Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

April 2026. Vol. 14, No. 2
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 504-520

Designing an Artificial Intelligence-Supported Research Project-Based Inquiry Model to Improve Scientific Reasoning in Academically Underprepared Students

[1]*Asriyadin, [1]Adi Apriadi Adiansha, [2]Khairil Anwar, [3]Syarifuddin
[1]STKIP Taman Siswa Bima, Bima, Indonesia
[2]Universitas Muhammadiyah Mataram, Mataram, Indonesia
[3]Universitas Muhammadiyah Bima, Bima, Indonesia
*Corresponding Author e-mail: asriyadin@gmail.com

Received: Januari 2026; Revised: February 2026; Published: April 2026

Abstract

This study aimed to (1) design and validate an artificial intelligence-assisted Research Project-Based Inquiry (RPBI) learning model and (2) examine its effectiveness in improving the scientific reasoning of academically underprepared students. A quantitative approach was employed using a non-equivalent control group quasi-experimental design involving 60 students assigned proportionally to experimental and control groups. The intervention was implemented over eight instructional sessions in a foundational science course. The AI component functioned to generate adaptive scaffolding questions, provide hypothesis-checking prompts, and deliver rubric-based formative feedback during project development. Scientific reasoning was measured using a test constructed around five indicators: Variable Identification, Hypothesis Formulation, Data Interpretation, Conclusion Drawing, and Procedure Evaluation. Rasch analysis indicated that the instrument demonstrated satisfactory measurement quality (Person Reliability = 0.82; Item Reliability = 0.91), with appropriate item-person alignment. The validity of the learning model itself was established through expert validation using a structured rubric, pilot implementation feedback, and monitoring of instructional fidelity. Independent sample t-tests revealed statistically significant differences between the experimental and control groups across all indicators ($p < 0.001$). Exploratory relationship analysis further indicated that Hypothesis Formulation showed the strongest association with other reasoning components, suggesting a hierarchical structure within students' scientific reasoning processes. These findings indicate that the AI-assisted RPBI model is pedagogically feasible and effective in enhancing scientific reasoning, particularly among students with initial academic gaps.

Keywords: Research project; Inquiry; Artificial intelligence; Scientific reasoning; Students lack academic readiness

How to Cite: Asriyadin, A., Adiansha, A. A., Anwar, K., & Syarifuddin, S. (2026). Designing an Artificial Intelligence-Supported Research Project-Based Inquiry Model to Improve Scientific Reasoning in Academically Underprepared Students. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *14*(2), 504–520. https://doi.org/10.33394/j-ps.v14i2.19159

https://doi.org/10.33394/j-ps.v14i2.19159

INTRODUCTION

The background of this study is based on the urgency of the increasing academic readiness gap among higher education students, particularly in developing countries. In this context, scientific reasoning skills are a key element in science education because they contribute significantly to academic success and readiness to face the challenges of the 21st century (OECD, 2023a, 2023b). Data from the World Bank indicates that approximately 43% of first-year university students in Southeast Asia struggle to develop systematic scientific thinking skills (Nielsen, 2007). is exacerbated by the lack of learning models capable of bridging the academic readiness gap through adaptive, contextual, and innovative approaches.

This research problem is reinforced by various empirical reports highlighting low scientific literacy among students, particularly among those considered academically unprepared. A study by Low et al. states that more than 60% of first-year students failed to demonstrate scientific reasoning in project-based assignments (Kah Choon Low et al., 2024). Diagnostic assessment results from two state universities in Indonesia indicate that students' scientific reasoning abilities are only in the "moderate to low" category based on indicators of critical thinking and evidence-based conclusion drawing (Lukito & Madzkiyah, 2025; Sahidun et al., 2023; Wang, 2022). This highlights the need for innovation in learning strategies that can boost these scientific abilities through planned interventions.

Preliminary observations conducted during the odd semester of the 2023/2024 academic year on 82 students in the science education study program showed that 73% of students were unable to develop a scientific framework in simple research projects. The observation instruments used referred to indicators of scientific reasoning, such as the ability to identify variables, formulate hypotheses, and interpret data (Abate et al., 2020; Maghfiroh & Shofiyah, 2023). These findings indicate that underprepared students lack adequate epistemological foundations, necessitating a learning design that is not only project-based but also integrates scientific inquiry in a gradual and adaptive manner.

As a solution to this problem, a learning approach that integrates the Research Project-Based Learning (RPBL) model with the Level of Inquiry approach was chosen as the basic framework for model development. RPBL has been proven effective in improving scientific and collaborative skills (Arantes do Amaral & Lino dos Santos, 2018; Yanti et al., 2019), while the Level of Inquiry approach allows learning to be customized according to students' cognitive readiness (Arifin et al., 2025; Gerhátová et al., 2021; Wiseman et al., 2020). However, these two approaches have not yet been optimized with the support of artificial intelligence (AI) technology, which has the potential to enhance learning interactions and personalized scaffolding. Therefore, the integration of RPBL, Level of Inquiry, and AI forms a new conceptual framework that is considered strategic for implementation in the context of academically underprepared students.

In the state of the art review, most studies combining artificial intelligence technology in learning still focus on general learning personalization (Chen, 2024; Wang et al., 2023) and have not explicitly linked AI as a pedagogical facilitator in the context of RPBL and scientific inquiry. Furthermore, existing RPBL models are still generic and have not been designed for students with low academic readiness (Agustina et al., 2023; Jia et al., 2023; Naviri et al., 2021). A study by Halaweh states that AI technology support for metacognitive and project-based learning is still very limited (Halaweh, 2023). Therefore, a completely new and synthetic learning model design is needed to address these challenges

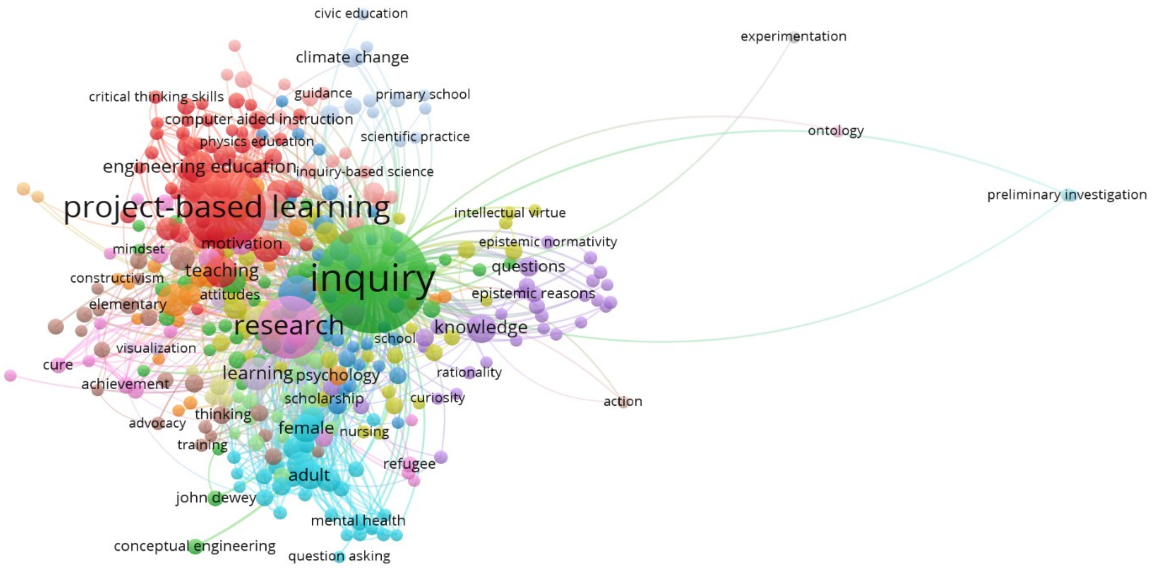

Figure 1. Bibliometric Map Visualization of Conceptual Relationships between Inquiry, Project-Based Learning, and Research

The novelty of this study is grounded in a systematic bibliometric analysis conducted to identify gaps in the integration of research-based learning, inquiry approaches, project-based learning, artificial intelligence, and academically underprepared students. The analysis was performed using the Scopus database, covering publications from 2015 to 2024. The search string combined the following keywords: (“project-based learning” OR “research-based learning”) AND (“inquiry learning”) AND (“artificial intelligence” OR “AI in education”) AND (“underprepared students” OR “academically at-risk students”). Only peer-reviewed journal articles in the fields of education and educational technology were included. After applying inclusion and exclusion criteria, records were exported and analyzed using VOSviewer with a minimum occurrence threshold of five keywords per term. Co-occurrence analysis and clustering were conducted using the default association strength normalization. Interpretation of clusters was based on keyword proximity, link strength, and thematic grouping. The resulting visualization indicates that project-based learning, inquiry learning, and research-based approaches form partially overlapping clusters; however, no dominant cluster systematically integrates these pedagogical approaches with artificial intelligence support targeted specifically at academically underprepared students. The AI-related cluster appears conceptually connected to adaptive learning and analytics, but remains weakly linked to inquiry-driven project research models for vulnerable learner populations. Therefore, the proposed AI-assisted Research Project-Based Inquiry model addresses an empirically identified conceptual gap rather than relying solely on visual impression, providing methodological justification for the study’s claimed novelty.

