**Jurnal Review Pendidikan Dasar:**
*Jurnal Kajian Pendidikan dan Hasil Penelitian*
*Vol 12(1), Februari 2026*

e-ISSN: 2460-8475
http://journal.unesa.ac.id/index.php/PD
https://doi.org/10.26740/jrpd.v12n1.p128-147

# Higher-Order Thinking Skills Outcomes in Integrated Natural and Social Sciences Learning Following the Implementation of Discovery Learning among Fifth-Grade Students

**Erlina** ✉, Kuningan University, Indonesia
**Leviana Putri,** Kuningan University, Indonesia
**Heryani,** Kuningan University, Indonesia
**Ina Setiawati,** Kuningan University, Indonesia
**Mia Zultrianti Sari,** Kuningan University, Indonesia
**M. Deni Ruhiyat,** Kuningan University, Indonesia
**Ilham Nur Ikhsan,** Kuningan University, Indonesia
**Frian Sumarya,** Kuningan University, Indonesia

✉ *mia.zultrianti.sari@uniku.ac.id*

**Abstract:** Discovery learning is widely promoted as a means of developing higher-order thinking skills (HOTS), and evidence on HOTS outcomes after its implementation in routine elementary classrooms remains limited, so this study aimed to describe the Natural and Social Sciences learning outcomes of fifth-grade students, their performance on essay items at the C2, C4, C5, and C6 levels, and the factors that the classroom teacher considered relevant. A descriptive quantitative design was applied to 20 fifth-grade students and complemented by a semi-structured teacher interview, and the data were analysed with descriptive statistics and a one-sample *t*-test against the Minimum Mastery Criterion (*Kriteria Ketuntasan Minimal*, KKM = 75). The mean score was 46.75 (*SD* = 18.73, very low category), one student (5%) reached mastery, the mean was reliably below the KKM ($t(19) = -6.75$, $p < .001$, $d = -1.51$), item-level correct response rates ranged from 7% (C6) to 28% (C4), and the teacher referred to limited literacy, unfamiliarity with essay questions, limited parental guidance, and heterogeneous learning needs. Because the design has no comparison group or pretest, the findings describe post-implementation outcomes and suggest that guided discovery, diagnostic attention to prior knowledge, and regular HOTS essay practice merit empirical testing in elementary IPAS classrooms.

**Keywords:** higher-order thinking skills; discovery learning; Natural and Social Sciences; elementary school; scaffolding.

**How to Cite**: Erlina, E., Putri, L., Heryani, H., Setiawati, I., Sari, M.Z., Ruhiyat, M.D, Ikhsan, I.N & Sumarya, F. (2026). Higher-Order Thinking Skills Outcomes in Integrated Natural and Social Sciences Learning Following the Implementation of Discovery Learning among Fifth-Grade Students. *Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan dan Hasil Penelitian, 12*(1), 128-147. https://doi.org/10.26740/jrpd.v12n1.p128-147

© The Author(s) 2026

**Article Info:**
Received: 12 January 2026; Accepted: 12 February 2026; Published Online: 28 February 2026

**Published by:**
**Pendidikan Dasar Universitas Negeri Surabaya**

## INTRODUCTION

Higher-order thinking skills (HOTS) refer to cognitive processes such as analysing, evaluating, and creating, which correspond to the C4, C5, and C6 levels of the revised Bloom's taxonomy (Ichsan et al., 2019). Studies of HOTS in Indonesian schools commonly arrange cognitive processes from remembering and understanding to applying, analysing, evaluating, and creating (Hadi et al., 2018; Ramdiah et al., 2019). In Indonesian elementary schools, the *Kurikulum Merdeka* (Independent Curriculum) presents science and social studies content in a combined subject, *Ilmu Pengetahuan Alam dan Sosial* (Natural and Social Sciences, IPAS). Teachers are therefore expected to develop students' reasoning through learning activities and assessment items that require analysis, evaluation, and the creation of solutions (Retnawati et al., 2018; Tiruneh et al., 2017).

Previous studies have reported that HOTS can be developed through instruction in science. Research by Yang et al. (2016) reported that students who received creative inquiry-based science teaching outperformed a comparison group on science inquiry and convergent thinking performances. In a study of project-based and problem-based learning, Anazifa and Djukri (2017) reported that both models affected students' creativity and critical thinking, and that the two models did not differ in their effect on critical thinking. Other studies have examined the conditions around HOTS, including the factors associated with college students' HOTS in a smart classroom environment (Lu et al., 2021) and the way teachers understand, plan, and implement HOTS-oriented learning (Ramdiah et al., 2019).

Discovery learning has been debated mainly in terms of the amount of guidance it provides. Research by Shieh and Yu (2016) with 98 secondary school students reported that guided discovery instruction affected learning achievement and learning retention. Weaver et al. (2018) reported that undergraduate physics students who solved problems before instruction showed better conceptual knowledge than students who received instruction first, which shows that exploration and instruction can be sequenced in different ways.

Evidence from primary classrooms points to the role of explicit support in inquiry. Kruit et al. (2018) reported that an eight-week intervention with explicit instruction on inquiry skills facilitated the acquisition of science inquiry skills among students in grades

5 and 6. Taken together, these reports suggest that the label discovery learning covers classroom practices with different levels of support. This variation makes it important to examine outcomes in specific classrooms where the level of guidance is uncertain.

Evidence from Indonesian and elementary classrooms has also been reported. In a quasi-experimental study in senior high schools, Yerimadesi et al. (2023) reported higher HOTS scores for students taught with a guided discovery learning-based chemistry e-module than for control students. In a study of 60 students in Surakarta, Sidiq et al. (2021) reported that habituation to HOTS-based science questions in essay format was associated with higher critical thinking scores than conventional instruction. Evidence from a grade 6 science classroom indicates that learning experiences focused on critical thinking had a significant influence on students' critical thinking and scientific literacy (Vieira & Tenreiro-Vieira, 2016).

