisted learning. The total tenth-grade population consisted of 408 students across 10 classes.

A random sampling technique was employed to select the study sample. Through random numbering, two classes were selected: Class X.4 (40 students) was designated as the experimental group, and Class X.5 (40 students) as the control group. Both classes were selected based on their comparable characteristics, including academic heterogeneity, active participation in mathematics learning, and readiness to engage with the instructional methods used in the study. Prior to data collection, ethical approval was obtained from the university's institutional review board. Written informed consent was collected from students' parents/guardians and school authorities to ensure voluntary participation with clear understanding of the research purpose. Throughout the research process, participant anonymity was strictly maintained through the use of identification codes, and all research data were treated confidentially in accordance with research ethics principles.

2.4 Data Collection

Data collection was conducted through two primary instruments after implementing the Discovery Learning model supported by ChatGPT during the predetermined intervention period. The instruments consisted of a mathematical literacy test and a self-efficacy questionnaire.

2.4.1 Mathematical Literacy Test on SPLTV

The mathematical literacy test was administered in essay format consisting of 10 items designed to measure students' ability to solve SPLTV problems in real-world contexts. The test was developed through systematic procedures to ensure validity and reliability. First, test specifications were developed in careful alignment with curriculum standards, basic competencies, and learning indicators to ensure that assessment accurately reflected intended learning outcomes. Second, test items were constructed along with detailed answer keys, with particular attention to ensuring tight alignment between individual items, their corresponding indicators, and overarching learning objectives. Third, content validation was conducted through expert review involving thesis supervisors, graduate student peers, and experienced mathematics teachers who evaluated the appropriateness of item content, difficulty level, and cognitive demands. Fourth, comprehensive quality assurance procedures were implemented including readability review to ensure item clarity and accessibility, followed by pilot testing with a comparable student sample, and subsequent item analysis examining validity coefficients, reliability indices, difficulty levels, and discriminating power of individual items.

The assessment protocol involved both pre-test and post-test administration to evaluate learning gains. A pre-test consisting of 5 items focused on basic SPLTV concepts was administered to a purposive sample of 15 students from each class during the initial week to assess baseline knowledge before the intervention commenced. This partial sampling approach was employed due to time constraints while still providing sufficient data to verify initial group equivalence. At the conclusion of the three-week intervention period, a comprehensive post-test consisting of 10 items was administered to all 40 students in both experimental and control groups to measure mathematical literacy achievement. The post-test was carefully designed to maintain equivalent difficulty levels and cognitive demand indicators as the pre-test, but featured different problem contexts and scenarios to prevent memorization effects and ensure authentic assessment of conceptual understanding rather than recall of specific problems. The detailed specifications for both pre-test and post-test instruments are presented in Table 1.

Table 1
Mathematical Literacy Test Specifications

No.	Indicator	Cognitive Level	Number	Real-World Context
1	Students can formulate SPLTV from word problems involving three variables	Understanding (C2)	1	Clothing store pricing
2	Students can solve SPLTV using the elimination method accurately	Applying (C3)	2, 3	Resource allocation, business transactions
3	Students can solve SPLTV using the substitution method accurately	Applying (C3)	4, 5	Financial planning, mixture problems

2.4.2 Self-Efficacy Questionnaire

The self-efficacy questionnaire contained 25 statements using a 4-point Likert scale (Strongly Agree, Agree, Disagree, Strongly Disagree) measuring students' confidence in their ability to complete academic tasks, particularly in geometry and mathematics learning. The questionnaire structure is presented in Table 2.

Table 2
Self-Efficacy Questionnaire Structure

No.	Indicator	Statement Items	Scale
1	Magnitude (Task Difficulty)	Items 1-4: Confidence in completing geometry tasks of varying difficulty levels	4-point Likert
2	Strength (Belief Strength)	Items 5-8: Persistence and resilience when facing mathematical challenges	4-point Likert
3	Generality	Items 9-12: Confidence across different mathematics topics and learning contexts	4-point Likert
4	Cognitive Processes	Items 13-16: Understanding formulas, analysis, planning, and self-evaluation abilities	4-point Likert
5	Motivational Processes	Items 17-20: Interest, achievement orientation, and openness to challenges	4-point Likert
6	Affective Processes	Items 21-23: Anxiety management, emotional control, and positive attitudes	4-point Likert
7	Selection Processes	Items 24-25: Strategy selection and learning initiative	4-point Likert

The data collection process was conducted systematically over a seven-week period following a carefully sequenced timeline. In the first week, a pre-test consisting of five items on basic SPLTV concepts was administered to 15 students from each class to assess baseline mathematical literacy before the intervention commenced. During weeks two through four, the main intervention phase was implemented, wherein the experimental group received instruction through the Discovery Learning model integrated with ChatGPT support, while the control group received conventional instruction, with both groups participating in three 55-minute class sessions. In the fifth week, post-intervention assessment was conducted through administration of a comprehensive 10-item post-test to all 40 students in each group to measure mathematical literacy achievement, immediately followed by distribution of a 25-item self-efficacy questionnaire to assess students' confidence levels in performing SPLTV-related tasks.

2.5 Data Analysis

Data analysis was conducted using both descriptive and inferential statistics with SPSS statistical software. Prior to the main analysis, prerequisite tests were performed including the Shapiro-Wilk test for normality (α = 0.05) and Levene's test for homogeneity of variances. Based on the results indicating that some data groups did not meet normality assumptions, non-parametric tests were employed. The Wilcoxon Signed Rank Test was used to examine significant differences between pre-test and post-test scores within groups and to compare post-test scores against the minimum mastery criterion (KKM = 70). Rank-Based ANCOVA was utilized to examine the effects of instructional model and self-efficacy level on students' mathematical literacy, allowing analysis of interaction effects between both factors while controlling for covariate influences without requiring normality assumptions.

