PEDAGOGIKA
Volume 17 (Nomor 01) 2026
HaL. 92-111

Science Learning Outcomes of Junior High School Students on The Topic of Global Warming : A Descriptive Study with Rasch-Logit Item Analysis

Rifda Nur Hikmahwati Arif, Aulia Aura Yusmaputri, Hasanuddin
Department of Science Education, Universitas Negeri Makassar, Indonesia
Email: rifdanha@unm.ac.id

ABSTRAK

Penelitian ini bertujuan mendeskripsikan hasil belajar IPA peserta didik kelas IX pada materi pemanasan global di Kabupaten Luwu Utara melalui statistik deskriptif dan analisis item berbasis logit Rasch. Penelitian menggunakan desain kuantitatif deskriptif dengan melibatkan 378 peserta didik dari empat SMP Negeri: SMPN 1 Sabbang, SMPN 1 Masamba, SMPN 3 Sabbang Selatan, dan SMPN 4 Masamba (Sugiyono, 2018). Instrumen penelitian berupa 20 soal pilihan ganda yang disusun berdasarkan level kognitif C1–C4 taksonomi Bloom revisi (Anderson & Krathwohl, 2015). Hasil hitung ulang menunjukkan 5.583 jawaban benar dengan rata-rata 14,77 dari skor maksimum 20, mengonfirmasi rata-rata skripsi sebesar 14,76. Hasil belajar berada pada kategori tinggi (Arikunto, 2018). SMPN 3 Sabbang Selatan memperoleh rata-rata tertinggi (17,20; sangat tinggi), sedangkan SMPN 1 Sabbang memperoleh rata-rata terendah (13,85; tinggi). Koefisien KR-20 sebesar 0,674 menunjukkan konsistensi internal sedang. Analisis tingkat kesukaran menunjukkan 13 item mudah, 7 item sedang, dan tidak ada item sulit. Item tersulit adalah Item 18 (p = 0,415; logit = +0,342), sedangkan item termudah adalah Item 11 (p = 0,902; logit = −2,221). Visualisasi Wright Map menunjukkan perlunya penambahan item yang lebih menantang, terutama pada level C2–C4 (Bond & Fox, 2015; Boone et al., 2014).

Kata Kunci: hasil belajar IPA; pemanasan global; logit Rasch; Wright Map; sekolah menengah pertama

ABSTRACT

This study aimed to describe the science learning outcomes of Grade IX junior secondary students on the topic of global warming in North Luwu Regency through descriptive statistics and Rasch-logit item analysis. A quantitative descriptive design was employed involving 378 students from four public junior secondary schools (SMPN 1 Sabbang, SMPN 1 Masamba, SMPN 3 Sabbang Selatan, and SMPN 4 Masamba). The instrument consisted of 20 multiple-choice items aligned with cognitive levels C1–C4 of the revised Bloom's taxonomy (Anderson & Krathwohl, 2015). Recalculation showed 5,583 correct responses, with an overall mean score of 14.77 out of 20, confirming the reported thesis mean of 14.76. Students' achievement was categorised as high (Arikunto, 2018). SMPN 3 Sabbang Selatan obtained the highest mean score (17.20; very high category), whereas SMPN 1 Sabbang obtained the lowest mean score (13.85; high category). The KR-20 coefficient was 0.674, indicating moderate internal consistency. Item difficulty analysis showed that 13 items were easy, 7 items were moderate, and no item was difficult. The most difficult item was Item 18 (p = 0.415; logit = +0.342), while the easiest was Item 11 (p = 0.902;

logit = −2.221). Wright Map visualisation indicates that the instrument needs more challenging items, particularly at C2–C4 levels (Bond & Fox, 2015; Boone et al., 2014).