In the fifth-grade IPAS classroom examined in the present study, the teacher had implemented discovery learning, and initial information from the school indicated that many students had not reached the *Kriteria Ketuntasan Minimal* (Minimum Mastery Criterion, KKM) of 75 on essay-based assessments targeting the C4, C5, and C6 levels. This pattern differs from the favourable outcomes reported in several of the studies above, although low HOTS scores among Indonesian elementary students have also been documented (Ichsan et al., 2019). Differences in study design, comparison conditions, educational level, assessment format, and the amount of guidance provided may be associated with such variation. The contribution of these differences in routine elementary classrooms has received limited empirical attention.

Existing studies have primarily focused on the effectiveness of discovery-based and inquiry-based approaches in experimental or quasi-experimental designs (Anazifa & Djukri, 2017; Yang et al., 2016). Yerimadesi et al. (2023) worked with senior high school students, Lu et al. (2021) with college students, and Aldalur and Perez (2023) reported a case study of gamified discovery learning in a university software engineering course. Less attention has been directed toward how elementary students perform on essay-format HOTS items after discovery learning in routine IPAS classrooms, particularly in descriptive, criterion-referenced evaluations that report performance by cognitive level and consider the factors teachers associate with low performance (Hadi et al., 2018; Ichsan et al., 2019).

This study addresses an aspect of discovery learning research that has received limited attention. The aim of this study is to describe the IPAS learning outcomes of fifth-grade students after the implementation of discovery learning, to describe their performance on essay items at the C2, C4, C5, and C6 levels, and to identify the factors that the classroom teacher considered relevant to these outcomes. This study extends previous research by examining HOTS outcomes through a criterion-referenced essay test, reported by cognitive level and complemented by a teacher interview, within a routine fifth-grade IPAS classroom. Because the study has no comparison group or pretest, it describes post-implementation outcomes and leaves causal questions to subsequent comparative research.

This study is guided by three research questions: 1) What are the IPAS learning outcomes of fifth-grade students on a HOTS-oriented essay test after discovery learning was implemented, and does the class mean differ from the KKM of 75? 2) How do students perform on essay items at the C2, C4, C5, and C6 levels, and which items show the lowest correct response rates? 3) Which factors does the classroom teacher identify as relevant to students' HOTS performance in IPAS? These questions are answered through the test scores, the item-level results, and the teacher interview reported in the sections that follow.

## METHODS

### Research Design

This study used a descriptive quantitative design, supplemented by a semi-structured interview with the classroom teacher. The design describes students' scores after discovery learning had been implemented in IPAS and compares the class mean with a fixed criterion. The researchers did not assign or manipulate any treatment, and no pretest or control group was used. For the one-sample comparison, the null hypothesis stated that the population mean is equal to or lower than the KKM ($H_0$: $\mu \leq 75$) and the alternative hypothesis stated that the mean exceeds the KKM ($H_1$: $\mu > 75$), tested one-tailed at $\alpha = .05$.

### Setting and Participants

The study was conducted at SDN Sukamulya (Sukamulya State Elementary School), a public elementary school in Garawangi, Kuningan Regency, Indonesia. All fifth-grade students with valid post-test records participated ($N = 20$; 12 male and 8 female students,

aged 11 to 12 years), so the sample coincided with the population of the class. The fifth-grade classroom teacher, who implemented discovery learning in IPAS lessons, was interviewed. Data are reported in aggregate form without student names, and the number of lessons and the fidelity of implementation were not systematically recorded in this study.

**Instruments and Data Sources**

The HOTS essay test consisted of five items aligned with the IPAS learning objectives and with the revised Bloom's taxonomy, namely one item at C2 (understanding), two items at C4 (analysing), one item at C5 (evaluating), and one item at C6 (creating). Each item was scored from 0 to 20, which produced a maximum total score of 100. Each student's HOTS percentage was calculated as $P = (X / N) \times 100\%$, where X is the total score obtained by the student and N is the maximum score of the test (100). Percentages were classified with the five-level scale in Table 1, and mastery was judged against the KKM of 75.

A semi-structured interview with the classroom teacher addressed the implementation of discovery learning and students' profiles, including their emotional, physical, and cognitive conditions and family backgrounds. Archived records of students' cognitive assessment scores served as documentation to support the test data. Interview responses were summarised descriptively and used to contextualise the test results. Item-level performance was reported as the percentage of the 20 students whose answers to an item were judged correct.

**Data Analysis**

Test data were processed in Microsoft Excel to obtain the maximum value, minimum value, range, class intervals, and frequency distribution. The class mean was compared with the KKM using a one-sample *t*-test ($\alpha = .05$, one-tailed, $df = 19$). The 95% confidence interval of the mean and the standardised mean difference ($d = (M - \mu_0) / SD$) were also calculated. Figure 1 summarises how these procedures connect the data sources to the research questions.

Table 1 presents the five-level percentage scale used to classify HOTS scores. The scale ranges from very low (0 to 49) to very good (80 to 100) and was applied to the class mean and to item-level correct response rates. It is placed in the Methods section so that the criteria are visible before any result is read. The table follows below.

**Table 1.** Percentage categories used to classify HOTS scores

| Percentage | Category |
|---|---|
| 80 to 100 | Very good |
| 70 to 79 | Good |
| 60 to 69 | Fair |
| 50 to 59 | Low |
| 0 to 49 | Very low |

Table 1 was included because every category label in the Results depends on this scale. Readers can use it to see that a mean of 46.75% falls in the very low category and that no item-level rate reaches the low category. It benefits the reader by making the interpretation of scores transparent and replicable in other classrooms. It also helps reviewers judge whether the category labels are applied consistently across the article.