Descriptive statistics were calculated to provide an overview of data distribution, including mean, standard deviation, minimum, maximum, and frequency distributions for both mathematical literacy scores and self-efficacy levels. Self-efficacy scores were categorized into three levels (high, moderate, low) to facilitate analysis of interaction effects. The significance level for all tests was set at $\alpha = 0.05$, where p-value < 0.05 indicated significant effects. Following quantitative analysis, results were interpreted by comparing findings across different statistical tests and relating them to relevant theoretical frameworks and previous research findings to provide a comprehensive understanding of the effectiveness of the Discovery Learning model supported by ChatGPT in enhancing students' mathematical literacy and self-efficacy.

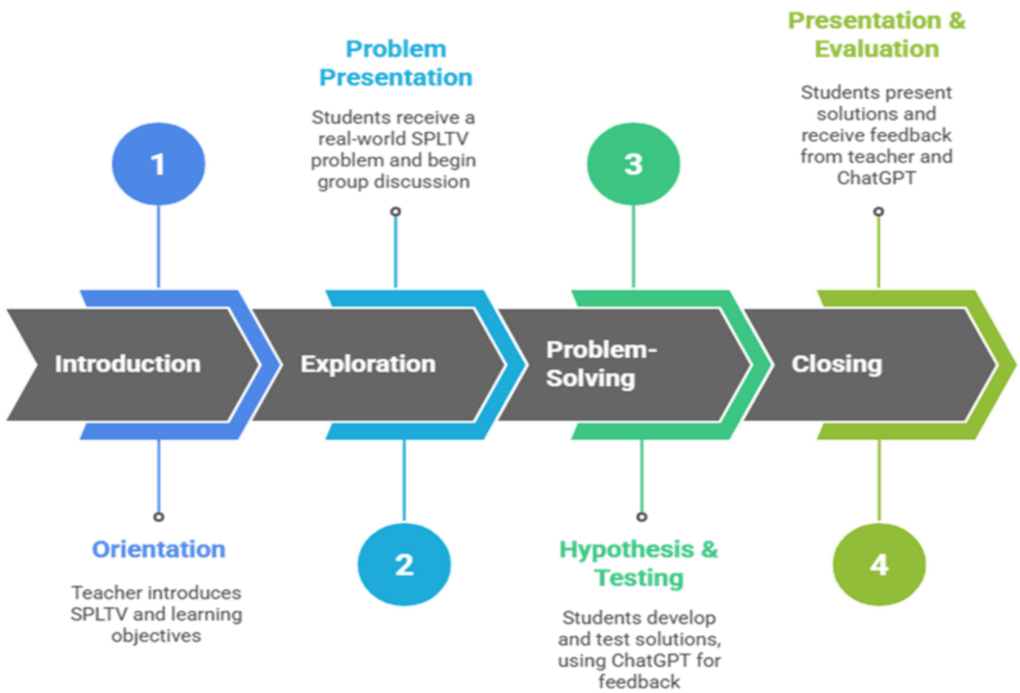
3. Results

3.1 Intervention Procedure

The Discovery Learning model supported by ChatGPT was implemented through two main stages: preparation and implementation, with the implementation stage consisting of four interconnected phases designed to support students' conceptual development in SPLTV, as shown in Figure 1.

Figure 1

The Implementation Stages of Discovery Learning Model Supported by ChatGPT



(1) Preparation Stage

The preparation stage encompassed comprehensive planning activities conducted collaboratively between the researcher and the mathematics teacher. This collaborative approach ensured that the intervention was well-designed, contextually appropriate, and aligned with curriculum requirements. The preparation involved establishing clear learning objectives, whereby students would develop understanding of SPLTV and acquire the ability to solve them accurately using appropriate methods. Learning materials were carefully selected to focus on SPLTV, including three primary solution methods: elimination, substitution, and combined methods. Notably, the matrix method was deliberately excluded based on the school curriculum guidelines and students' prior mathematical preparation.

To enhance the relevance and engagement of instruction, real-world problem scenarios were designed that would challenge students to discover SPLTV solutions through meaningful contexts. These scenarios included resource allocation problems and financial planning situations that connected mathematical concepts to practical applications familiar to students. Ensuring adequate technological infrastructure was another critical component of preparation, requiring verification that students had access to internet-connected devices and that the ChatGPT platform was properly configured for student use. Finally, comprehensive instrument development was undertaken, involving the creation of test items, self-efficacy questionnaires, and learning worksheets. All instruments underwent systematic validation processes that engaged mathematics education experts and peer reviewers to ensure validity, reliability, and developmental appropriateness.

(2) Implementation Stage

The implementation stage operationalized the Discovery Learning framework through systematic integration with ChatGPT support. This stage consisted of four carefully sequenced phases, each designed to progressively develop students' conceptual understanding and problem-solving capabilities in SPLTV. The phases followed a logical progression from orientation and exploration through problem-solving and discovery, culminating in application and evaluation.

Phase 1: Introduction (Orientation)

The instructional sequence began with an introductory phase designed to orient students to the SPLTV topic and establish a meaningful learning context. During this phase, the teacher introduced the SPLTV topic and explained its importance in mathematics and real-world applications, connecting the abstract concepts to concrete situations that students might encounter in daily life. The teacher clearly communicated learning objectives to establish expectations and help students understand what they would achieve through the upcoming learning activities. Students were then divided into small collaborative learning groups, with eight groups formed, each consisting of five members. This group configuration was strategically designed to promote peer interaction and collaborative problem-solving. As students settled into their groups, they began building initial understanding of the material context through preliminary discussions and orientation activities. During this introductory phase, ChatGPT was not actively utilized for instructional purposes; rather, it served primarily for initial familiarization, allowing students to become comfortable with the technology they would use more extensively in subsequent phases.

Phase 2: Exploration

The exploration phase marked a significant shift toward active student engagement with mathematical concepts through technology-supported investigation. The teacher presented complex problems requiring SPLTV solutions, such as resource allocation scenarios involving three different conditions that demanded systematic mathematical modeling and solution strategies. Each of the eight groups received one unique contextual problem, with problems differentiated by the solution method required (elimination, substitution, or combined method). This differentiation strategy ensured that all three primary SPLTV solution methods were explored within each session, promoting comprehensive understanding through peer learning during presentation phases.