Keywords: *science learning outcomes; global warming; Rasch logit; Wright Map; junior secondary school*

INTRODUCTION

Science learning outcomes are central indicators of the effectiveness of science instruction because they reflect the extent to which students master concepts, interpret scientific information, and apply knowledge to contextual problems (Annisa & Habibi, 2020; Arif et al., 2024). In junior secondary science education, learning outcomes are not merely intended to represent students' ability to recall facts, but also their capacity to understand relationships among concepts, apply scientific principles, and analyse phenomena related to daily life (Anderson & Krathwohl, 2015; Sufinasa et al., 2023). According to Arikunto (2018), achievement assessment is the primary mechanism through which teachers can identify the degree of instructional effectiveness and areas requiring improvement.

The topic of global warming is especially relevant because it connects scientific concepts with environmental problems that are increasingly visible in local communities. Global warming involves the interaction of greenhouse gases, human activities, climate change, and environmental impacts (Wildan et al., 2020). Students' mastery of this topic is important not only for academic achievement but also for strengthening evidence-based environmental awareness (Hifjir & Agustizar, 2020). In North Luwu Regency, the region has experienced severe hydrometeorological disasters directly linked to climate change, making this topic particularly meaningful for local science education (Prabawati, 2021).

Many classroom-based studies report students' achievement using only mean scores and categories. Although useful, mean scores alone are insufficient to show which test items are too easy, moderately difficult, or difficult, and which cognitive levels require improvement (Nafiati, 2021). Item-level analysis is needed to identify the diagnostic quality of a test (Azizah et al., 2021). A Rasch-logit transformation of item difficulty can provide an additional measurement perspective by expressing item difficulty on a logit scale, where positive values indicate more difficult items and negative values indicate easier items (Bond & Fox, 2015; Boone et al., 2014). This approach enables teachers and researchers to go beyond mean scores and examine the measurement properties of the instrument itself (Creswell & Creswell, 2018).

Contemporary science education frameworks also emphasise that science instruction should develop students' capacity to argue, evaluate evidence, and construct explanations — competencies that extend well beyond factual recall (Osborne, 2010; National Research Council, 2012). In this sense, describing students' learning outcomes at the item and cognitive-level is essential to identify where these higher-order competencies are or are not being developed. Based on this background, the research questions are: (1) What is the overall level of Grade IX students' science learning outcomes on global warming in North Luwu Regency? (2) How do learning outcomes differ across schools and cognitive levels? (3) What does item difficulty

analysis using proportion correct, Rasch-logit transformation, and Wright Map visualisation reveal about the quality of the learning outcome instrument?

METHOD

This study employed a quantitative descriptive design to describe students' cognitive achievement profile on global warming without experimental manipulation (Creswell & Creswell, 2018). The participants were 378 Grade IX students from four public junior secondary schools in North Luwu Regency: SMPN 1 Sabbang, SMPN 1 Masamba, SMPN 3 Sabbang Selatan, and SMPN 4 Masamba. The sample was selected using purposive sampling based on school accessibility, data completeness, and adequacy of student numbers (Sugiyono, 2018). Table 1 presents the sample distribution.

Table 1. Sample distribution by school

School	N	Proportion (%)
SMPN 1 Sabbang	78	20.60
SMPN 1 Masamba	128	33.90
SMPN 3 Sabbang Selatan	39	10.30
SMPN 4 Masamba	133	35.20

The instrument consisted of 20 multiple-choice items covering global warming causes, mechanisms, impacts, and mitigation strategies. Each item was scored dichotomously (1 = correct, 0 = incorrect; maximum score = 20). Items were distributed across cognitive levels C1–C4 of the revised Bloom's taxonomy (Anderson & Krathwohl, 2015) as shown in Table 2.