This study addresses this gap by developing and testing an AI-assisted Research Project-Based Inquiry model specifically designed for academically underprepared students, operationally defined in this study as first-year undergraduates who obtained a readiness score of ≤55 on a standardized academic readiness scale (0–100) administered prior to the intervention. Within the context of this research, “AI-based” does not merely refer to the presence of digital tools, but to the integration of artificial intelligence as an adaptive scaffolding mechanism grounded in scaffolding theory and cognitive apprenticeship. The AI system functions as a text-based facilitator that generates Socratic prompts, supports hypothesis refinement, provides rubric-aligned formative feedback on students’ research drafts, and assists metacognitive monitoring through reflective questioning. In addition, an instructor dashboard aggregates students’ response patterns to support formative analytics and instructional adjustment. These mechanisms are embedded within structured digital worksheets connected to an AI chatbot, ensuring that AI operates as a cognitive scaffold rather than as a content-delivery substitute. By systematically integrating research projects, inquiry processes, and AI-driven adaptive scaffolding for learners with low academic readiness, this model extends beyond the fragmented treatment of these themes in existing literature. The bibliometric mapping indicates that artificial intelligence, epistemic normativity, and underprepared learners remain weakly connected to the core project-based learning cluster, suggesting the absence of a coherent pedagogical framework that unifies these domains. Therefore, this study contributes not only empirical evidence regarding intervention effectiveness but also a theoretically grounded integration of inquiry pedagogy and intelligent educational technology targeted at a clearly defined vulnerable learner population.

Based on the above description, the research problems can be formulated as follows: (1) How to design an AI-supported Research Project-Based Inquiry learning model to improve the scientific reasoning of academically unprepared students? (2) How effective is the model in improving aspects of students' scientific reasoning, such as hypothesis formulation, data interpretation, and conclusion drawing? The objectives of this study are: (1) to design and validate a research project-based inquiry learning model supported by AI; and (2) to test the effectiveness of the model in improving the scientific reasoning of academically underprepared students in higher education.

METHOD

Research Types and Designs

This study is a development research using a Design-Based Research (DBR) approach that aims to design, implement, and evaluate an artificial intelligence-assisted Research Project-Based Inquiry learning model to improve the scientific reasoning of academically unprepared students. The DBR approach was chosen based on the need to produce an applicable and contextual learning model that allows for model revision through iterative cycles based on the results of implementation in the field (Reilly & Reeves, 2024; Tinoca et al., 2022).

The research design follows Reeves' development model which consists of four main stages: (1) Analysis of practical problems by researchers and practitioners, (2) Development of solutions informed by existing design principles and technological innovations, (3) Iterative cycles of testing and refinement of solutions in practice, dan (4) Reflection to produce design principles and enhance implementation (Guisasola Aranzabal et al., 2021; Zhao et al., 2021). The research design is illustrated in the following diagram:

Figure 2. Design-Based Research (DBR) according to Reeves

Research Procedures

This study was conducted in four stages. The first stage was a needs analysis involving observation, documentation, and interviews with lecturers and students to identify difficulties in scientific reasoning. The second stage was the development of a preliminary model design based on a synthesis of RPBL theory, levels of inquiry, and the use of AI, particularly in the form of a natural language processing (NLP)-based tutoring system. The third stage involved the limited implementation of the model in an experimental class to assess the functionality of the model structure and the effectiveness of AI as a learning facilitator. The fourth stage was formative evaluation through quantitative and qualitative data analysis to refine the model and formulate final design principles.

Research Subjects/Participants

The subjects in this study were third-semester students majoring in Science Education at a teacher training college in Indonesia. A total of 60 students were divided into two groups: an experimental group (30 students) who received the AI-supported Research Project-Based Inquiry learning model, and a control group (30 students) who followed conventional discussion-based learning. The subjects were selected using purposive sampling based on academic readiness assessment results that showed scores below the national average (≤55 on a scale of 0-100).

Data Collection Techniques and Procedures

Data collection was conducted within a Design-Based Research (DBR) framework consisting of four iterative stages—Problem Analysis, Solution Development, Testing and Refinement, and Reflection and Implementation—while embedding a non-equivalent control group quasi-experimental design during the testing phase to evaluate effectiveness. In the problem analysis stage, baseline data were gathered through a scientific reasoning pretest and academic readiness screening to identify learning gaps among academically underprepared students, which informed the initial design principles emphasizing structured inquiry scaffolding and adaptive AI feedback. During solution development, Prototype 1 of the AI-assisted Research Project-Based Inquiry model was created, integrating AI-generated Socratic prompts, hypothesis-verification feedback, and digital research worksheets connected to a chatbot system; expert validation and limited pilot observations were used to evaluate content alignment and usability, leading to revisions in prompt sequencing and feedback clarity. In the testing and refinement stage, the model was implemented in classroom settings using experimental and control groups, and data were collected through posttests, structured observations, AI interaction logs, and semi-structured interviews; quantitative learning gains and qualitative feedback informed modifications from Prototype 1 to Prototype 2, including adjustments to task complexity and metacognitive guidance. Finally, in the reflection and implementation stage, triangulated evidence from statistical results and qualitative data was used to derive refined design principles—such as positioning AI as adaptive scaffolding rather than answer-provider and strengthening hypothesis formulation as the core cognitive driver—thereby demonstrating iterative refinement grounded in empirical evidence and aligning the DBR framework with experimental evaluation procedures.

Data Collection Instruments

The main instrument is a scientific reasoning test that has been validated and reliability checked. The Table 1 shows the instrument specifications.

Table 1. Data Collection Instruments

Scientific Reasoning Indicators	Specific Description	Context of Question	Cognitive Domain
Identify variables (IV)	Determine independent and dependent variables in a simple experiment.	Microecosystem biology experiment	C4 (Analysis)
Formulating hypotheses (PH)	Formulating predictive relationships between two variables.	Chemical reactions in science learning	C5 (Evaluation)
Data interpretation (PD)	Analyzing tables or graphs of experimental results.	Temperature and plant growth data	C4 (Analysis)
Draw conclusions (PK)	Logically conclude from empirical findings.	Practical results report	C5 (Evaluation)
Evaluate procedures (ED)	Review the steps in the experiment.	Water quality testing protocol	C6 (Creation)

Data Analysis Techniques

Data analysis was conducted quantitatively and qualitatively. Quantitative data in the form of pretest and posttest results were analyzed using independent t-tests and ANCOVA tests to determine the effect of the model on the improvement of scientific reasoning. This analysis was carried out with the help of Jamovi, JASP, Ministep Rasch, and SmartPLS software. Qualitative data from observations and interviews were analyzed using thematic analysis techniques to uncover perceptions, responses, and the effectiveness of AI support in the learning process. The validity of qualitative data was ensured through source and method triangulation. With a systematic method and an adaptive research-based design approach to the context, this study is believed to be able to produce an effective learning model that can be replicated in various higher education settings with similar characteristics.