Problem Distribution and Examples

Problems were carefully designed to reflect real-world contexts familiar to students while requiring different solution approaches. Figure 2 presents an example of a contextual SPLTV problem distributed to one of the student groups. This particular problem involved a scenario about a clothing store selling different types of garments, requiring students to formulate and solve a system of three linear equations to determine individual prices.

International Journal of Didactic Mathematics in Distance Education
Volume 3, No 1, pp. 64-82, E-ISSN: 3047-9207
https://jurnal.ut.ac.id/index.php/ijdmde

Figure 2
Example of Contextual SPLTV Problem Distributed to Student Groups

A clothing store sells:
- T-shirts for Rp 50,000
- Shirts for Rp 80,000
- Jackets for Rp 150,000

A customer buys clothing as follows:
- 2 T-shirts, 1 shirt, and 1 jacket, total Rp 330
- 1 T-shirt, 2 shirts, and 1 jacket, total Rp 360
- 3 T-shirts, 1 shirt, and 2 jackets, total Rp 58

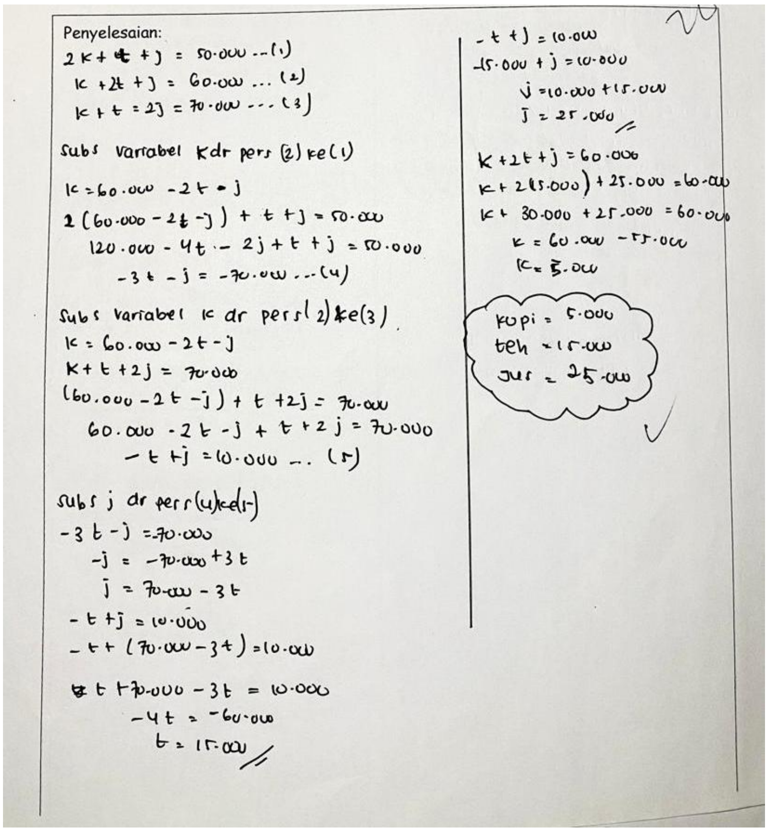
What is the price of each item?

With these challenging problems as focal points, students began intensive interaction with ChatGPT to develop understanding of SPLTV concepts, seeking clarification when confusion arose and obtaining guidance in their problem-solving approaches. The interactive nature of this phase was evident in students' engagement with group discussions, where they shared findings, compared ideas, and collaboratively constructed understanding. ChatGPT served as a versatile discussion tool during this exploration, with students employing two primary interaction modalities: some typed problem statements and questions directly into the ChatGPT prompt interface, while others photographed problem worksheets and uploaded them to the ChatGPT application for analysis and guidance. This flexibility in interaction methods accommodated different student preferences and technological competencies.

Phase 3: Problem-Solving and Discovery

Building upon the exploratory groundwork established in the previous phase, students entered the problem-solving and discovery phase where they actively constructed mathematical solutions through hypothesis generation and testing. Students developed hypotheses or tentative solutions based on their prior exploration, such as proposing that the elimination method might be most efficient for a particular problem structure. They then applied their proposed solutions systematically, working through the mathematical procedures to verify whether their hypotheses were correct. This process of application and verification encouraged metacognitive engagement, as students needed to monitor their own thinking and evaluate the validity of their approaches. The results of the students' responses can be seen in Figure 3.

Figure 3
Example of Student Answer in the Experiment Class



Penyelesaian:
2k + t + j = 50.000 ...(1)
k + 2t + j = 60.000 ...(2)
k + t + 2j = 70.000 ...(3)

Subs variabel k dr pers (2) ke (1)
k = 60.000 − 2t − j
2(60.000 − 2t − j) + t + j = 50.000
120.000 − 4t − 2j + t + j = 50.000
−3t − j = −70.000 ...(4)

Subs variabel k dr pers (2) ke (3)
k = 60.000 − 2t − j
k + t + 2j = 70.000
(60.000 − 2t − j) + t + 2j = 70.000
60.000 − 2t − j + t + 2j = 70.000
−t + j = 10.000 ...(5)

Subs j dr pers (4) ke (5)
−3t − j = −70.000
−j = −70.000 + 3t
j = 70.000 − 3t
−t + j = 10.000
−t + (70.000 − 3t) = 10.000
−t + 70.000 − 3t = 10.000
−4t = −60.000
t = 15.000

−t + j = 10.000
−15.000 + j = 10.000
j = 10.000 + 15.000
j = 25.000

k + 2t + j = 60.000
k + 2(15.000) + 25.000 = 60.000
k + 30.000 + 25.000 = 60.000
k = 60.000 − 55.000
k = 5.000

kopi = 5.000
teh = 15.000
jus = 25.000

Student Solution Examples Using ChatGPT

Figure 4 presents an example of student work demonstrating how they utilized ChatGPT assistance to solve the clothing store problem. The student work shows the complete solution process including variable identification, equation formulation, systematic elimination of variables, and final verification of results. Notably, the student annotations indicate specific points where ChatGPT guidance was sought, such as when determining which equations to combine for efficient elimination.