Table 2. Item distribution by cognitive level

Cognitive Level	Description	Number of Items	Item Numbers
C1	Remembering	2	1, 2
C2	Understanding	8	3, 4, 5, 8, 9, 11, 12, 18
C3	Applying	7	7, 10, 13, 14, 15, 17, 19
C4	Analysing	3	6, 16, 20
Total	–	20	–

The data were analysed using descriptive statistics, including total correct responses, mean score, standard deviation, variance, and learning outcome category. Students' achievement was classified using Arikunto's (2018) criteria, as presented in Table 3. The internal consistency of the dichotomously scored test items was estimated using KR-20. Item difficulty was calculated based on the proportion of correct responses (p), where higher values indicate easier items and lower values indicate more difficult items. To support Rasch-based interpretation, the proportion correct was transformed into Rasch-logit item difficulty using the formula $b = \ln[(1 - p) / p]$. Positive logit values indicate more difficult items, while negative values indicate easier items (Bond & Fox, 2015). Wright Map visualisations were then produced to show the alignment between students' estimated ability and item difficulty.

Table 3. Learning outcome categories

Score Range	Category
17–20	Very High
13–16	High
9–12	Moderate
5–8	Low
0–4	Very Low

Source: Arikunto (2018)

RESULTS AND DISCUSSION

Result

The overall descriptive analysis was conducted to provide a general overview of students' science learning outcomes on the topic of global warming. The analysis involved 378 Grade IX junior secondary students from four public schools in North Luwu Regency. Students' responses were scored based on 20 multiple-choice items, with each correct answer assigned a score of 1 and each incorrect answer assigned a score of 0. Therefore, the maximum possible score was 20.

The recalculation showed that the total number of correct responses across all students and all items was 5,583, producing a mean score of 14.77 out of 20. This result is almost identical to the mean score reported in the original thesis, namely 14.76, indicating that the recalculated data are consistent with the previous report. Based on the learning outcome categorization, this mean score falls into the high category. The standard deviation of 3.07 and variance of 9.47 indicate that students' scores were moderately dispersed, meaning that although most students achieved good learning outcomes, there were still meaningful differences in mastery levels among students.

The reliability estimate using KR-20 produced a coefficient of 0.674, which indicates that the 20-item test had a moderate level of internal consistency (Haw et al., 2022). This suggests that the instrument was sufficiently acceptable for describing students' science learning outcomes, although further item refinement may be needed to strengthen its measurement quality. The complete recalculated descriptive statistics are presented in Table 4.

Overall Learning Outcomes

Table 4 presents recalculated descriptive statistics for overall science learning outcomes.

Table 4. Recalculated descriptive statistics of science learning outcomes

Statistic	Value
N	378
Highest Score	20
Lowest Score	6
Ideal Maximum	20

Statistic	Value
Ideal Minimum	0
Total Correct Responses	5,583
Mean Score	14.77
Reported Mean (thesis)	14.76
Standard Deviation	3.07
Variance	9.47
KR-20	0.674
Category	High

The recalculation produced 5,583 correct responses across 20 items and 378 students. The mean score of 14.77 is consistent with the reported thesis mean of 14.76 (Arikunto, 2018), confirming that the overall science learning outcome was in the high category. The standard deviation of 3.07 shows moderate dispersion, indicating that students generally performed well but with meaningful variation in mastery levels.

Learning Outcomes by School

Table 5. Descriptive statistics by school

School	N	Highest	Lowest	Mean	Mean (%)	SD	Variance	Category
SMPN 1 Sabbang	78	20	7	13.85	67.40	3.25	10.59	High
SMPN 1 Masamba	128	19	6	14.17	70.70	3.08	9.52	High
SMPN 3 Sabbang Selatan	39	20	11	17.20	85.45	2.26	5.11	Very High
SMPN 4 Masamba	133	19	7	15.16	76.25	2.73	7.50	High

Table 5 shows differences in science learning outcomes among the four schools. The highest mean score was obtained by SMPN 3 Sabbang Selatan, with a mean score of 17.20 or 85.45%, placing it in the very high category. This school also had the smallest standard deviation, namely 2.26, indicating that students' scores were relatively more homogeneous compared with the other schools.