RESULTS AND DISCUSSION

Designing and validating an artificial intelligence-supported Research Project-Based Inquiry learning model for academically unprepared students

The results of the design and validation analysis show that the developed model consists of five main components, namely: (1) structured RPBL learning syntax based on inquiry stages (guided to open inquiry); (2) the role of AI as a text-based facilitator in the project planning and reporting stages; (3) digital worksheets integrated with an AI chatbot that can provide adaptive feedback; (4) a scientific reasoning assessment rubric developed based on Lawson's indicators; and (5) a student performance monitoring tool through an interactive dashboard.

Table 2. Descriptive Statistics and Person Reliability Output from Rasch Model Analysis

Statistik	Total Score	Count	Measure	Model S.E.	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	377.8	5.0	-.05	.10	1.03	.04	1.03	.04
SEM	3.3	.0	.04	.00	.09	.13	.09	.13
P.SD	25.5	.0	.28	.00	.66	1.03	.66	1.03
S.SD	25.7	.0	.28	.00	.66	1.04	.66	1.04
MAX.	415.0	5.0	.35	.11	2.58	2.02	2.71	2.12
MIN.	334.0	5.0	-.53	.10	.14	-1.96	.14	-1.97
Person Reliability and Separation Indices								
Estimation Type		**RMSE**	**True SD**	**Separation**	**Person Reliability**		**Value**	
REAL		.12	.25	2.14	.82		-	
MODEL		.10	.26	2.48	.86		-	
S.E. of Person Mean		-	-	-	-		.04	

Based on the results of Rasch analysis in Table 2, it was found that the Person Reliability value was 0.82 (real) and 0.86 (model), which is classified as high and indicates that the instrument is capable of consistently measuring differences in ability among respondents. The person separation value of 2.14 indicates that participants can be grouped into at least two strata of significantly different abilities. The average infit and outfit MNSQ values of 1.03 are within the tolerance range of 0.5–1.5, indicating that the participants' response patterns to the items are productive and in line with the model's expectations. These findings align with the study by Yanto et al., who stated that instruments with reliability ≥0.8 and stable infit/outfit values indicate consistency with the unidimensionality assumption in the Rasch Model (Yanto et al., 2019). Furthermore, research by Bhaw et al. in Measurement: Interdisciplinary Research and Perspectives states that high separation and reliability values indicate the effectiveness of an instrument in classifying respondents accurately (Bhaw et al., 2023). With a standard error of person mean of 0.04, the estimated average ability of participants is also considered highly accurate. Thus, these results support the validity of the instrument used in measuring students' scientific reasoning accurately and in a standardized manner.

Table 3. Descriptive Statistics and Item Reliability Output from Rasch Model Analysis

Statistik	Total Score	Count	Measure	Model S.E.	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4534.0	60.0	.00	.03	1.03	.14	1.03	.15
SEM	58.8	.0	.05	.00	.10	.50	.10	.53
P.SD	117.6	.0	.11	.00	.19	1.00	.21	1.06
S.SD	131.5	.0	.12	.00	.22	1.12	.23	1.19
MAX.	4715.0	60.0	.13	.03	1.31	1.55	1.33	1.63
MIN.	4390.0	60.0	-.16	.03	.83	-.92	.83	-.96

Person Reliability and Separation Indices					
Estimation Type	**RMSE**	**True SD**	**Separation**	**Person Reliability**	**Value**
REAL	.03	.10	3.23	.91	-
MODEL	.03	.10	3.41	.92	-
S.E. of Person Mean	-	-	-	-	.05

Based on Table 3, the Item Reliability value obtained was 0.91 (real) and 0.92 (model), indicating that the instrument has a high ability to distinguish the level of difficulty between items consistently. The item separation value of 3.23 indicates that the items can be grouped into more than three different levels of difficulty, strengthening the evidence of the construct validity of the instrument. The average infit and outfit MNSQ values of 1.03 are within the ideal range (0.5–1.5), indicating that the items function as predicted by the Rasch model. The standard error of the item mean of only 0.05 indicates high precision in the estimation of item ability. These findings reinforce the view of Bezci et al., who stated that item reliability above 0.90 indicates high stability in measurement and the ability of items to generalize across different learning contexts (Bezci & Sungur, 2021). Research by Wati et al. in Measurement and Evaluation in Counseling and Development also confirms that instruments achieving item separation >2.0 and item reliability >0.9 are considered highly effective for evaluating complex cognitive aspects such as scientific reasoning (M. Wati et al., 2019). Thus, this instrument is deemed highly valid and reliable in measuring students' learning outcomes based on the implemented learning model.

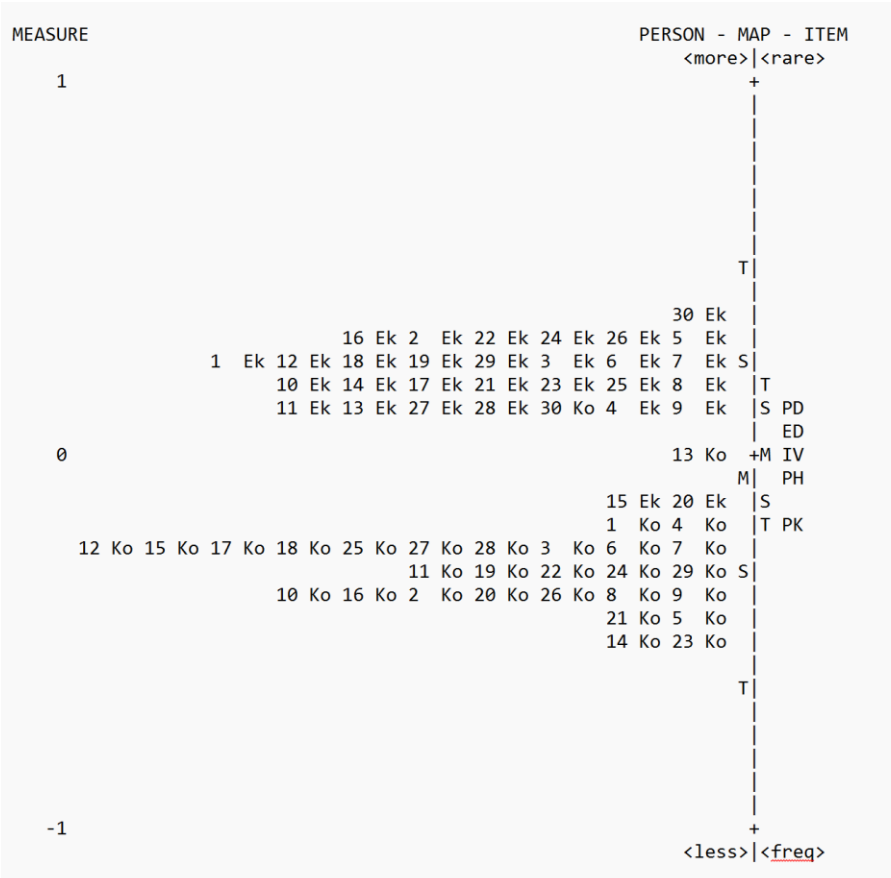

Figure 3. Rasch Model Person-Item Map between Students and Scientific Reasoning Indicators

Figure 3. Shows the distribution of student abilities (left) and the level of difficulty of the indicators (right) on a logit scale, visualized through a Person-Item Map. The data indicate that most students from the experimental group (Ek) are in the upper level (logit > 0), reflecting high scientific reasoning ability, while the majority of students from the control group (Ko) are concentrated in logit < 0, indicating lower ability. On the item side, indicators IV (Variable

Identification) and PH (Hypothesis Formulation) are at the midpoint (logit 0), indicating moderate difficulty, while PK (Conclusion Drawing) is below logit 0, suggesting that these items are easier for students to answer. This is consistent with the findings of Coleman et al. in Applied Psychological Measurement, which state that person-item maps are highly effective in visualizing the fit between individual abilities and item difficulty, and can assist in evaluating the diagnostic accuracy of instruments (Coleman et al., 2021). The proportional distribution between persons and items also indicates that the test was well-designed, as it covers various levels of student ability and avoids item difficulty bias. Therefore, it can be concluded that the instrument is capable of detecting ability differences with precision and supports the construct validity of the learning model being tested.