Figure 1 depicts the flow of the study from participants to data sources, analysis, and research questions. It shows that the essay test, the teacher interview, and the documentation each feed specific analyses. The figure is placed after the description of the procedures so that it summarises them in one view. The figure appears below.

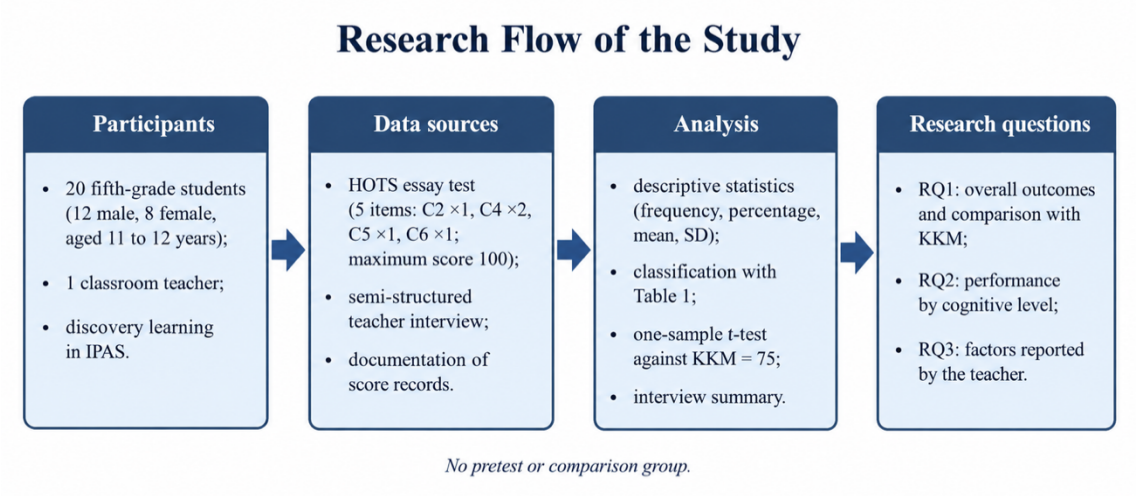



**Figure 1.** Research flow of the study

Figure 1 was included because the study combines test scores with interview data, and readers need to see which data answer which question. It reduces the effort of tracing the link between instruments and research questions across several paragraphs. The benefit for readers is a clear picture of the scope of the design, including the absence of a pretest and comparison group. The benefit for reviewers is a quick check that every research question is matched with a data source and an analysis.

## RESULTS

### Overall Learning Outcomes

Table 2 presents the frequency distribution of post-test scores for the 20 students. Four students (20%) scored between 20 and 33, six (30%) between 34 and 47, four (20%) between 48 and 61, and five (25%) between 62 and 75. No student scored between 76 and 89, and one student (5%) scored between 90 and 103, so 19 students (95%) did not reach the KKM. Half of the students (50%) scored between 20 and 47, and the class mean percentage (46.75%) falls in the very low category.

**Table 2.** Frequency distribution of fifth-grade IPAS post-test scores ($N = 20$)

| Interval | *f* | % | Mastery status |
|---|---|---|---|
| 20-33 | 4 | 20 | Not yet mastered |
| 34-47 | 6 | 30 | Not yet mastered |
| 48-61 | 4 | 20 | Not yet mastered |
| 62-75 | 5 | 25 | Not yet mastered |
| 76-89 | 0 | 0 | Not yet mastered |
| 90-103 | 1 | 5 | Mastered |
| Total | 20 | 100 | |

Table 2 was included because the class mean alone cannot show how scores were spread across the class. It lets readers see that scores were concentrated in the lower intervals and that mastery was reached by one student only. The benefit is a direct view of the proportion of students below the criterion, which is the quantity that matters for classical mastery decisions. Teachers and school leaders can use the distribution to identify how many students need support and how large the gap to the KKM is.

Figure 2 displays the same distribution as Table 2 in the form of a bar chart. The bars show the number of students in each score interval, with the one student in the 90 to 103 interval highlighted as the only case of mastery. A note on the figure marks the class mean (46.75) and the KKM (75). The figure appears below.