Meanwhile, SMPN 4 Masamba achieved a mean score of 15.16 or 76.25%, followed by SMPN 1 Masamba with a mean score of 14.17 or 70.70%. Both schools were categorized as high. The lowest mean score was found in SMPN 1 Sabbang, with a mean score of 13.85 or 67.40%, although it still remained in the high category. This result indicates that students from all schools generally demonstrated good learning outcomes on the topic of global warming.

The variation in mean scores suggests that although the overall achievement was high, there were differences in students' mastery levels across schools. SMPN 3 Sabbang Selatan showed the strongest performance, while SMPN 1 Sabbang showed the lowest average achievement and the largest variation in scores. To provide a clearer comparison of the mean learning outcomes among schools, the data are visualized in Figure 1.

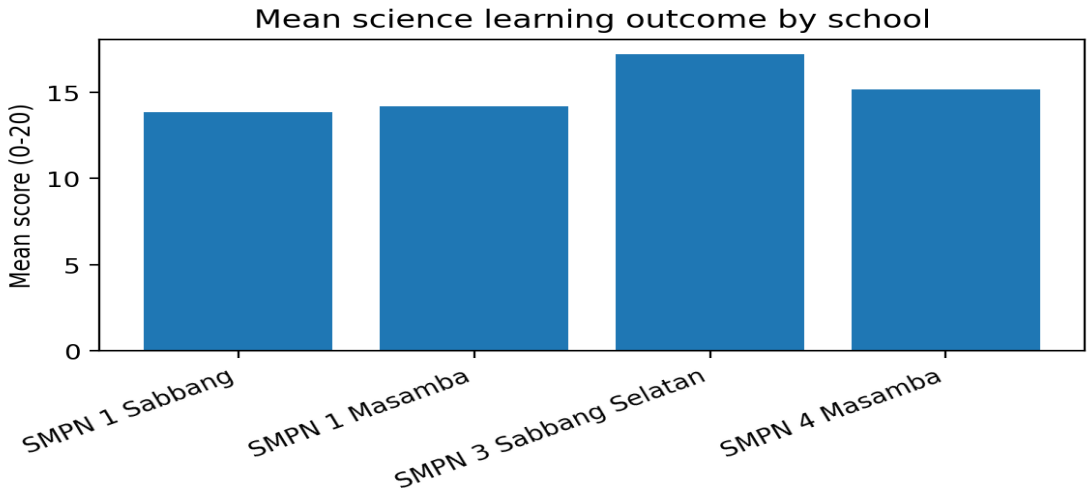

Figure 1. Diagram of mean science learning outcome by school

School-level analysis showed that SMPN 3 Sabbang Selatan obtained the highest mean score (17.20; very high category), followed by SMPN 4 Masamba (15.16), SMPN 1 Masamba (14.17), and SMPN 1 Sabbang (13.85; all high category). All four schools reached at least the high category, indicating an adequate general level of mastery. The variation across schools may reflect differences in class size, instructional quality, student readiness, and school-level academic culture, including the use of learning media that influence student motivation (Arif et al., 2023). However, because this study is descriptive, causal interpretation of school-level differences is not warranted (Creswell & Creswell, 2018). Further research is required to examine teacher practices, learning resources, and classroom climate as possible explanatory variables.

Learning Outcomes by Cognitive Level

Table 6 presents outcomes disaggregated by cognitive level.

Table 6. Learning outcomes by cognitive level

Cognitive Level	Number of Items	Total Correct	Mean Score	Mean Value (%)
C1 (Remembering)	2	630	1.67	83.33
C2 (Understanding)	8	2,048	5.42	67.72
C3 (Applying)	7	2,080	5.50	78.61
C4 (Analysing)	3	825	2.18	72.75

Table 6 presents students' science learning outcomes based on cognitive levels. The highest achievement was found at the C1 remembering level, with a mean value of 83.33%. This indicates that students were generally able to recall factual information and basic concepts related to global warming. The strong performance at this level suggests that students had good mastery of basic knowledge, such as the definition, causes, and examples of global warming phenomena.