Table 4. Results of Person Fit Statistics Based on Rasch Analysis between Experimental and Control Group

Entry Num.	Total Score	Total Count	JMLE Measure	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PT Meas. Corr.	PT Meas. Exp.	Exact Match OBS%	Exact Match EXP%	Person
30	415	5	.35	.10	.25	-1.46	.26	-1.45	.85	.42	20.0	8.0	30 Exp.
26	412	5	.32	.10	.41	-.95	.42	-.94	.82	.42	20.0	9.3	26 Exp.
5	411	5	.31	.10	.34	-1.19	.33	-1.20	.29	.42	.0	8.9	5 Exp.
2	410	5	.30	.10	1.29	.63	1.30	.64	.89	.42	.0	10.8	2 Exp.
16	410	5	.30	.10	1.04	.27	1.05	.29	.02	.42	.0	10.8	16 Exp.
22	409	5	.28	.10	.53	-.67	.53	-.66	.11	.42	.0	9.0	22 Exp.
24	409	5	.28	.10	2.19	1.60	2.17	1.58	.86	.42	20.0	9.0	24 Exp.
7	408	5	.27	.10	.84	-.04	.85	-.03	.62	.42	40.0	9.0	7 Exp.
12	408	5	.27	.10	.68	-.34	.68	-.35	-.07	.42	20.0	9.0	12 Exp.
18	408	5	.27	.10	.31	-1.30	.31	-1.29	.89	.42	20.0	9.0	18 Exp.
29	408	5	.27	.10	.69	-.33	.68	-.34	.87	.42	.0	9.0	29 Exp.
6	407	5	.26	.10	2.56	1.93	2.55	1.92	-.80	.42	.0	9.0	6 Exp.
19	405	5	.24	.10	.50	-.75	.50	-.75	.59	.42	20.0	8.0	19 Exp.
1	404	5	.23	.10	.34	-1.22	.34	-1.22	.52	.42	20.0	9.7	1 Exp.
3	403	5	.22	.10	1.11	.39	1.10	.37	.24	.42	20.0	9.7	3 Exp.
21	402	5	.21	.10	1.51	.91	1.52	.92	-.18	.42	.0	9.7	21 Exp.
17	401	5	.20	.10	1.35	.71	1.36	.73	.82	.42	20.0	10.0	17 Exp.
25	401	5	.20	.10	.54	-.66	.55	-.66	.76	.42	20.0	10.0	25 Exp.
8	400	5	.19	.10	1.51	.91	1.52	.92	.05	.42	.0	8.8	8 Exp.
10	400	5	.19	.10	.26	-1.52	.26	-1.52	.84	.42	20.0	8.8	10 Exp.
23	399	5	.18	.10	1.20	.51	1.19	.50	.27	.42	.0	8.2	23 Exp.
14	397	5	.16	.10	1.88	1.34	1.88	1.33	-.24	.43	.0	8.2	14 Exp.
9	396	5	.15	.10	.50	-.80	.50	-.80	.27	.43	.0	7.8	9 Exp.
13	396	5	.15	.10	.20	-1.81	.20	-1.82	.74	.43	40.0	7.8	13 Exp.
60	394	5	.13	.10	2.49	1.93	2.48	1.92	.40	.43	.0	9.0	30 Con.
11	392	5	.11	.10	1.13	.41	1.13	.41	.29	.43	20.0	9.0	11 Exp.
27	392	5	.11	.10	.93	.09	.94	.10	.35	.43	.0	9.0	27 Exp.
28	392	5	.11	.10	.70	-.33	.71	-.32	.18	.43	20.0	9.0	28 Exp.
4	391	5	.10	.10	1.50	.91	1.49	.89	.52	.43	.0	7.7	4 Exp.
43	382	5	.00	.10	.70	-.33	.68	-.36	.71	.42	20.0	9.2	13 Con.
20	371	5	-.12	.11	.75	-.20	.72	-.26	.87	.41	10.0	10.5	20 Exp.
15	368	5	-.15	.11	1.01	.23	1.00	.21	.89	.41	20.0	10.8	15 Exp.
31	363	5	-.21	.11	.65	-.39	.64	-.41	-.24	.41	.0	10.2	1 Con.
34	363	5	-.21	.11	.14	-1.96	.14	-1.97	.92	.41	20.0	10.2	4 Con.
42	360	5	-.24	.11	1.69	1.09	1.68	1.08	.84	.40	.0	9.2	12 Con.
45	360	5	-.24	.11	2.21	1.60	2.19	1.58	-.31	.40	20.0	9.2	15 Con.
33	359	5	-.25	.11	1.09	.35	1.10	.36	.08	.40	.0	8.9	3 Con.
37	359	5	-.25	.11	1.09	.35	1.10	.37	.47	.40	20.0	8.9	7 Con.
55	359	5	-.25	.11	1.79	1.20	1.79	1.20	-.48	.40	20.0	8.9	25 Con.
58	359	5	-.25	.11	1.07	.33	1.06	.31	-.11	.40	.0	8.9	28 Con.
47	358	5	-.26	.11	.77	-.16	.77	-.16	.72	.40	.0	8.7	17 Con.
57	358	5	-.26	.11	2.28	1.66	2.25	1.64	.87	.40	.0	8.7	27 Con.
36	357	5	-.27	.11	.59	-.52	.59	-.51	.29	.40	.0	8.8	6 Con.
48	357	5	-.27	.11	1.40	.77	1.41	.77	.28	.40	20.0	8.8	18 Con.
54	356	5	-.29	.11	.84	-.03	.84	-.04	.32	.40	.0	9.3	24 Con.
59	354	5	-.31	.11	.45	-.84	.45	-.85	.59	.40	.0	8.3	29 Con.
41	353	5	-.32	.11	2.25	1.64	2.21	1.60	-.57	.40	.0	9.1	11 Con.
49	351	5	-.34	.11	.30	-1.29	.30	-1.31	.92	.40	.0	8.8	19 Con.
52	351	5	-.34	.11	.37	-1.08	.37	-1.06	.76	.40	.0	8.8	22 Con.

Entry Num.	Total Score	Total Count	JMLE Measure	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PT Meas. Corr.	PT Meas. Exp.	Exact Match OBS%	Exact Match EXP%	Person
39	350	5	-.35	.11	1.68	1.09	1.71	1.12	.76	.40	.0	9.6	9 Con.
46	350	5	-.35	.11	1.49	.87	1.46	.83	.38	.40	.0	9.6	16 Con.
32	349	5	-.37	.11	1.10	.37	1.07	.33	.20	.41	20.0	8.9	2 Con.
38	348	5	-.38	.11	.17	-1.85	.17	-1.84	.73	.41	20.0	9.5	8 Con.
40	348	5	-.38	.11	.84	-.04	.87	.00	.44	.41	.0	9.5	10 Con.
50	347	5	-.39	.11	.93	.10	.94	.13	.52	.41	.0	9.5	20 Con.
56	346	5	-.40	.11	.83	-.06	.84	-.05	-.26	.41	.0	8.5	26 Con.
51	341	5	-.45	.10	.36	-1.17	.37	-1.14	.64	.41	.0	10.5	21 Con.
35	340	5	-.47	.10	.31	-1.36	.31	-1.34	.52	.41	20.0	12.7	5 Con.
53	339	5	-.48	.10	2.58	2.02	2.71	2.12	-.26	.41	20.0	9.9	23 Con.
44	334	5	-.53	.10	1.21	.54	1.27	.61	.92	.42	20.0	10.6	14 Con.
MEAN	377.8	5.0	-.05	.10	1.03	.04	1.03	.04			10.0	9.2	
P.SD	25.5	.0	.28	.00	.66	1.03	.66	1.03			11.3	.9	

Table 4 presents individual statistics (person fit) based on Rasch measurement results from 60 students, consisting of experimental and control groups. The Measure value shows that the majority of students from the experimental group occupied positions with positive logit scores (≥ 0), indicating a higher level of scientific reasoning ability compared to students from the control group, which was dominated by negative logit values (≤ 0). The average Infit and Outfit MNSQ values of 1.03 and ZSTD of 0.04 indicate that the participants' response data are still within reasonable limits (range 0.5–1.5), thus fulfilling the Rasch model assumptions. The Point Measure Correlation (PtMeas Corr) values were mostly above 0.40, meaning that all items had a positive correlation with participants' abilities, indicating good item functioning. These findings support the study by Muna et al., who stated that person fit examination is important to ensure that each respondent responds to items consistently with the theoretical model used, especially in the context of inquiry-based and project-based learning evaluation (Muna & Aziz, 2021). Thus, the data in this table strengthen the empirical validity of the implemented learning model, where the experimental group students showed greater consistency and achievement of abilities compared to the control group.