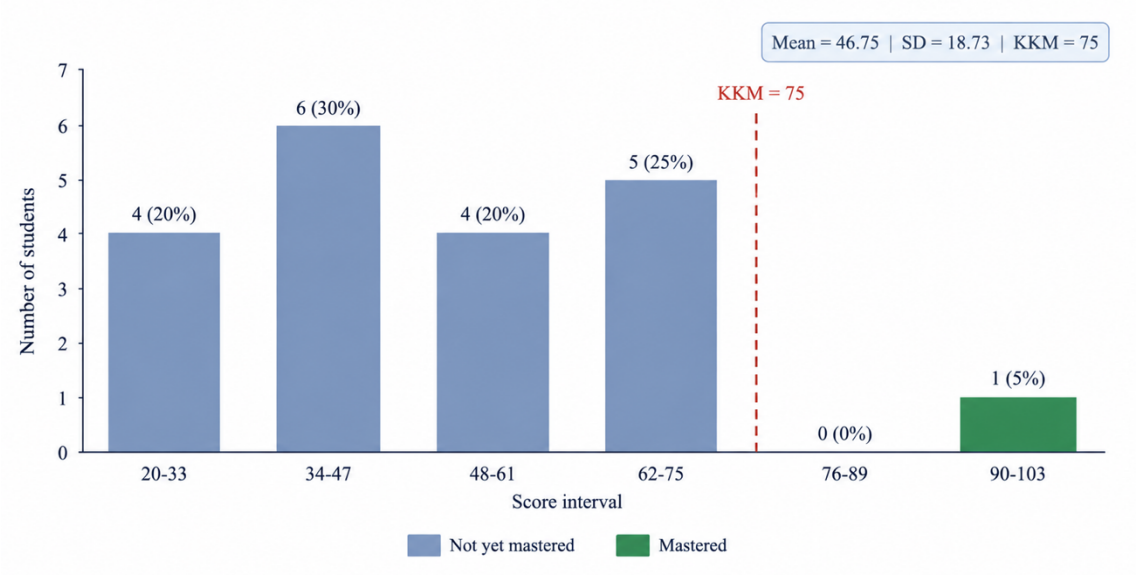



**Figure 2.** Distribution of fifth-grade IPAS post-test scores by interval ($N = 20$)

Figure 2 was included because a visual display makes the concentration of scores in the lower half of the scale visible at a glance. Readers can compare the position of the mean and the KKM without calculating anything. The benefit is faster comprehension for readers who do not read tables closely. The figure also supports the later discussion, because the long gap between the bars and the KKM line illustrates the size of the shortfall.

**One-Sample *t*-Test against the KKM**

Table 3 reports the one-sample *t*-test comparing the class mean with the KKM. The class mean was 46.75 ($SD$ = 18.73, $SE$ = 4.19), which is 28.25 points below the KKM of 75. The calculated $t$ value (-6.75, $df$ = 19) lies below the one-tailed critical value (+1.729), so $H_0$ was not rejected and the data provided no evidence that the class mean exceeded the KKM. The 95% confidence interval of the mean [37.99, 55.51] lies entirely below 75, the two-tailed probability was below .001, and Cohen's $d$ was -1.51.

**Table 3.** One-sample *t*-test of post-test scores against the KKM

| Parameter | Value |
|---|---|
| $N$ | 20 |
| Mean ($M$) | 46.75 |
| Standard deviation ($SD$) | 18.73 |
| Standard error ($SE$) | 4.19 |
| KKM ($\mu_0$) | 75 |
| $t$ | -6.75 |
| $df$ | 19 |
| Critical $t$ (one-tailed, $\alpha = .05$) | 1.729 |

| | |
|---|---|
| *p* (two-tailed) | < .001 |
| 95% CI of the mean | [37.99, 55.51] |
| Cohen's *d* | -1.51 |
| Decision | $H_0$ not rejected |

Table 3 was included because the comparison with a fixed criterion needs transparent statistics. Readers can verify the calculation from the mean, standard deviation, and sample size, and they can see the size of the difference through the confidence interval and *d*. The benefit is that the conclusion about mastery rests on inferential evidence in addition to percentages. It also allows other researchers to replicate or compare the analysis in similar classes.

**Performance by Item and Cognitive Level**

Table 4 presents the percentage of students who answered each essay item correctly. Correct response rates ranged from 7% to 28%, and all rates fall in the very low category. The C4 item on environmental problems and their impact on living things showed the highest rate (28%), followed by the C2 item (27%), while the C4 mountain-community item and the C5 item each showed 19%. The C6 item, which asked students to propose solutions to problems encountered in science learning, showed the lowest rate (7%).

**Table 4.** Percentage of students answering each essay item correctly ($N = 20$)

| Item | Cognitive level | Focus of the item | Correct (%) | Category |
|---|---|---|---|---|
| Item 1 | C2 (understanding) | Explaining and giving examples | 27 | Very low |
| Item 2 | C4 (analysing) | Environmental problems in mountain communities | 19 | Very low |
| C5 item | C5 (evaluating) | Evaluating a science-related condition | 19 | Very low |
| Item 4 | C4 (analysing) | Environmental problems in the surrounding community and their impact on living things | 28 | Very low |
| Item 5 | C6 (creating) | Solutions to problems faced in science learning | 7 | Very low |

Table 4 was included because the total score hides which cognitive levels were most difficult. It shows readers that low performance appeared on every item, including the C2 item, and that creating solutions was the weakest area. The benefit is a diagnostic profile that teachers can use to target instruction at specific levels. It also supplies the

evidence for the discussion of how comprehension and higher-order processes relate in this class.

Figure 3 shows the item-level results of Table 4 in the chart taken from the original analysis. Each segment represents one essay item with its correct response percentage as labelled in the figure. The chart allows the reader to compare the five items visually. The figure appears below.