The second-highest result was found at the C3 applying level, with a mean value of 78.61%. This shows that students were quite capable of applying concepts of global warming to familiar

situations or contextual problems. Meanwhile, the C4 analysing level obtained a mean value of 72.75%, indicating that students had a fairly good ability to identify relationships, interpret causes and effects, and analyse environmental problems related to global warming.The lowest mean value was found at the C2 understanding level, with a percentage of 67.72%. Although this value still indicates a relatively good level of achievement, it suggests that some students may still face difficulties in explaining concepts, interpreting information, or connecting scientific ideas in their own words. Overall, the findings show that students' learning outcomes were generally positive across all cognitive levels, but conceptual understanding at C2 still requires further instructional reinforcement. To provide a clearer comparison of achievement across cognitive levels, the results are presented visually in Figure 2.

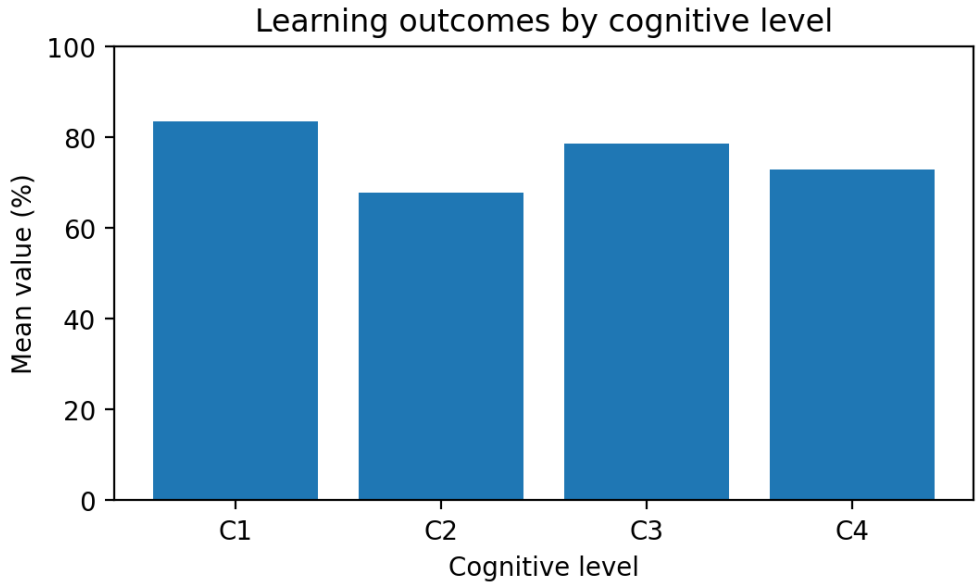

Figure 2. Diagram of science learning outcomes by cognitive level

C1 remembering obtained the highest percentage (83.33%), indicating that students were strongest in recalling facts and basic global warming concepts (Anderson & Krathwohl, 2015). C3 applying was also relatively high (78.61%), suggesting that many students could use global warming concepts in familiar contextual problems. C4 analysing obtained 72.75%. In contrast, C2 understanding had the lowest percentage (67.72%). This finding is consistent with Nafiati's (2021) observation that the understanding level is a critical bridge between lower and higher-order cognition; when C2 is not well developed, deeper analytical comprehension tends to remain constrained. The implication is that instructional strategies should explicitly target conceptual explanation and cause-effect reasoning around global warming, moving students beyond factual recall toward genuine conceptual understanding.