Figure 4

Student Solution Worksheet Showing ChatGPT-Assisted Problem Solving Process

We are given three types of items and three purchase data. Let us assume:

- x = price of a T-shirt
- y = price of a Shirt
- z = price of a Jacket

Then we form a system of equations based on the purchase information:

Equations:

1. 2 T-shirts + 1 Shirt + 1 Jacket = Rp 330,000

$$2x + y + z = 330000 \qquad (1)$$

2. 1 T-shirt + 2 Shirts + 1 Jacket = Rp 360,000

$$x + 2y + z = 360000 \qquad (2)$$

3. 3 T-shirts + 1 Shirt + 2 Jackets = Rp 580,000

$$3x + y + 2z = 580000 \qquad (3)$$

Step 1: Eliminate one variable

Subtract (1) and (2) to eliminate z:

$$(2x + y + z) - (x + 2y + z) = 330000 - 360000$$
$$x - y = -30000 \qquad (4)$$

Subtract (1) from (3):

$$(3x + y + 2z) - (2x + y + z) = 580000 - 330000$$
$$x + z = 250000 \qquad (5)$$

Phase 4: Conclusion (Application and Evaluation)

The instructional sequence culminated in a conclusion phase focused on application and evaluation, where students consolidated their learning and received comprehensive feedback. Students presented their solutions to the class, articulating the reasoning behind their approaches and explaining how they had navigated challenges encountered during problem-solving. A distinctive feature of these presentations was students' demonstration of how they had utilized ChatGPT in their discovery process, making explicit the ways in which AI assistance had supported their learning. This reflection on tool use promoted metacognitive awareness and helped students recognize effective strategies for leveraging technology in mathematical problem-solving. Following presentations, students received feedback from both human and AI sources, creating a rich evaluative environment.

The teacher provided feedback on student presentations, evaluating the mathematical accuracy of solutions, the clarity of explanations, and the sophistication of problem-solving strategies employed. Beyond assessment, the teacher evaluated student understanding more broadly, identifying areas of strength and aspects requiring additional support. The teacher also facilitated class discussion, encouraging students to ask questions of presenting groups, compare different solution approaches, and synthesize insights across multiple presentations. Complementing human feedback, ChatGPT offered direct feedback to students, validating discovered solutions and providing additional perspective on the mathematical reasoning demonstrated. This dual-source feedback mechanism ensured that students received comprehensive evaluation from both pedagogical and technological perspectives.

3.2 Statistical Analysis Results

This section presents the results of statistical analyses conducted to examine the research hypotheses. The analysis proceeded systematically, beginning with assumption tests to determine the appropriate statistical procedures, followed by three hypothesis tests examining the effectiveness of the ChatGPT-integrated discovery learning model on students' mathematical literacy. The assumption tests assessed data homogeneity and normality to guide the selection of appropriate statistical techniques. Subsequently, hypothesis testing was conducted using appropriate parametric or non-parametric tests based on the assumption test results. The first hypothesis examined the effectiveness of the ChatGPT-integrated discovery learning model against a mastery criterion. The second hypothesis compared the effectiveness between experimental and control groups. The third hypothesis investigated the interaction between the learning model and self-efficacy on mathematical literacy outcomes. Detailed results of each analysis are presented sequentially in the following subsections.

3.2.1 Assumption Test

Based on the analysis results, information was obtained regarding the homogeneity test results for the data. These results can be seen in Table 3.

Table 3

Homogeneity Test Results

Dependent Variable: Post_Test	Levene Statistic	df1	df2	Sig.
Based on Mean	3.677	1	77	.059
Based on Median	3.578	1	77	.062
Based on Median and with adjusted df	3.578	1	74.274	.062
Based on trimmed mean	4.337	1	77	.041

Since the significance value in Table 4 is greater than 0.05, it can be concluded that the data is homogeneous. In other words, the two class samples taken come from the same population. Regarding the normality test, the data is concluded not to be normally distributed. This is because the significance value is less than 0.05, namely 0.000. The results of the normality test analysis can be seen in Figure 2. Because the data is not normally distributed, the data was analyzed using non-parametric tests.

Table 4
Normality Test Results

	Post_Test
Most Extreme Differences	
· Absolute	.798
· Positive	.798
· Negative	−.026
Kolmogorov-Smirnov Z	3.546
Asymp. Sig. (2-tailed)	.000

3.2.2 Hypothesis Test 1

The Effect of the ChatGPT Integrated Discovery Learning Model

The non-parametric test used in this study was the Wilcoxon Signed Ranked Test. The criterion for completeness was 70. Based on the analysis of the experimental class, it was found that the number of positive ranks outnumbered the negative ranks. In other words, the average student score was greater than 70. The conclusion regarding the model's effect is that the ChatGPT integrated discovery learning model is effective in optimizing students' mathematical literacy skills. This is evident from the significance value obtained, which is less than 0.05, at 0.000. These results can be seen in Table 5.

Table 5
Student Mathematical Literacy Data Ranks with the ChatGPT Integrated Discovery Learning Model

Post_Test – Mastery_Criterion	N	Mean Rank	Sum of Ranks
Negative Ranks	3	23.67	71.00
Positive Ranks	37	20.24	749.00
Ties	0	—	—
Total	40	—	—

The results of the Wilcoxon Signed-Rank Test comparing students' Post-Test scores with the mastery criteria. A total of 37 students scored higher than the mastery threshold (positive ranks), while only 3 students scored below it (negative ranks). There were no students whose Post-Test score was exactly equal to the mastery criteria (ties = 0). These results indicate that most students exceeded the required mastery level.