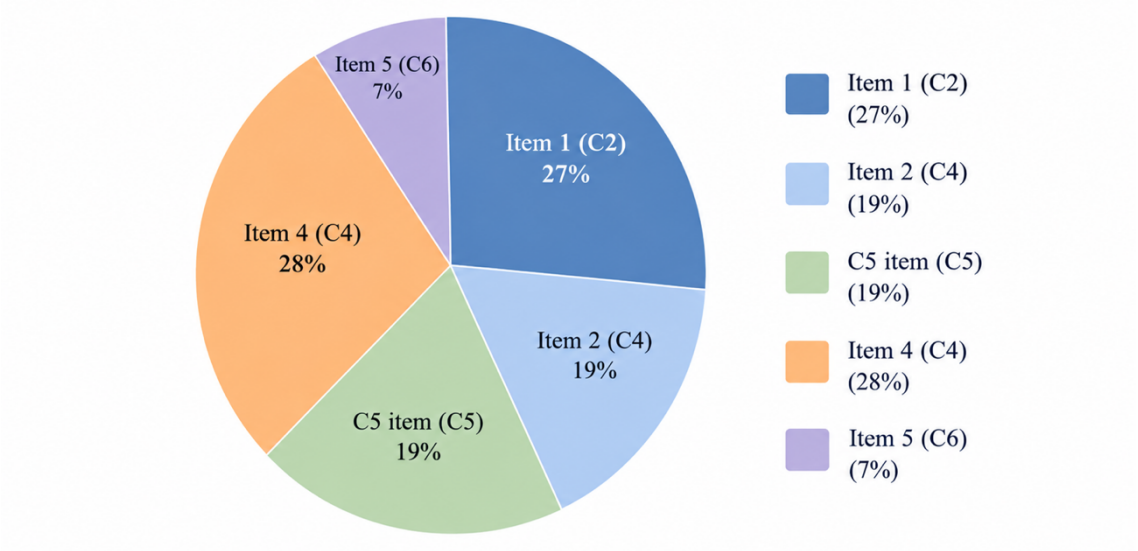



**Figure 3.** Percentage of students answering each HOTS essay item correctly

Figure 3 was included because it presents the item comparison in a form that is quick to read. Readers can see immediately that the C6 segment is the smallest and that the C2 and C4 segments are similar in size. The benefit is that the pattern of weak creating performance and uniformly low rates is grasped without reading each number. The figure also complements Table 4 for readers who prefer visual summaries.

**Teacher Interview**

The semi-structured interview with the classroom teacher provided additional information about students' learning characteristics and the factors that the teacher considered relevant to their HOTS performance in IPAS. The teacher described the students as heterogeneous in terms of academic ability, literacy skills, learning readiness, family background, physical conditions, and socio-emotional characteristics. According to the teacher, these differences influenced students' participation in learning activities and their ability to respond to assessment tasks. The interview findings were organised into three themes: students' literacy and assessment experiences, family circumstances

and individual learning needs, and the challenges of accommodating student diversity through classroom instruction.

**Students' literacy and assessment experiences.** The teacher reported that students generally demonstrated limited HOTS and literacy skills. Many students tended to provide very brief answers to essay questions, while some experienced difficulty understanding the questions they read. The teacher perceived that students had not yet developed sufficient proficiency in analysing information, thinking critically, and evaluating situations. These observations provide context for the quantitative findings, which showed that correct response rates across the five essay items ranged from 7% to 28%. The low percentage on the C2 item (27%), which required students to explain concepts and provide examples, suggests that difficulties were not limited to the higher cognitive levels.

The teacher also reported that students appeared more enthusiastic and comfortable when answering multiple-choice questions than when responding to essay questions. This difference may reflect students' familiarity with particular assessment formats or the additional demands involved in expressing ideas in writing. However, the interview did not establish whether students performed better on multiple-choice questions assessing the same content and cognitive processes. Therefore, the teacher's account is interpreted as a contextual observation rather than evidence that students possessed stronger understanding when answering multiple-choice items.

The teacher's account is particularly relevant to the C6 item, which required students to propose solutions to problems encountered in science learning. This item recorded the lowest correct response rate, at 7%. Formulating a solution requires students to understand the problem, identify relevant information, and develop an appropriate response. The combination of limited literacy skills, difficulties understanding questions, and brief written answers may help contextualise this result. Nevertheless, the interview findings do not establish a direct causal relationship between literacy difficulties and performance on the C6 item.

**Family circumstances and individual learning needs.** The teacher described differences in students' family circumstances and the learning support available at home. Several students reportedly lived with their grandparents because their parents worked outside the area. The teacher associated these circumstances with limited parental

attention and guidance in students' everyday learning. This account suggests that the teacher considered the home learning environment relevant to understanding differences in students' educational experiences. However, the study did not directly measure parental involvement or the quality of learning support provided by other family members, so the reported circumstances cannot be interpreted as evidence that particular family arrangements caused low achievement.

The teacher also described variation in students' learning readiness, physical conditions, cognitive development, surrounding environments, and socio-emotional characteristics. According to the teacher, most students had no apparent physical limitations or disabilities, while some required additional support because of difficulties in learning. The teacher further referred to below-average IQ scores among some students. These statements represent the teacher's reported perceptions and should not be treated as independently verified diagnoses. Without supporting assessment documentation, the findings are best interpreted as indicating that the teacher perceived differences in students' cognitive and learning support needs.

These reported differences are important when considering students' performance on assessment tasks with varying cognitive demands. Students who experience difficulty understanding written instructions may require additional support before they can demonstrate their reasoning through an essay response. Similarly, students with different levels of prior knowledge may require different amounts of guidance when analysing problems or proposing solutions. These possibilities highlight the relevance of examining students' individual learning needs, although the present study did not analyse individual student profiles in relation to their test scores.

**Instructional practices and heterogeneous learning needs.** The teacher acknowledged that the teaching methods used in the classroom had not fully accommodated the diversity of students' learning characteristics. This observation highlights a perceived challenge in providing learning experiences that respond to differences in academic ability, literacy, readiness, and socio-emotional circumstances. It also provides an important contextual consideration for interpreting the post-test results, in which only one student reached the Kriteria Ketuntasan Minimal (KKM) of 75 and the class mean was 46.75.

The teacher's account suggests that students may need structured opportunities to develop their understanding before being expected to complete more complex analytical, evaluative, and creative tasks independently. In the context of discovery learning, possible forms of support include clarifying task instructions, activating prior knowledge, providing guiding questions, modelling how to analyse information, and offering feedback as students develop their answers. Regular practice with essay questions may also help students become more familiar with explaining concepts, evaluating situations, and constructing solutions. These are potential instructional implications arising from the reported classroom circumstances, rather than strategies whose effectiveness was tested in the present study.

The interview findings should also be interpreted in light of the study design. The research examined students' post-test performance after discovery learning had been implemented but did not include a pre-test, a comparison group, or a systematic assessment of implementation fidelity. Consequently, the teacher's observations cannot establish that discovery learning caused the low HOTS outcomes or that a particular instructional limitation was responsible for students' performance. Instead, the interview provides contextual information about the classroom and identifies factors that warrant further investigation.