Item Difficulty and Rasch-Logit Analysis

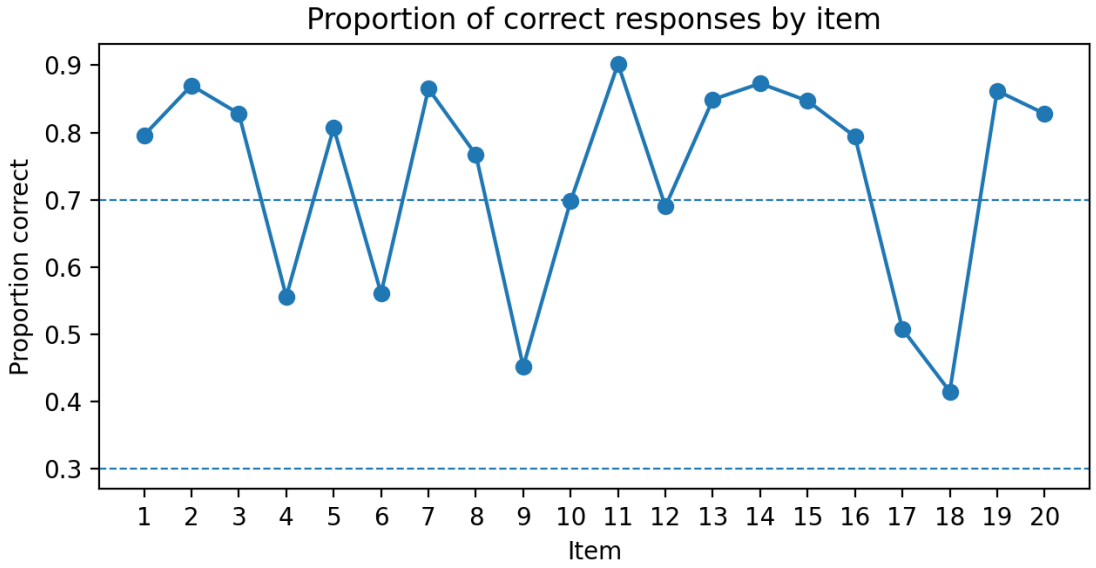

Figure 3. Diagram of item-level proportion of correct responses

Figure X shows that the proportion of correct responses across the 20 items varied substantially, indicating differences in item difficulty. The highest correct response proportion was found in Item 11, while the lowest was found in Item 18. Most items had proportions above 0.70, suggesting that they were relatively easy for students. However, Items 9, 17, and 18 showed lower proportions of correct responses, indicating that these items were more difficult and may require higher conceptual understanding or analytical reasoning. These findings suggest that although students' overall science learning outcomes on global warming were generally high, several concepts still require further instructional reinforcement, particularly those involving deeper reasoning and analysis.

Table 7. Recalculated item difficulty and Rasch-logit values (Items 1–20)

Item	Level	Cognitive Process	Correct	Incorrect	p	Logit	Category
1	C1	Remembering	301	77	0.796	−1.363	Easy
2	C1	Remembering	329	49	0.870	−1.904	Easy
3	C2	Understanding	313	65	0.828	−1.572	Easy
4	C2	Understanding	210	168	0.556	−0.223	Moderate
5	C2	Understanding	305	73	0.807	−1.430	Easy
6	C4	Analysing	212	166	0.561	−0.245	Moderate
7	C3	Applying	327	51	0.865	−1.858	Easy
8	C2	Understanding	290	88	0.767	−1.193	Easy
9	C2	Understanding	171	207	0.452	+0.191	Moderate
10	C3	Applying	264	114	0.698	−0.840	Moderate
11	C2	Understanding	341	37	0.902	−2.221	Easy

Item	Level	Cognitive Process	Correct	Incorrect	p	Logit	Category
12	C2	Understanding	261	117	0.690	−0.802	Moderate
13	C3	Applying	321	57	0.849	−1.728	Easy
14	C3	Applying	330	48	0.873	−1.928	Easy
15	C3	Applying	320	58	0.847	−1.708	Easy
16	C4	Analysing	300	78	0.794	−1.347	Easy
17	C3	Applying	192	186	0.508	−0.032	Moderate
18	C2	Understanding	157	221	0.415	+0.342	Moderate
19	C3	Applying	326	52	0.862	−1.836	Easy
20	C4	Analysing	313	65	0.828	−1.572	Easy

Item difficulty analysis showed that no item fell into the difficult category. Thirteen items were categorised as easy ($p > 0.70$) and seven as moderate ($0.30 \leq p \leq 0.70$). The most difficult item was Item 18 ($p = 0.415$; logit = +0.342), a C2 Understanding item, followed by Item 9 ($p = 0.452$; logit = +0.191) and Item 17 ($p = 0.508$; logit = −0.032). The easiest item was Item 11 ($p = 0.902$; logit = −2.221), also a C2 Understanding item.