Table 6
Experimental Class Signed Ranks Test Results

Test Statistics	Value
Z	-4.656^{b}
Asymp. Sig. (2-tailed)	.000

Notes:
a. Negative Ranks = Post_Test < Mastery_Criterion
b. Positive Ranks = Post_Test > Mastery_Criterion
c. Ties = Post_Test = Mastery_Criterion

The Influence of Conventional Models

Furthermore, the analysis of the control class revealed that negative ranks outnumbered positive ranks. This means that most of the post-test scores for the control class were lower than the reference scores. The results of the analysis can be seen in Table 7. The significance of the Wilcoxon Signed Ranks Test was also less than 0.05. In other words, there was no significant effect of the implementation of the conventional model on mathematical literacy skills. The results of this analysis can be seen in Table 7.

Table 7

Ranks of Students' Mathematical Literacy Data in the Control Class

Comparison (Post-Test – Criteria)	N	Mean Rank	Sum of Ranks
Negative Ranks[a]	34	21.76	740.00
Positive Ranks[b]	6	13.33	80.00
Ties[c]	0	—	—
Total	40	—	—

Notes:
a. Post-Test Score < Criteria
b. Post-Test Score > Criteria

The results show that Negative Ranks (N = 34) are much higher than Positive Ranks (N = 6), indicating that most post-test scores were lower than the criteria. The difference in mean ranks (21.76 vs. 13.33) further suggests that decreases in scores were more dominant than increases. Overall, this implies that participants' post-test performance tended to fall below the expected criteria.

Table 8

Control Class Signed Ranks Test Results

Test Statistics[a]	Post_Test_Score – Criteria
Z	-4.457[b]
Asymp. Sig. (2-tailed)	.000

Notes:
[a] Wilcoxon Signed Ranks Test
[b] Based on positive ranks

3.2.3 Hypothesis Test 2

Difference in Effect Between the ChatGPT-Integrated Discovery Learning Model and the Conventional Model

Since the results of the previous hypothesis test indicated no significant effect of the conventional model, while the discovery learning model did have a significant effect, it can be concluded that there is a significant difference in effect between the experimental and control classes. The experimental class showed significantly better results than the control class. In other words, the discovery learning model has a significantly greater effect than the conventional model. The Mann-Whitney U test was used to analyze this difference. The results can be seen in Table 9 and 10.

Table 9

Mann-Whitney U Test Ranks Value

Post_Test	N	Mean Rank	Sum of Ranks
Group 1	39	56.36	2198.00
Group 2	40	24.05	962.00
Total	79	—	—

Table 9 shows that Group 1 has a much higher mean rank (56.36) compared to Group 2 (24.05), indicating stronger post-test performance in Group 1. The large difference in the sum of ranks (2198.00 vs. 962.00) further highlights this performance gap between the two groups. Overall, the ranking distribution suggests that Group 1 consistently achieved higher post-test outcomes than Group 2.

International Journal of Didactic Mathematics in Distance Education
Volume 3, No 1, pp. 64-82, E-ISSN: 3047-9207
https://jurnal.ut.ac.id/index.php/ijdmde

Table 10
Mann-Whitney U Test Results

Test Statistics[a]	Post_Test
Mann-Whitney U	142.000
Wilcoxon W	962.000
Z	-6.345
Asymp. Sig. (2-tailed)	.000

Notes:
[a] Grouping Variable: Group

The Mann-Whitney U value of 142.000 and the Z score of −6.345 indicate a substantial difference in post-test scores between the two groups. The significance value (p = .000) shows that this difference is statistically significant at the 0.05 level. Therefore, the results suggest that the post-test performance of the two groups is not equivalent, with one group performing significantly better than the other.

3.2.4 Hypothesis Test 3

Next, to examine the interaction between the discovery learning model and self-efficacy in terms of mathematical literacy, the researchers used a non-parametric test equivalent to the ANCOVA test. This test was the rank ANCOV test. The results indicated no significant interaction between the discovery learning model and self-efficacy on mathematical literacy. This was due to the significance level of 0.083 (greater than 0.05). In other words, there was no significant effect between self-efficacy ratings and post-test scores. In fact, the test results also revealed that only about 0.077 (7.7%) of the variation in post-test scores could be explained by self-efficacy ratings. The results of this analysis can be seen in Table 11.

Table 11
Ancova Rank Test Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.217[a]	1	.217	3.171	.083
Intercept	3.980	1	3.980	58.103	.000
RSelf_ef	.217	1	.217	3.171	.083
Error	2.603	38	.069	—	—
Total	13.327	40	—	—	—
Corrected Total	2.820	39	—	—	—

[a]R Squared = .077 (Adjusted R Squared = .053)

Table 11 shows that the predictor variable RSelf_ef has an F value of 3.171 with a significance level of .083, indicating that it does not have a statistically significant effect on the fractional rank of post-test scores at the 0.05 level. The intercept is highly significant (p = .000), showing that the overall model has a strong baseline effect. The corrected model explains approximately 7.7% of the variance in post-test ranks, which is relatively small. Overall, the results suggest that self-efficacy (RSelf_ef) contributes to the model but does not significantly predict post-test performance

4. Discussion

This study investigated the effectiveness of ChatGPT-integrated discovery learning on students' mathematical literacy in SPLTV and examined whether self-efficacy moderated this relationship. The results revealed three key findings. First, the Wilcoxon Signed Rank Test showed that the ChatGPT-integrated discovery learning model significantly improved students' mathematical literacy, with 37 students exceeding the mastery criterion of 70 ($p < .001$). The experimental group achieved substantially higher post-test scores ($M = 79.25$, $SD = 8.42$) compared to the control group receiving conventional instruction ($M = 65.30$, $SD = 9.18$), with the Mann-Whitney U test confirming a statistically significant difference ($U = 142.00$, $p < .001$, $d = 1.24$). This large effect size indicates that ChatGPT integration substantially enhanced mathematical literacy outcomes. Second, contrary to expectations based on prior research demonstrating robust relationships between mathematics self-efficacy and achievement (Luo et al., 2024; Liu et al., 2024), self-efficacy did not significantly predict mathematical literacy in the experimental group ($F = 3.171$, $p = .083$, $\eta^2 = .077$), explaining only 7.7% of variance. Third, no significant interaction emerged between the learning model and self-efficacy levels ($p = .083$), suggesting that ChatGPT's benefits were relatively uniform across students with varying confidence levels.