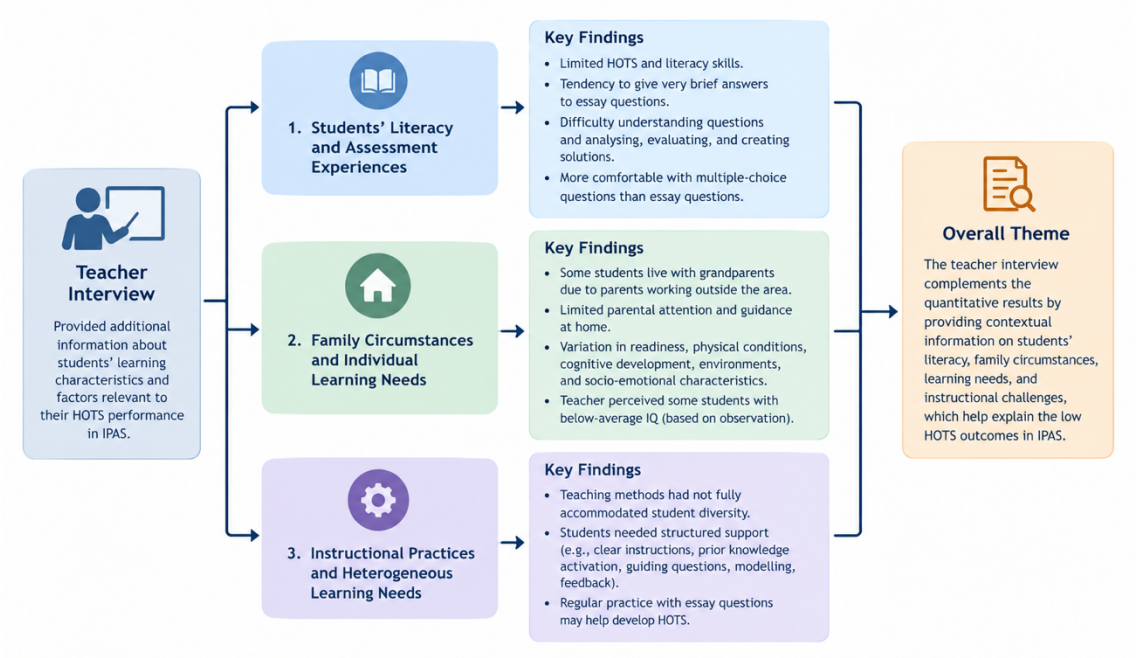



**Figure 4.** Thematic Concept Map of the Teacher Interview Findings

Overall, the teacher interview complements the quantitative results by describing perceived difficulties in literacy, responses to essay-based assessment, differences in family circumstances, and heterogeneous learning needs. These findings help contextualise the low correct response rates across cognitive levels and the particularly weak performance on the C6 item. Although the interview does not establish causal relationships, it points to the importance of examining students' understanding of written tasks, their opportunities to practise HOTS-oriented questions, and the guidance available to support diverse learners. Future research could investigate these factors using classroom observations, student interviews, documented assessment records, and pre-test and post-test comparisons.

## DISCUSSION

### Learning Outcomes after the Implementation of Discovery Learning

The main finding is that the class mean (46.75) was reliably below the KKM (d = -1.51) and that only one of 20 students reached mastery. Because the study has no pretest or comparison group, this pattern shows where students stood after discovery learning was implemented and cannot show what discovery learning produced. Studies that compared guidance conditions in science classrooms offer a reference point, because van Riesen et al. (2018) reported that learners with low prior knowledge showed a higher learning gain when they worked with a scaffolding tool than in two control conditions. If the discovery activities in this classroom provided limited guidance, the pattern would be consistent with the idea that students with low prior knowledge benefit from structured support, although implementation data were not collected to confirm this possibility.

Different patterns have been reported in other contexts, because Mahanal et al. (2019) found that a structured learning model improved critical thinking skills for students with different academic abilities. Wijesekera and Hameed (2025) reported that collaborative questioning strategies in a secondary science classroom were accompanied by greater curiosity and willingness to engage with complex scientific concepts, although the setting and the school level differ from the present class. The findings should be interpreted in relation to differences in educational level, comparison groups, duration, and the structure of learning materials, which suggests that the label discovery learning covers practices with very different levels of support. The practical implication is that schools adopting discovery learning could document the guidance provided in each

lesson, and future evaluations could add a pretest and a comparison class so that outcomes can be attributed to instruction with more confidence.

**Performance across Cognitive Levels**

Correct response rates were low on every item and lowest on the C6 item (7%), while the C2 item (27%) and the C4 item (28%) showed similar rates. This pattern suggests that difficulties were present at the level of understanding as well as at higher levels, which agrees with the teacher's report that students sometimes had difficulty understanding the questions they read. O'Reilly et al. (2019) reported that comprehension was impeded when readers' background knowledge fell below a threshold, which suggests that limited knowledge can constrain the understanding of written tasks. Retnawati et al. (2017) examined why mathematics national examination items were difficult for students and what strategies teachers used to overcome it, and although that study concerns mathematics, it shows that item difficulty can be examined together with teachers' instructional strategies, so question comprehension and classroom practice deserve joint attention in IPAS lessons.

The lowest rate on the C6 item is consistent with the greater cognitive demand of creating, which draws on the processes that precede it. Dong et al. (2020) examined how prior knowledge influences learning engagement through cognitive load and help-seeking, which suggests that students with limited prior knowledge may find open-ended creating tasks heavier to handle. The teacher also reported that students answered multiple-choice questions with ease, and Greving and Richter (2018) reported that question format was relevant to the benefit that students obtained from practice questions. Because that study was conducted in a university course and no multiple-choice comparison was included here, familiarity with the question format is one possible contributor among several, and C4 to C6 essay items with feedback could be introduced regularly in formative assessment.