The item difficulty analysis showed that students' performance varied across the 20 test items. Based on the proportion of correct responses, most items were answered correctly by more than half of the students, indicating that the test items were generally within students' ability range. However, several items had lower correct-response proportions and therefore required closer attention.

The Rasch-logit transformation provided a clearer interpretation of item difficulty. Items with positive logit values indicate relatively more difficult items, whereas items with negative logit values indicate easier items. The results showed that only two items had positive logit values, namely Item 18 and Item 9, suggestig that these items were the most challenging for students. In contrast, most other items had negative logit values, meaning that they were relatively easier for the students. Among the 20 items, the five most difficult items were identified based on the lowest proportion of correct responses. These items are presented in Table 8.

Table 8. Five most difficult items

Item	Level	Cognitive Process	p	Logit	Category
18	C2	Understanding	0.415	+0.342	Moderate
9	C2	Understanding	0.452	+0.191	Moderate
17	C3	Applying	0.508	−0.032	Moderate
4	C2	Understanding	0.556	−0.223	Moderate
6	C4	Analysing	0.561	−0.245	Moderate

As shown in Table 8, the most difficult item was Item 18, with a correct-response proportion of 0.415 and a logit value of +0.34. This indicates that fewer than half of the students answered the item correctly. The second most difficult item was Item 9, with a correct-response proportion of 0.452 and a logit value of +0.19. Both items had positive logit values, confirming that they were relatively more difficult than the other items.

The remaining difficult items, namely Items 17, 4, and 6, had correct-response proportions ranging from 0.508 to 0.561. Although their logit values were negative, the proportions indicate that these items were still more challenging compared with the majority of the test items. These findings suggest that students experienced greater difficulty in items requiring conceptual understanding and analytical reasoning, especially those related to understanding relationships, interpreting causes, and analysing global warming phenomena. Now the following table presents the five easiest items based on the highest proportion of correct responses.

Table 9. Five easiest items

Item	Level	Cognitive Process	p	Logit	Category
11	C2	Understanding	0.902	−2.221	Easy
14	C3	Applying	0.873	−1.928	Easy
2	C1	Remembering	0.870	−1.904	Easy
7	C3	Applying	0.865	−1.858	Easy
19	C3	Applying	0.862	−1.836	Easy

Table 9 shows the five easiest items in the test. The easiest item was Item 11, with a correct-response proportion of 0.902 and a logit value of −2.221, indicating that more than 90% of students answered the item correctly. Other easy items included Items 14, 2, 7, and 19, with correct-response proportions ranging from 0.862 to 0.873. These items were easier because most students were able to respond correctly.

The dominance of negative logit values among the easiest items indicates that these items were located below the average difficulty level. In other words, they measured concepts that were more accessible to students, such as remembering basic information, recognizing familiar examples, or applying concepts in relatively simple contexts. Although easy items are useful for identifying basic mastery, too many easy items may reduce the ability of the test to distinguish students with higher levels of understanding (Purnomo et al., 2021). Therefore, future revisions of the instrument should consider adding more items with moderate to high difficulty, especially at the analytical level.

Wright Map Analysis

The Wright Map analysis was conducted to examine the alignment between students' ability and item difficulty on the same logit scale. In this map, the left panel represents the distribution of students' estimated ability, while the right panel represents the difficulty level of the test items. Items positioned higher on the logit scale indicate greater difficulty, whereas items positioned lower indicate easier items. The Wright Map therefore provides a visual representation of whether the test items were appropriately targeted to the students' ability levels (Irmayanti et al., 2023).