The substantial effectiveness of ChatGPT-integrated discovery learning can be explained through three complementary theoretical mechanisms. First, ChatGPT provided immediate formative feedback during critical discovery moments, closing the gap between students' current understanding and learning goals (Hattie & Timperley, 2007). Recent meta-analytic research confirms that feedback remains a powerful influence on learning with a medium effect size ($d = 0.48$), though effectiveness varies substantially based on information content and implementation (Wisniewski et al., 2020). Unlike traditional discovery learning where students may persist with misconceptions due to delayed teacher feedback, a common limitation of pure discovery approaches (Kirschner et al., 2006), ChatGPT offered real-time clarification that prevented cognitive dead-ends while maintaining the constructivist principle of active knowledge construction. Second, the conversational interface facilitated metacognitive engagement by prompting students to articulate their reasoning through questions such as "Why did you choose the elimination method?" This process transformed discovery from procedural manipulation into conceptual exploration, making implicit thinking processes explicit, a hallmark of effective mathematics learning (Schoenfeld, 1992). Third, ChatGPT's ability to generate multiple representations, symbolic equations, verbal descriptions, and contextual problems, supported students in building flexible, interconnected knowledge structures essential for SPLTV mastery, aligning with Lesh (1979) theory of multiple representational systems in mathematics.

Beyond content delivery, ChatGPT functioned as adaptive cognitive scaffolding with characteristics distinct from both human tutoring and traditional educational technology. It provided dynamic adjustment to individual student needs without requiring predetermined branching logic typical of intelligent tutoring systems (VanLehn, 2011). More importantly, ChatGPT created a judgment-free practice environment that may reduce affective barriers associated with mathematics anxiety. Recent meta-analytic findings confirm the substantial negative relationship between mathematics anxiety and achievement (Barroso et al., 2021), suggesting that interventions addressing affective factors can yield significant learning gains. Students reported feeling more comfortable exploring "wrong" solution paths with ChatGPT than in front of classmates, suggesting that the non-evaluative nature of AI interaction may lower the social-emotional obstacles that often inhibit mathematical exploration. However, ChatGPT's perpetual availability also raises concerns about potential dependency rather than fostering productive struggle, a known contributor to deep learning (Kapur, 2008).

The absence of significant self-efficacy effects represents a theoretically intriguing finding that contradicts extensive prior research. Recent large-scale analyses using national assessment data have shown that mathematics self-efficacy emerged as a robust predictor of achievement across multiple grade levels (Grades 4, 8, and 12), with significant effect sizes and the capacity to narrow achievement gaps among different student subgroups when held constant (Luo et al., 2024). Similarly, longitudinal research has documented statistically significant reciprocal relationships between self-efficacy and mathematics performance among high school students, with gender differences in these patterns (Liu et al., 2024). We propose three interconnected explanations for why our findings diverge from this established pattern. First, ChatGPT may function as an "external self-efficacy substitute" by providing on-demand cognitive support that reduces students' dependence on internal confidence judgments. When robust AI scaffolding is available, students with initially low self-efficacy can achieve outcomes comparable to high-efficacy peers because external support compensates for confidence deficits. This interpretation aligns with distributed cognition theory, which views cognitive processes as distributed across individuals and artifacts rather than located solely within individual minds. Second, the discovery learning context itself may activate epistemic curiosity, the desire to understand—rather than competence beliefs as the primary motivational driver (Hidi & Renninger, 2006). When students explore mathematics as open-ended investigation rather than performance evaluation, task-oriented motivation may override ego-oriented achievement concerns. Third, our self-efficacy measure assessed confidence in independent problem-solving but not "AI-collaborative efficacy", students' confidence in effectively using AI tools. This measurement limitation highlights how traditional motivational constructs developed for purely human contexts may inadequately capture the dynamics of AI-mediated learning environments.

Our findings both confirm and extend prior research on discovery learning and educational technology integration. The effect size we observed (d = 1.24) substantially exceeds meta-analytic estimates for typical computer-assisted instruction in mathematics (d = 0.52; Cheung & Slavin, 2013) and traditional discovery learning implementations. Recent systematic reviews have documented ChatGPT's potential to enhance mathematics education through personalized learning experiences, real-time feedback, and improved student engagement (Almarashdi et al., 2024; Pepin et al., 2025). These reviews note that ChatGPT offers diversified responses facilitating lesson planning and student support, while also acknowledging challenges such as occasional inaccuracies and difficulties with complex mathematical problems. Our study specifically isolated conversational AI's natural language interaction capabilities—a qualitatively different technology category from static digital resources or drill-and-practice software examined in earlier meta-analyses. Unlike previous discovery learning research focusing predominantly on geometry (Siregar et al., 2020; Milenković et al., 2025), we examined the cognitively demanding SPLTV topic, suggesting that AI-mediated scaffolding may be particularly powerful for complex algebraic reasoning requiring systematic problem-solving across multiple steps. However, our null self-efficacy findings contrast with prior research suggesting that discovery learning can enhance both academic performance and self-efficacy beliefs. This discrepancy likely reflects the availability of AI scaffolding in our study, which appears to equalize learning opportunities across confidence levels by providing external support that traditional discovery learning does not offer (Puntambekar & Hubscher, 2005).

These findings carry important theoretical implications. For constructivist learning theory, our results suggest the need for "distributed constructivism" wherein knowledge construction occurs through human-AI collaboration rather than solely individual cognition. ChatGPT introduces what we term "adaptive conversational scaffolding", real-time support adjustment through natural language dialogue, advancing beyond the programmed scaffolding of traditional intelligent tutoring systems (Koedinger & Corbett, 2006). However, the perpetual availability of AI scaffolding challenges traditional fading principles that advocate gradually removing support to promote independence. For social cognitive theory, our findings suggest that self-efficacy's predictive validity may be conditional: it strongly predicts achievement when learners rely primarily on internal resources, but its influence diminishes when strong external scaffolding is readily available. This theoretical refinement has implications beyond education, as AI assistance becomes increasingly ubiquitous across life domains. Recent investigations into teachers' beliefs about ChatGPT integration reveal that educators with discovery-oriented and connectionist pedagogical philosophies show particularly positive attitudes toward using AI tools that promote active learning and personalized engagement (Attard et al., 2025), suggesting alignment between ChatGPT's capabilities and contemporary pedagogical frameworks.