**Factors Reported by the Teacher**

The teacher referred to limited literacy, unfamiliarity with essay questions, limited parental guidance for several students, cognitive delays in some students, and diverse learning needs that instruction had not fully accommodated. These statements come from a single informant and describe the teacher's perspective, so they do not establish causal relationships with the test scores. In a sample of 3,076 children in rural China, Fu et al.

(2017) reported that left-behind children were more likely to obtain low scores in academic examinations than non-left-behind children. The report that several students lived with grandparents while parents worked out of town is consistent with reduced supervision for some students, and Li et al. (2024) found that parental educational involvement partly mediated the link between parents' expectations and left-behind children's achievement, but parental involvement was not measured here, so this connection is a hypothesis for later investigation.

Regarding cognitive delays and diverse learning needs, Suprayogi et al. (2017) examined how teachers implement differentiated instruction in the classroom, and Pozas et al. (2020) examined teachers' differentiation practices for addressing student diversity. The implication is that differentiated support can be considered for this class while the cognitive demand of tasks is maintained, although this study did not test differentiation. Regarding learning styles, Rogowsky et al. (2020) reported that instruction matched to learning-style preferences did not improve learning, and Newton and Miah (2017) examined the persistence of learning-styles beliefs in higher education. Differentiation in this classroom might therefore be directed toward students' prior knowledge and the type of support needed for each task, and the usefulness of style-based matching should be examined with caution.

**Guidance, Prior Knowledge, and Implications for Discovery Learning in IPAS**

The teacher's statement that methods had not fully accommodated diverse needs points toward guidance that is adjusted to students. Prior knowledge has been examined as a predictor of achievement in introductory biology and physics programmes (Binder et al., 2019), which supports diagnostic attention to what students already know. Contingent support, gradual fading, and transfer of responsibility to learners are common features of scaffolding in classroom research, and these features can be observed in discovery lessons. These characteristics offer a practical vocabulary for observing whether discovery lessons adjust support to students and gradually release responsibility.

Prior knowledge also shapes how much guidance learners need, because in the study by van Riesen et al. (2018) the gain from the scaffolding tool was larger for learners who started with low prior knowledge. This evidence agrees with the expectation that students with more knowledge need less support than novices, and Dong et al. (2020) linked prior knowledge to engagement through cognitive load and help-seeking. For fifth-

grade students whose literacy was described as low, structured prompts and gradual release of responsibility could be expected to fit well, and this expectation requires testing in this setting. Studies of teachers' differentiation practices (Pozas et al., 2020; Suprayogi et al., 2017) show that teacher practice is a relevant object of study, but teacher knowledge was not measured here, so professional learning on guided discovery and HOTS item design is a direction for future evaluation.

**Limitations and Future Research**

Several limitations qualify the findings. The sample comprised one class of 20 students, which limits generalisation. The design had no pretest or comparison group and implementation fidelity was not recorded, so the findings cannot be attributed to discovery learning. The teacher interview involved one informant, and classroom observations and student perspectives were not included.

The psychometric properties of the test were not examined in the analysis reported here, and the normality assumption of the t-test was not checked with this small sample. Future studies could use pre-post designs with comparison classes and observe the level of guidance in discovery lessons. They could also measure teachers' knowledge of HOTS and compare essay and multiple-choice formats for the same content. Mixed-methods work could then examine how literacy, family support, and prior knowledge relate to HOTS performance.

## CONCLUSION

This study described the HOTS outcomes of 20 fifth-grade students in IPAS after discovery learning was implemented. The class mean (46.75) was below the KKM of 75, 95% of students did not reach mastery, and the mean percentage falls in the very low category. Correct response rates ranged from 7% on the C6 item to 28% on a C4 item, and the C2 item also showed a low rate (27%). The teacher referred to limited literacy, unfamiliarity with essay questions, limited parental guidance for several students, and heterogeneous learning needs that instruction had not fully accommodated.

Because the design has no comparison group or pretest, the findings describe outcomes after implementation and leave the effect of discovery learning to comparative research. The results suggest that teachers and schools could examine the level of guidance in discovery lessons and assess students' prior knowledge. They also suggest that HOTS essay items with feedback could be used regularly in formative assessment.

Comparative and mixed-methods studies are needed to test these suggestions in elementary IPAS classrooms.

**AI Use Statement**

AI-assisted tools were used, if applicable, for language refinement and proofreading. The authors reviewed the manuscript and take responsibility for its content.

**Author Contributions**

E: coordinated the study and prepared the initial manuscript. LP; H: contributed to data collection and analysis. IS: contributed to research methodology and interpretation of findings. MZS: supervised the research and revised the manuscript. MDR; INI; FS: contributed to data interpretation and critical manuscript review. All authors approved the final manuscript.

**Funding Statement**

This study received no external funding.

**Conflict of Interest**

The authors declare no competing interests.