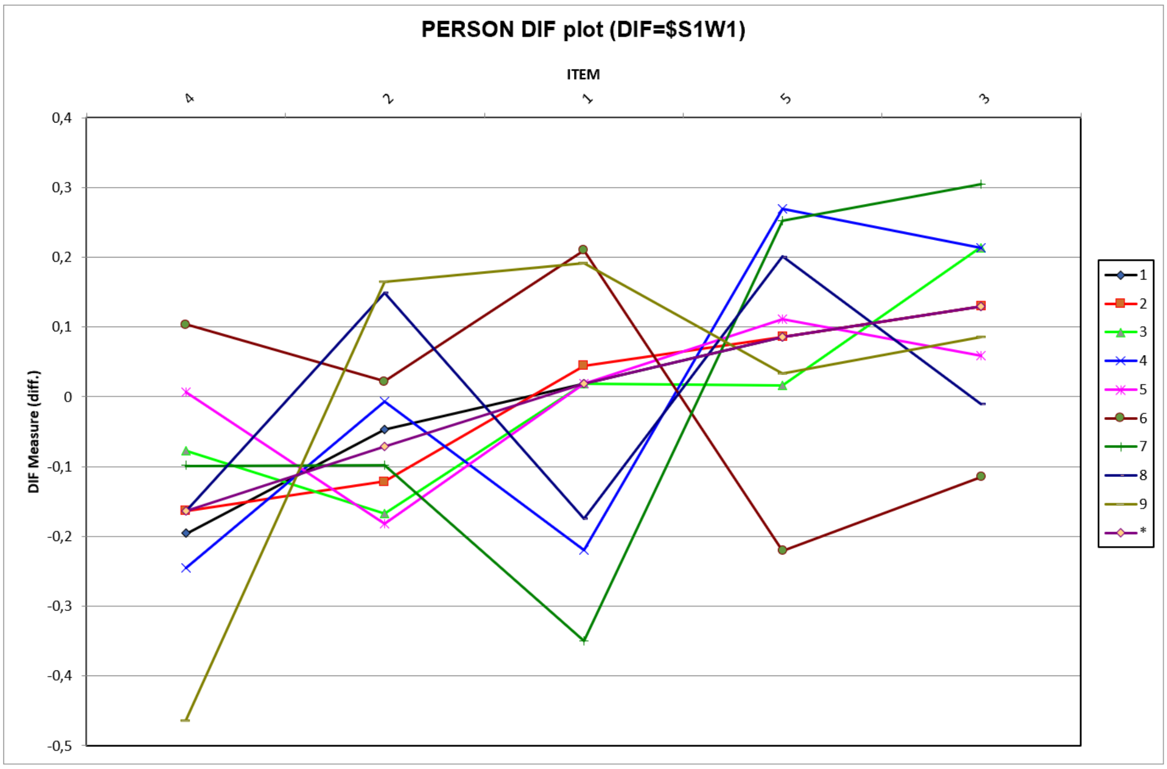

Figure 4. DIF (Differential Item Functioning) Graph Between Items Based on Student Groups

Figure 4 shows the Person DIF Plot graph, which illustrates differences in item functioning (Differential Item Functioning) based on student groups or specific respondent characteristics. Variations in the lines between items indicate that some questions show significant differences in responses depending on the student group. Items above the zero line have positive DIF, indicating that the item is more favorable to a particular group, while items below the zero line show negative DIF, meaning they are more difficult or less favorable to other groups. Although most differences are still within the tolerance range (±0.5 logit), this pattern is important to note to ensure measurement fairness. These findings are consistent with the research by Syawaludin et al. in the journal Frontiers in Psychology, which emphasizes that DIF analysis is important in identifying potential item bias toward subgroups of participants and ensuring that instruments are able to measure equitably across groups (Syawaludin et al., 2022). Therefore, although the instrument is generally valid and reliable, the presence of some items with DIF tendencies should be further considered for revision or adjustment in the broader use of the scale.

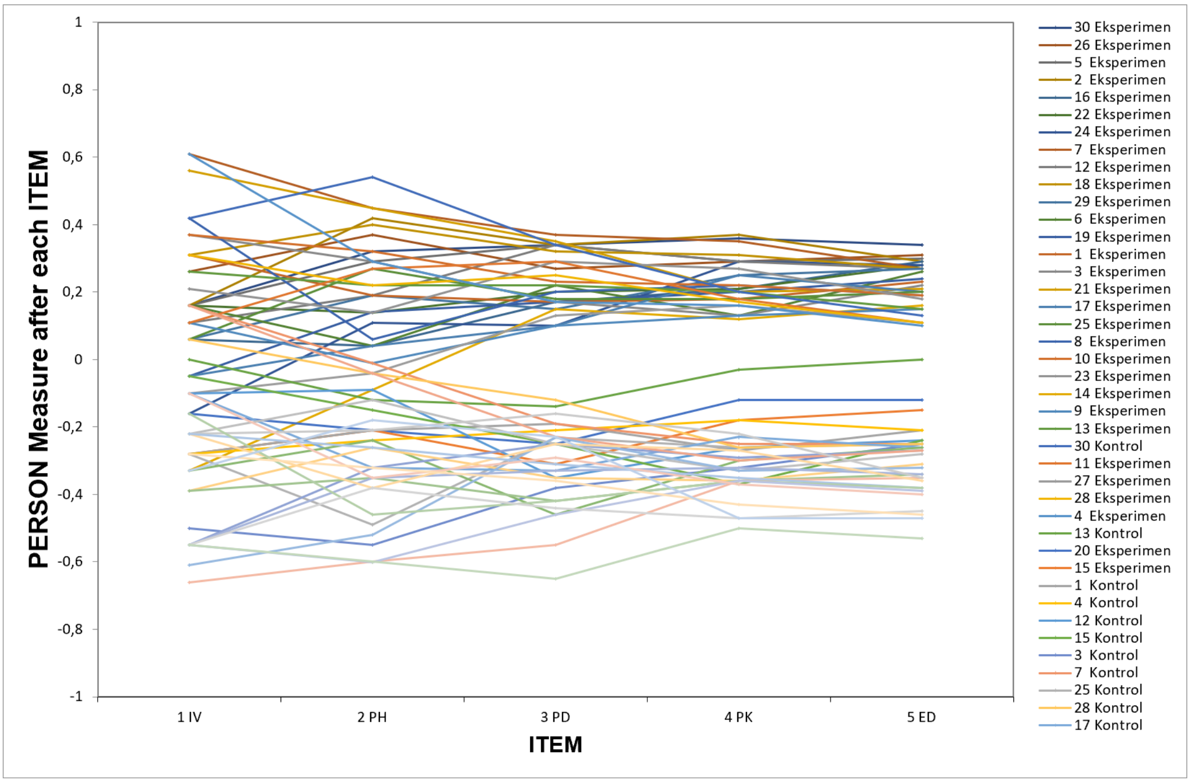

Figure 5. Comparison Chart of Person Measure Scores After Answering Each Indicator (Person Measure After Each Item)

Figure 5 shows the visualization of the development of individual ability scores (person measure) after answering each scientific reasoning indicator: IV (Identification of Variables), PH (Formulation of Hypotheses), PD (Interpretation of Data), PK (Drawing Conclusions), and ED (Evaluation of Procedures). The graph pattern indicates that most students in the experimental group consistently scored in the positive range (≥ 0), indicating improved and stable abilities across all five indicators. In contrast, students in the control group showed fluctuating scores that tended to be negative and unstable, particularly on the PD and ED indicators, which are aspects requiring higher cognitive levels. This trend confirms that the AI-based Research Project-Based Inquiry learning model not only improves overall performance but also has a significant impact on strengthening cognitive consistency across indicators. These findings align with the study by Vassilakopoulou et al. in Computers & Education, which stated that project-based learning supported by adaptive AI technology can enhance performance consistency and metacognition in students on complex tasks (Vassilakopoulou et al., 2023). Thus, this graph reflects the success of the model intervention in maintaining the continuity of students' scientific thinking processes in a more structured and meaningful manner.