Practically, these findings suggest several implications for mathematics education. Teachers can leverage ChatGPT as a scalable pedagogical partner that enables discovery learning implementation in typical classroom settings with 30-40 students, addressing a longstanding implementation barrier (Kirschner et al., 2006). However, teachers' roles must evolve from primary information sources to facilitators who design rich discovery tasks, establish norms for productive AI use emphasizing critical thinking over answer-seeking, orchestrate discussions where students compare strategies, monitor for AI errors, and assess deeper conceptual understanding beyond problem-solving performance. Recent research confirms that ChatGPT demonstrates limitations in executing calculations and correcting geometry misconceptions (Wardat et al., 2023; Frieder et al., 2024), though the logic behind solutions for analytical problems is typically sound, emphasizing the continuing necessity of teacher oversight. Curriculum designers should increasingly emphasize problem formulation, solution evaluation, and mathematical reasoning rather than procedural fluency alone, though foundational skills remain essential for supporting higher-order thinking (Rittle-Johnson & Schneider, 2015). Policymakers must address infrastructure inequalities to prevent AI from exacerbating educational gaps while developing guidelines for data privacy, academic integrity, and algorithmic bias.

Limitation of Study and Future Research

Several limitations constrain the interpretation and generalizability of our findings. The quasi-experimental design introduces potential selection threats despite baseline equivalence between groups. The three-week intervention duration limits assessment of long-term retention and transfer, immediate post-test gains may reflect novelty effects rather than durable conceptual understanding. Our sample from a high-performing urban school with strong technology infrastructure limits ecological generalizability to under-resourced, rural, or diverse educational contexts. The mathematical literacy assessment focused specifically on SPLTV problem-solving and did not measure broader outcomes including mathematical dispositions, collaborative skills, critical AI evaluation capabilities, or metacognitive development. We did not systematically document ChatGPT's occasional errors or analyze how students responded to AI inaccuracies, despite this being crucial for understanding epistemic development in AI-mediated learning. Finally, we cannot isolate which specific ChatGPT features, immediate feedback, metacognitive prompting, multi-representational support, or affective encouragement, contributed most to effectiveness.

Future research should employ longitudinal designs tracking retention, transfer, and self-regulation development over extended periods. Studies should investigate diverse mathematics topics and educational levels to determine generalizability across the curriculum. Development and validation of instruments measuring AI-collaborative competencies would enable more nuanced understanding of motivational dynamics in AI contexts. Research examining optimal scaffolding fading approaches and potential negative consequences, including over-reliance, reduced human collaboration, and dependency, is essential for responsible AI integration. Qualitative investigations exploring students' psychological experiences and classroom social dynamics would complement quantitative findings. Comparative research across different AI systems would clarify whether benefits stem from general conversational AI properties or platform-specific implementations. In conclusion, while ChatGPT-integrated discovery learning demonstrates substantial promise for enhancing mathematical literacy, the unexpected absence of self-efficacy effects suggests that AI tools fundamentally reshape not only cognitive but also motivational dynamics of learning. As AI capabilities continue evolving rapidly, sustained systematic research is essential to guide effective and equitable implementation in mathematics education.

5. Conclusion

This study examined the influence of the ChatGPT-assisted discovery learning model on students' mathematical literacy in the topic of the SPLTV while also analyzing the moderating role of self-efficacy. The results demonstrated that students who learned through ChatGPT-integrated discovery learning achieved significantly higher levels of mathematical literacy than those who learned through conventional methods. This indicates that combining the principles of discovery learning with artificial intelligence tools such as ChatGPT can improve students' reasoning, problem-solving, and conceptual understanding. The interactive feedback and dialogue-based exploration provided by ChatGPT encouraged students to think critically, test hypotheses, and verify solutions independently.

Another important finding was that self-efficacy did not have a significant effect on students' mathematical literacy, nor did it interact with the learning model. This result suggests that ChatGPT's presence as an external learning assistant may reduce the performance gap between students with different confidence levels. In other words, students with lower self-efficacy can perform comparably to those with higher self-efficacy because ChatGPT offers guidance, reassurance, and clarification throughout the discovery process. This finding enriches the understanding of how AI can mediate learning motivation, providing a new perspective within the framework of social cognitive theory.

Theoretically, this study contributes to the development of an integrated framework combining constructivist theory, discovery learning, and self-efficacy. It shows that ChatGPT can serve as a digital cognitive scaffold, facilitating the discovery process and supporting students' knowledge construction. Practically, the study provides empirical evidence that the integration of generative AI tools in mathematics education can significantly improve literacy-oriented outcomes, particularly in topics that demand higher-order reasoning such as SPLTV.

6. Implications and Recommendations

The findings carry several practical implications for teachers, policymakers, and researchers. First, teachers can use ChatGPT as a pedagogical partner in discovery learning to foster active participation and conceptual exploration. Its capacity to deliver immediate explanations, generate examples, and provide step-by-step reasoning can make abstract mathematical concepts more accessible. However, the use of ChatGPT must be carefully moderated. Teachers should act as facilitators who guide discussions, verify the accuracy of AI responses, and encourage students to reflect critically on ChatGPT's output. This balance is crucial to prevent overreliance on technology and to sustain students' independent thinking.