## REFERENCES

Aldalur, I., & Perez, A. (2023). Gamification and discovery learning: Motivating and involving students in the learning process. *Heliyon, 9*(1), Article e13135. https://doi.org/10.1016/j.heliyon.2023.e13135

Anazifa, R. D., & Djukri. (2017). Project-based learning and problem-based learning: Are they effective to improve student's thinking skills? *Jurnal Pendidikan IPA Indonesia, 6*(2), 346-355. https://doi.org/10.15294/jpii.v6i2.11100

Binder, T., Sandmann, A., Sures, B., Friege, G., Theyssen, H., & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. *International Journal of STEM Education, 6*, Article 33. https://doi.org/10.1186/s40594-019-0189-9

Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology, 11*, Article 591203. https://doi.org/10.3389/fpsyg.2020.591203

Fu, M., Bo, W. V., Xue, Y., & Yuan, T.-F. (2017). Parental absence accompanies worse academic achievements: Evidence based upon a sample of left-behind children in rural China. *Frontiers in Education, 2*, Article 38. https://doi.org/10.3389/feduc.2017.00038

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology, 9*, Article 2412. https://doi.org/10.3389/fpsyg.2018.02412

Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century, 76*(4), 520-532. https://doi.org/10.33225/pec/18.76.520

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research, 8*(4), 935-942. https://doi.org/10.12973/eu-jer.8.4.935

Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education. *International Journal of Science Education, 40*(4), 421-441. https://doi.org/10.1080/09500693.2018.1428777

Li, J., Xue, E., & You, H. (2024). Parental educational expectations and academic achievement of left-behind children in China: The mediating role of parental involvement. *Behavioral Sciences, 14*(5), Article 371. https://doi.org/10.3390/bs14050371

Lu, K., Yang, H. H., Shi, Y., & Wang, X. (2021). Examining the key influencing factors on college students' higher-order thinking skills in the smart classroom environment. *International Journal of Educational Technology in Higher Education, 18*, Article 1. https://doi.org/10.1186/s41239-020-00238-7

Mahanal, S., Zubaidah, S., Sumiati, I. D., Sari, T. M., & Ismirawati, N. (2019). RICOSRE: A learning model to develop critical thinking skills for students with different academic abilities. *International Journal of Instruction, 12*(2), 417-434. https://doi.org/10.29333/iji.2019.12227a

Nancekivell, S. E., Shah, P., & Gelman, S. A. (2020). Maybe they're born with it, or maybe it's experience: Toward a deeper understanding of the learning style myth. *Journal of Educational Psychology, 112*(2), 221-235. https://doi.org/10.1037/edu0000366

Newton, P. M., & Miah, M. (2017). Evidence-based higher education: Is the learning styles "myth" important? *Frontiers in Psychology, 8*, Article 444. https://doi.org/10.3389/fpsyg.2017.00444

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science, 30*(9), 1344-1351. https://doi.org/10.1177/0956797619862276

Pozas, M., Letzel, V., & Schneider, C. (2020). Teachers and differentiated instruction: Exploring differentiation practices to address student diversity. *Journal of Research in Special Educational Needs, 20*(3), 217-230. https://doi.org/10.1111/1471-3802.12481

Ramdiah, S., Abidinsyah, Royani, M., Husamah, & Fauzi, A. (2019). Understanding, planning, and implementation of HOTS by senior high school biology teachers in Banjarmasin-Indonesia. *International Journal of Instruction, 12*(1), 425-440. https://doi.org/10.29333/iji.2019.12128a

Retnawati, H., Kartowagiran, B., Arlinwibowo, J., & Sulistyaningsih, E. (2017). Why are the mathematics national examination items difficult and what is teachers' strategy to overcome it? *International Journal of Instruction, 10*(3), 257-276. https://doi.org/10.12973/iji.2017.10317a

Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century, 76*(2), 215-230. https://doi.org/10.33225/pec/18.76.215

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2020). Providing instruction based on students' learning style preferences does not improve learning. *Frontiers in Psychology, 11*, Article 164. https://doi.org/10.3389/fpsyg.2020.00164

Shieh, C.-J., & Yu, L. (2016). A study on information technology integrated guided discovery instruction towards students' learning achievement and learning retention. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(4), 833-842. https://doi.org/10.12973/eurasia.2015.1554a

Sidiq, Y., Ishartono, N., Desstya, A., Prayitno, H. J., Anif, S., & Hidayat, M. L. (2021). Improving elementary school students' critical thinking skill in science through HOTS-based science questions: A quasi-experimental study. *Jurnal Pendidikan IPA Indonesia, 10*(3), 378-386. https://doi.org/10.15294/jpii.v10i3.30891

Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in the classroom. *Teaching and Teacher Education, 67*, 291-301. https://doi.org/10.1016/j.tate.2017.06.020

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education, 15*(4), 663-682. https://doi.org/10.1007/s10763-016-9723-0

van Riesen, S. A. N., Gijlers, H., Anjewierden, A., & de Jong, T. (2018). Supporting learners' experiment design. *Educational Technology Research and Development, 66*(2), 475-491. https://doi.org/10.1007/s11423-017-9568-4

Vieira, R. M., & Tenreiro-Vieira, C. (2016). Fostering scientific literacy and critical thinking in elementary science education. *International Journal of Science and Mathematics Education, 14*(4), 659-680. https://doi.org/10.1007/s10763-014-9605-2

Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology, 52*, 36-47. https://doi.org/10.1016/j.cedpsych.2017.12.003

Wijesekera, H. D., & Hameed, R. (2025). From rote learning to critical inquiry: Fostering higher order thinking skills through collaborative questioning in a rural secondary English-medium science classroom. *Thinking Skills and Creativity, 60*, Article 102093. https://doi.org/10.1016/j.tsc.2025.102093

Yang, K.-K., Lee, L., Hong, Z.-R., & Lin, H.-S. (2016). Investigation of effective strategies for developing creative science thinking. *International Journal of Science Education, 38*(13), 2133-2151. https://doi.org/10.1080/09500693.2016.1230685

Yerimadesi, Y., Warlinda, Y. A., Rosanna, D. L., Sakinah, M., Putri, E. J., Guspatni, G., & Andromeda, A. (2023). Guided discovery learning-based chemistry e-module and its effect on students' higher-order thinking skills. *Jurnal Pendidikan IPA Indonesia, 12*(1), 168-177. https://doi.org/10.15294/jpii.v12i1.42130