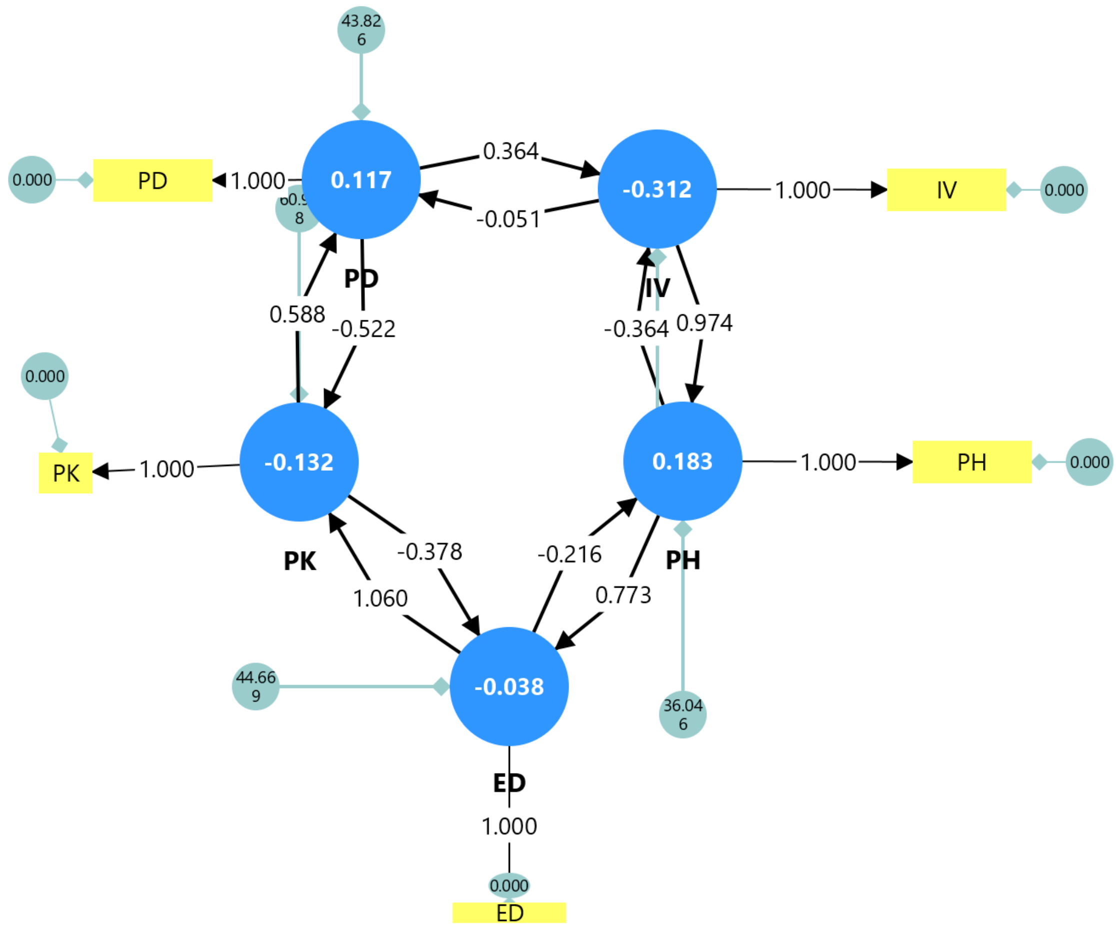

Figure 6. Relational Structure Model between Scientific Reasoning Indicators of Students (Path Analysis Diagram)

Figure 6 shows a path analysis model that illustrates the structural relationships between scientific reasoning indicators: Variable Identification (IV), Hypothesis Formulation (PH), Data Interpretation (PD), Conclusion Drawing (PK), and Procedure Evaluation (ED). The diagram shows that the PH indicator has the greatest direct influence on IV ($\beta = 0.974$), while PD is influenced by PK ($\beta = 0.588$). However, some relationships, such as PK → ED and PH → ED, show weak negative coefficient values, indicating insignificant paths or indirect effects that do not strengthen the model. The largest coefficient in the positive influence path is PH → IV, indicating that strong hypothesis formation is the primary predictor in the identification of scientific variables. This result reinforces the findings of Luo et al. in the Journal of Science Education and Technology, which affirm that the structure of scientific reasoning is hierarchical and interactive, with hypothesis formulation as a critical foundation that influences other scientific activities (Luo et al., 2020). Thus, this model demonstrates that the relationships among indicators in project-based learning are not linear but form a cognitive network that supports one another and can be utilized as a foundation for developing more precise intervention modules.

Based on the Rasch analysis results above, the AI-assisted Research Project-Based Inquiry model that was designed has been strongly validated with high reliability (Person Reliability = 0.82; Item Reliability = 0.91) and the ability to effectively differentiate individuals and test items (Person Separation = 2.14; Item Separation = 3.23). The person-item and person fit statistics indicate that the majority of experimental students are in the positive logit range with consistent response patterns, while the DIF graphs and person scores after each indicator show fair distribution and stability in cognitive performance. These findings confirm that the developed model is valid, fair, and adaptive for improving the scientific reasoning of academically underprepared students in a systematic and evidence-based manner.

Measuring the effectiveness of models in improving scientific reasoning among academically unprepared students.

To test the effectiveness of the AI-based Research Project-Based Inquiry model, a comparative analysis of scientific reasoning scores between the experimental and control groups was conducted to identify statistically significant improvements. This analysis aims to ensure that the developed model has a real impact on strengthening the scientific thinking skills of academically unprepared students.

Table 5. Descriptive Statistics of Students' Scientific Reasoning Scores Based on Groups and Indicators

	Group	N	Mean	Median	SD	SE
IV	Experiment	30	79.1	79.0	4.58	0.836
	Control	30	71.4	71.0	4.70	0.858
PH	Experiment	30	81.4	81.5	4.67	0.853
	Control	30	72.4	73.0	5.22	0.953
PD	Experiment	30	78.1	78.0	5.03	0.918
	Control	30	68.2	68.0	5.18	0.945
PK	Experiment	30	83.8	84.0	4.54	0.829
	Control	30	73.3	72.5	5.80	1.058
ED	Experiment	30	78.4	79.0	5.20	0.949
	Control	30	69.5	70.0	4.58	0.836

Table 5 shows the differences in the average scientific reasoning scores of students on five indicators between the experimental and control groups. The experimental group consistently scored higher on all indicators, with the most significant difference observed in the Drawing Conclusions (DC) indicator, which was 10.5 points (Experimental: 83.8; Control: 73.3), followed by Hypothesis Formulation (HF) at 9 points. The relatively small standard deviation (SD) and standard error (SE) in both groups indicate homogeneous data and precise mean estimates. These differences indicate that the application of the AI-based Research Project-Based Inquiry model can significantly improve students' scientific thinking quality. These findings align with the research results by Zudaire et al., which confirm that project-based and inquiry-based learning can strengthen the integration of concepts and higher-order thinking skills, especially when supported by interactive technology capable of providing adaptive feedback tailored to participants' needs (Zudaire et al., 2022). Thus, the tested model has proven effective in enhancing the scientific reasoning abilities of academically unprepared students.