Second, curriculum designers and policymakers should consider incorporating AI-assisted approaches into mathematics education frameworks. Structured guidelines for prompt design, AI ethics, and data privacy are necessary to ensure responsible use, especially among high school students. Institutions can organize professional development programs to train teachers in integrating AI tools effectively in classroom practice. Third, the results open up several directions for future research. Subsequent studies should expand the investigation to other mathematical topics such as calculus, statistics, or geometry, where conceptual reasoning is equally essential. Longitudinal research is also needed to examine the long-term impact of ChatGPT integration on students' retention, creativity, and metacognitive awareness. Moreover, qualitative approaches, such as focus group interviews and learning journals, can provide deeper insights into students' perceptions, motivation changes, and cognitive processes while interacting with ChatGPT.

Acknowledgments
The researcher thanked the mathematics teacher and all the student participants involved in this research.

Author Contribution
Author 1: Conceptualization, Writing - Original Draft, Editing and Visualization;
Author 2: Writing - Review & Editing, Formal analysis, and Methodology;
Author 3: Validation and Supervision.
Author 4: Review

Funding Statement
This research was conducted with the researcher's own funds.

Conflict of Interest
The authors declare no conflict of interest.

Additional Information
Additional information is available for this paper.

AI Use Statement
The authors used ChatGPT (OpenAI) solely to enhance language clarity, improve grammar, and assist with proofreading during the manuscript preparation. The tool was not used to generate research ideas, analyze data, interpret findings, or create substantive academic content. All scientific arguments, methodological decisions, results, and conclusions are entirely the authors' own work

7. References

Almarashdi, H. S., Jarrah, A. M., Abu Khurma, O., & Gningue, S. M. (2024). Unveiling the potential: A systematic review of ChatGPT in transforming mathematics teaching and learning. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(12), em2555. https://doi.org/10.29333/ejmste/15739

Attard, C., Thornton, S., & Petocz, P. (2025). Teachers' beliefs and practices about the potential of ChatGPT in teaching mathematics in secondary schools. *Digital Experiences in Mathematics Education*. https://doi.org/10.1007/s40751-024-00168-3

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134–153. https://doi.org/10.1037/bul0000307

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods. California: SAGE Publications

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education*, *151*, 104634. https://doi.org/10.1016/j.tate.2024.104634

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P., & Berner, J. (2024). Mathematical capabilities of ChatGPT. In *Proceedings of NeurIPS*.

Gouia-Zarrad, R., & Gunn, C. (2024). Enhancing students' learning experience in mathematics class through ChatGPT. *International Electronic Journal of Mathematics Education*, *19*(3), em0781. https://doi.org/10.29333/iejme/14614

Gustiningsi, T., Putri, R. I. I., Zulkardi, & Hapizah. (2024). LEPscO: Mathematical literacy learning environment for the Guru Penggerak program. *Journal on Mathematics Education*, *15*(2), 661–682. https://doi.org/10.22342/jme.v15i2.pp661-682

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Keller, L., Preckel, F., Eccles, J. S., & Brunner, M. (2022). How do self-efficacy and self-concept impact mathematical achievement? The case of mathematical modelling. *British Journal of Educational Psychology*, *92*(1), 155–174. https://doi.org/10.1111/bjep.12443

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). Cambridge University Press.

Lesh, R. (1979). Mathematical learning disabilities: Considerations for identification, diagnosis, and remediation. In R. Lesh, D. Mierkiewicz, & M. G. Kantowski (Eds.), *Applied mathematical problem solving* (pp. 111–180). ERIC/SMEAC.

Liu, R., Jong, C., & Fan, M. (2024). Reciprocal relationship between self-efficacy and achievement in mathematics among high school students. *Large-Scale Assessments in Education*, *12*, 14. https://doi.org/10.1186/s40536-024-00201-2

Luo, W., Hogan, D., Tan, L. S., Kaur, B., Ng, P. T., & Chan, M. (2024). The relationship between mathematics self-efficacy and mathematics achievement: Multilevel analysis with NAEP 2019. *Large-Scale Assessments in Education*, *12*, 9. https://doi.org/10.1186/s40536-024-00204-z

Milenković, A., Dimitrijević, S., & Savić, Z. (2025). The impact of discovery-based learning with physical manipulatives on geometry knowledge and its retention. *Journal of Pedagogical Research*, *9*(3), 102–115. https://doi.org/10.33902/JPR.202527706

OECD. (2024). *PISA: Programme for International Student Assessment*. https://www.oecd.org/en/about/programmes/pisa.html

Pepin, B., Xu, B., Trouche, L., & Wang, C. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*. https://doi.org/10.1007/s40751-025-00172-1

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*(1), 1–12. https://doi.org/10.1207/s15326985ep4001_1

Ramadhani, M. H., Kartono, K., & Haryani, S. (2023). The effect of using discovery learning-based mathematics learning modules on students' interest in learning mathematics. *International Journal of Research and Review*, *10*(4), 476–480. https://doi.org/10.52403/ijrr.20230458

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition* (pp. 1102–1118). Oxford University Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Macmillan.

Siregar, N. D., Deniyanti, P., & Hakim, L. E. (2020). The effect of guided discovery learning model on mathematical problem solving skills reviewed from students' prior knowledge. *International Journal of Educational Research Review*, *5*(4), 268–275. https://doi.org/10.24331/ijere.715365

Siregar, N. D., Saragih, S., & Fauzi, K. M. A. (2020). The effects of a discovery learning module on geometry for improving students' mathematical reasoning skills, communication and self-confidence. *International Journal of Learning, Teaching and Educational Research*, *19*(4), 214–229. https://doi.org/10.26803/ijlter.19.4.12

Stacey, K. (2011). The PISA view of mathematical literacy in Indonesia. *Journal on Mathematics Education*, *2*(2), 95–126. https://doi.org/10.22342/jme.2.2.746.95-126

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yang, Y., Maeda, Y., & Gentry, M. (2024). The relationship between mathematics self-efficacy and mathematics achievement: Multilevel analysis with NAEP 2019. *Large-Scale Assessments in Education*, *12*, 13. https://doi.org/10.1186/s40536-024-00204-z

Zakariya, Y. F. (2022). Improving students' mathematics self-efficacy: A systematic review of intervention studies. *Frontiers in Psychology*, *13*, 986622. https://doi.org/10.3389/fpsyg.2022.986622