Table 6. Normality Test of Scientific Reasoning Score Distribution (Shapiro-Wilk Test)

	W	p
IV	0.979	0.390
PH	0.944	0.009
PD	0.989	0.857
PK	0.981	0.488
ED	0.985	0.665

Note. A low p-value suggests a violation of the assumption of normality

Table 6 shows the results of the Shapiro-Wilk normality test for each scientific reasoning indicator. Four of the five indicators, namely IV ($p = 0.390$), PD ($p = 0.857$), PK ($p = 0.488$), and ED ($p = 0.665$), have p-values > 0.05, which means that their distributions meet the normality assumption. However, the PH (Hypothesis Formulation) indicator shows a p-value of $0.009 < 0.05$, indicating that the data for this indicator are not normally distributed. This suggests that further analysis of the PH indicator should consider non-parametric methods or data transformation. This finding is consistent with the argument by Indahsari et al., who state that normality tests are crucial in determining the appropriate statistical approach, especially when the sample size is small or the data distribution tends to be skewed (Indahsari et al., 2020). Therefore, although most of the data can be analyzed using parametric approaches, caution is still needed for indicators with violations of the normality assumption, such as PH.

Table 7. Variance Homogeneity Test (Levene's Test) on Scientific Reasoning Scores

	F	df	df2	p
IV	0.0215	1	58	0.884
PH	0.0108	1	58	0.918
PD	0.0619	1	58	0.804
PK	28.283	1	58	0.098
ED	14.353	1	58	0.236

Note. A low p-value suggests a violation of the assumption of equal variances

Table 7 presents the results of the Levene test to examine the assumption of homogeneity of variance between groups on five scientific reasoning indicators. All p-values for the indicators IV ($p = 0.884$), PH ($p = 0.918$), PD ($p = 0.804$), PK ($p = 0.098$), and ED ($p = 0.236$) are greater than 0.05, indicating that the variances between the experimental and control groups are homogeneous and there is no violation of the assumption of equality of variances. Although the p-value for the PK indicator is close to the significance threshold, it remains within the tolerance limit. This result allows for valid parametric statistical analysis, such as an independent t-test, to compare scores between groups. These findings are supported by the literature from Wati et al., who emphasize that the fulfillment of the assumption of homogeneity of variances is a prerequisite for the validity of mean difference analysis between two groups (D. A. Wati et al., 2023). Thus, the Levene test provides a strong statistical foundation for proceeding with parametric inferential analysis of the effectiveness of the developed learning model.

Table 8. Results of Independent t-tests on Scientific Reasoning Scores of Students between the Experimental and Control Groups

		Statistic	df	p
IV	Student's t	6.43	58.0	<.001
PH	Student's t	7.06	58.0	<.001
PD	Student's t	7.49	58.0	<.001
PK	Student's t	7.81	58.0	<.001
ED	Student's t	7.04	58.0	<.001

Note. H_a $\mu_{Eksperimen} \neq \mu_{Kontrol}$

Table 8 presents the results of independent sample t-tests comparing the posttest scores of the experimental and control groups across all scientific reasoning indicators. The analysis shows statistically significant differences for IV, PH, PD, PK, and ED (all $p < 0.001$), indicating that students exposed to the AI-based Research Project-Based Inquiry model achieved higher performance than those in the comparison group. The largest mean differences were observed in the PK ($t = 7.81$) and PD ($t = 7.49$) indicators, suggesting a substantial impact of the intervention on data interpretation and conclusion drawing skills. Prior to conducting the t-tests, assumptions of normality and homogeneity of variance were examined and met, supporting the appropriateness of the analysis. These findings are consistent with previous research demonstrating that project-based learning enriched with digital technology enhances higher-order cognitive processes, including scientific reasoning and problem-solving (Lupión-Cobos et al., 2022). Therefore, based on the independent t-test results, the intervention demonstrates statistically significant effectiveness in improving core components of scientific reasoning among academically underprepared students.

CONCLUSION

This study demonstrates that the AI-supported Research Project-Based Inquiry model meaningfully changes students' learning processes and outcomes by strengthening their capacity to identify variables, formulate hypotheses, interpret data, draw conclusions, and evaluate procedures in an integrated manner. Rather than merely improving test scores, the model restructures classroom practice by embedding inquiry cycles within research-oriented projects and augmenting them with AI-based scaffolding that provides guided questioning, iterative feedback, and metacognitive prompts. A plausible mechanism underlying these improvements is the combination of structured inquiry phases with adaptive AI support, which helps academically underprepared students regulate their reasoning processes and progressively refine their scientific arguments. However, these findings must be interpreted within defined boundaries: the study was conducted at a single institution using purposive, non-randomized groups; the intervention duration was limited; and the implementation relied on a specific AI tool, which may introduce dependency effects or novelty influences. The model appears to work under conditions where students demonstrate low initial readiness, instructors receive adequate orientation, and inquiry tasks are systematically aligned with AI-mediated scaffolding. For replication, several design principles emerge: (1) explicit alignment between inquiry stages and AI-generated prompts, (2) iterative feedback loops embedded within project milestones, (3) structured hypothesis formulation as a central reasoning anchor, and (4) continuous monitoring of student engagement through analytics dashboards. Future research should test the model across multiple disciplines and institutions, employ longer follow-up periods to examine sustainability of reasoning gains, and compare AI-supported scaffolding with non-AI inquiry facilitation to isolate the added value of intelligent assistance.

RECOMMENDATION

Based on the findings of this study, the AI-supported Research Project-Based Inquiry model is recommended for use in foundational science courses, particularly for academically underprepared students who require structured guidance in developing scientific reasoning. Its implementation should be accompanied by careful instructional planning, including the alignment of inquiry stages with research project activities, the preparation of AI-generated scaffolding prompts, and the use of formative feedback to support students in identifying variables, formulating hypotheses, interpreting data, drawing conclusions, and evaluating procedures. Lecturers are encouraged to position AI as a cognitive scaffold rather than as a substitute for direct instruction, so that students remain actively involved in reasoning, argument construction, and reflective decision-making. Institutions intending to adopt this model should provide adequate lecturer training, reliable digital infrastructure, and ethical guidelines for the responsible use of AI in learning. Future studies are recommended to test the model across different disciplines, institutions, and student populations, using larger samples, longer intervention periods, and comparative designs that distinguish the specific contribution of AI-based scaffolding from conventional inquiry facilitation.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) for its support through a Fundamental Research Grant, awarded based on contract No. 110/E5/PG.02.00.PL/2024. This grant has made a significant contribution to the implementation of our research in 2024, which aims to develop and improve the quality of education in Indonesia. We hope that the results of this research will have a positive impact on the academic world and contribute to the advancement of knowledge

FUNDING INFORMATION
This research was funded by Ministry of Education, Culture, Research, and Technology (Kemendikbudristek), grant number (contract): 110/E5/PG.02.00.PL/2024

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Asriyadin	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓
Adi Apriadi Adiansha	✓	✓	✓		✓	✓	✓	✓		✓	✓			✓
Khairil Anwar		✓		✓	✓	✓	✓	✓		✓	✓			✓
Syarifuddin				✓		✓			✓			✓	✓	✓

CONFLICT OF INTEREST STATEMENT
The authors state no conflict of interest.

INFORMED CONSENT
We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL
The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY
The data that support the findings of this study are available on request from the corresponding author.

REFERENCES

Abate, T., Michael, K., & Angell, C. (2020). Assessment of Scientific Reasoning: Development and Validation of Scientific Reasoning Assessment Tool. *Eurasia Journal of Mathematics, Science and Technology Education*. https://doi.org/10.29333/ejmste/9353

Agustina, W., Degeng, I. N. S., Praherdhiono, H., & Lestaric, S. R. (2023). The effect of blended project-based learning for enhancing student’s scientific literacy skills: An experimental study in University. *Pegem Journal of Education and Instruction*, *13*(1). https://doi.org/10.47750/pegegog.13.01.24

Arantes do Amaral, J. A., & Lino dos Santos, R. J. R. (2018). Combining Project-Based Learning and Community-Based Research in a Research Methodology Course: The Lessons Learned. *International Journal of Instruction*, *11*(1), 47–60. https://doi.org/10.12973/iji.2018.1114a

Arifin, Z., Sukarmin, S., Saputro, S., & Kamari, A. (2025). The effect of inquiry-based learning on students’ critical thinking skills in science education: A systematic review and meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(3), em2592. https://doi.org/10.29333/ejmste/15988

Bezci, F., & Sungur, S. (2021). How is Middle School Students’ Scientific Reasoning Ability Associated with Gender and Learning Environment? *Science Education International*, *32*(2), 96–106. https://doi.org/10.33828/sei.v32.i2.2

Bhaw, N., Kriek, J., & Lemmer, M. (2023). Insights from coherence in students’ scientific reasoning skills. *Heliyon*, *9*(7), e17349. https://doi.org/10.1016/j.heliyon.2023.e17349

Chen, J. (2024). Construction and Application of an Economic Intelligent Decision-Making Platform Based on Artificial Intelligence Technology. *Informatica (Slovenia)*, *48*(9), 89–106. https://doi.org/10.31449/inf.v48i9.5705

Coleman, B. M., Bunch, J. C., Thoron, A. C., & Roberts, T. G. (2021). Examining the Effects of Reflection Type and Abstraction Order on Students’ Scientific Reasoning Skills During Experiential Learning. *Journal of Agricultural Education*. https://doi.org/10.5032/jae.2021.02013

Gerhátová, Ž., Perichta, P., Drienovský, M., & Palcut, M. (2021). Temperature Measurement—Inquiry-Based Learning Activities for Third Graders. *Education Sciences*, *11*(9), 506. https://doi.org/10.3390/educsci11090506

Guisasola Aranzabal, J., Ametller, J., & Zuza, K. (2021). Investigación basada en el diseño de Secuencias de Enseñanza-Aprendizaje: una línea de investigación emergente en Enseñanza de las Ciencias. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, *18*(1), 1–18. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2021.v18.i1.1801

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, *15*(2), ep421. https://doi.org/10.30935/cedtech/13036

Indahsari, S. N., Supeno, & Maryani. (2020). Student worksheet based on inquiry with vee map to improve students' scientific reasoning ability in physics learning in senior high school. *Journal of Physics: Conference Series*, *1465*, 012036. https://doi.org/10.1088/1742-6596/1465/1/012036

Jia, L., Jalaludin, N. A., & Rasul, M. S. (2023). Design Thinking and Project-Based Learning (DT-PBL): A Review of the Literature. *International Journal of Learning, Teaching and Educational Research*. https://doi.org/10.26803/ijlter.22.8.20

Kah Choon Low, Siti Syuhadah Mohamad, Say Lee Chong, Muhammad Azwan Abd Rahman, Eko Priyo Purnomo, & Ardi Gunsuh. (2024). Improving University Students' Critical Thinking and Problem-Solving Skills: How Problem-Based Learning Works during COVID-19 Pandemic? *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *37*(2), 165–176. https://doi.org/10.37934/araset.37.2.165176

Lukito, A., & Madzkiyah, A. F. (2025). Students' Critical Thinking Dispositions in View of Cognitive Styles: Exhibiting Confidence in Reasoning and Inquisitive. *Perspectives of Science and Education*, *73*(1), 583–594. https://doi.org/10.32744/pse.2025.1.37

Luo, X., Wei, B., Shi, M., & Xiao, X. (2020). Exploring the impact of the reasoning flow scaffold (RFS) on students' scientific argumentation: Based on the structure of observed learning outcomes (SOLO) taxonomy. *Chemistry Education Research and Practice*. https://doi.org/10.1039/c9rp00269c

Lupión-Cobos, T., Girón-Gambero, J., & García-Ruiz, C. (2022). Building STEM Inquiry-Based Teaching Proposal Through Collaborations Between Schools and Research Centres: Students' and Teachers' Perceptions. *European Journal of Educational Research*. https://doi.org/10.12973/eu-jer.11.2.899

Maghfiroh, L., & Shofiyah, N. (2023). Exploring the Influence of the Evidence-Based Reasoning Model in the Inquiry Approach to Enhancing Students' Scientific Reasoning. *IJIS Edu : Indonesian Journal of Integrated Science Education*, *5*(2), 136. https://doi.org/10.29300/ijisedu.v5i2.10737

Muna, F., & Aziz, I. N. (2021). Mastering Students' Speaking Skill using Inquiry Online Project-based Strategy. *Indonesian Journal of Instructional Media and Model*. https://doi.org/10.32585/ijimm.v3i1.984

Naviri, S., Sumaryanti, S., & Paryadi, P. (2021). Explanatory Learning Research: Problem-Based Learning or Project-Based Learning? *Acta Facultatis Educationis Physicae Universitatis Comenianae*, *61*(1), 107–121. https://doi.org/10.2478/afepuc-2021-0010

Nielsen, H. D. (2007). Empowering communities for improved educational outcomes: Some evaluation findings from the world bank. *Prospects*, *37*(1), 81–93. https://doi.org/10.1007/s11125-007-9018-x

OECD. (2023a). *PISA 2022 Results (Volume I)*. OECD. https://doi.org/10.1787/53f23881-en

OECD. (2023b). *PISA 2022 Results (Volume II)*. OECD. https://doi.org/10.1787/a97db61c-en

Reilly, C., & Reeves, T. C. (2024). Refining active learning design principles through design-based research. *Active Learning in Higher Education*. https://doi.org/10.1177/14697874221096140

Sahidun, M., Suyitno, A., & Pujiastuti, E. (2023). Mathematical critical thinking ability through

Brain based learning model in view of self-regulated learning. *International Journal of Trends in Mathematics Education Research*, *6*(2), 177–185. https://doi.org/10.33122/ijtmer.v6i2.213

Syawaludin, A., Prasetyo, Z. K., Jabar, C. S. A., & Retnawati, H. (2022). The Effect of Project-based Learning Model and Online Learning Settings on Analytical Skills of Discovery Learning, Interactive Demonstrations, and Inquiry Lessons. *Journal of Turkish Science Education*. https://doi.org/10.36681/tused.2022.140

Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-Based Research in the Educational Field: A Systematic Literature Review. In *Education Sciences*. https://doi.org/10.3390/educsci12060410

Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, I. O. (2023). Developing human/AI interactions for chat-based customer services: lessons learned from the Norwegian government. *European Journal of Information Systems*, *32*(1), 10–22. https://doi.org/10.1080/0960085X.2022.2096490

Wang, S. (2022). Critical Thinking Development Through Project-Based Learning. *Journal of Language Teaching and Research*, *13*(5), 1007–1013. https://doi.org/10.17507/jltr.1305.13

Wang, S., Sun, Z., & Chen, Y. (2023). Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Education and Information Technologies*, *28*(5), 4919–4939. https://doi.org/10.1007/s10639-022-11338-4

Wati, D. A., Widodo, W., & Indana, S. (2023). Identify Student's Scientific Reasoning Skill in Straight Motion Material. *International Journal of Current Educational Research*, *2*(1), 39–47. https://doi.org/10.53621/ijocer.v2i1.195

Wati, M., Mahtari, S., Hartini, S., & Amalia, H. (2019). A Rasch model analysis on junior high school students' scientific reasoning ability. *International Journal of Interactive Mobile Technologies*. https://doi.org/10.3991/ijim.v13i07.10760

Wiseman, E., Carroll, D. J., Fowler, S. R., & Guisbert, E. (2020). Iteration in an Inquiry-Based Undergraduate Laboratory Strengthens Student Engagement and Incorporation of Scientific Skills. *Journal of the Scholarship of Teaching and Learning*, *20*(2). https://doi.org/10.14434/josotl.v20i2.26794

Yanti, F. A., Kuswanto, H., Mundilarto, & Habibi. (2019). Development of cooperative research project based learning models to improve research and communication skills for prospective physics teachers in Indonesia. *International Journal of Engineering and Advanced Technology*. https://doi.org/10.35940/ijeat.E1105.0585C19

Yanto, B. E., Subali, B., & Suyanto, S. (2019). Improving students' scientific reasoning skills through the three levels of inquiry. *International Journal of Instruction*. https://doi.org/10.29333/iji.2019.12444a

Zhao, L., He, W., & Su, Y.-S. (2021). Innovative Pedagogy and Design-Based Research on Flipped Learning in Higher Education. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.577002

Zudaire, I., Buil, R., Uriz, I., & Napal, M. (2022). Mars Explorers: A Science Inquiry-Based Learning Project in Preschool. *International Journal of Early Childhood*. https://doi.org/10.1007/s13158-021-00308-5