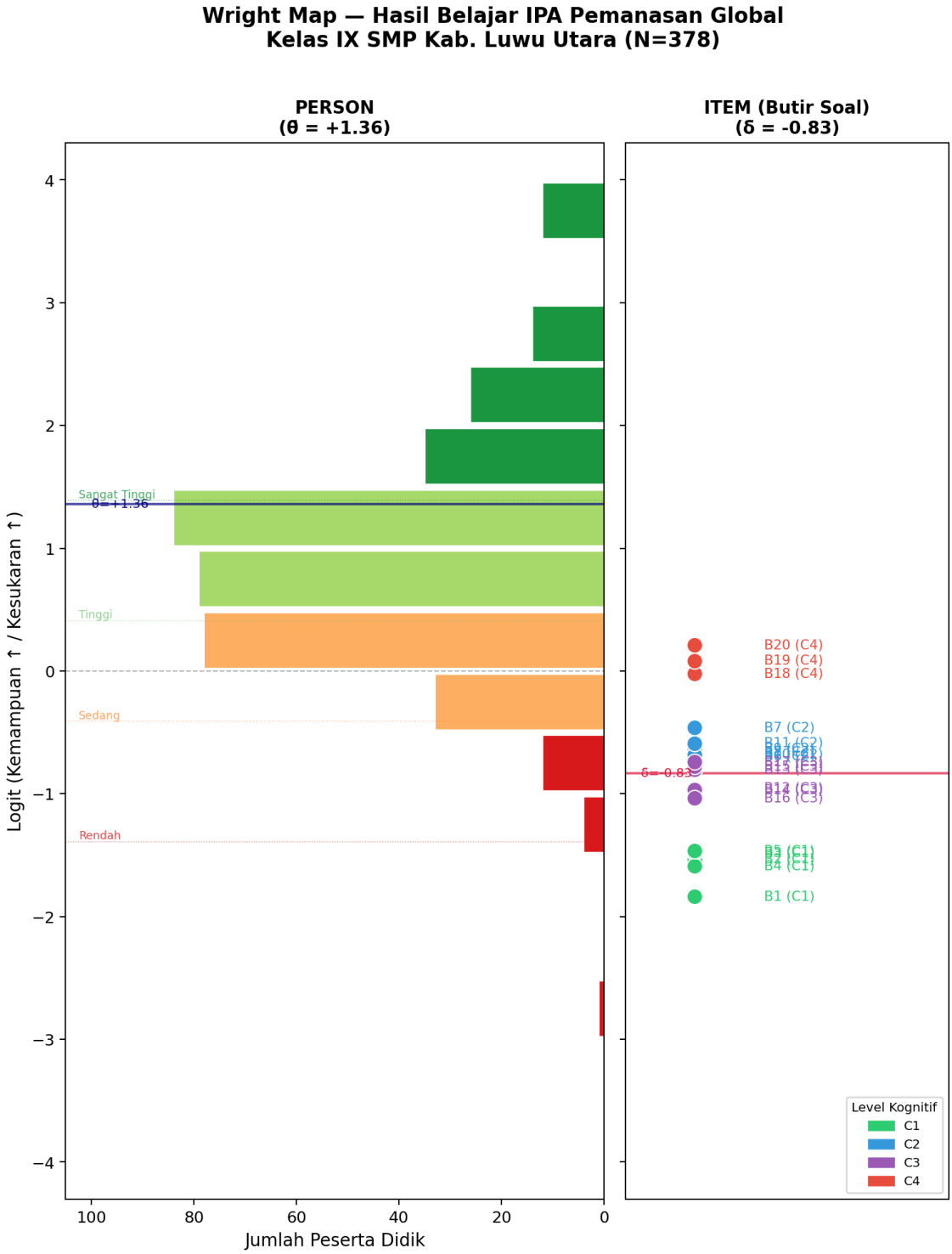

Figure 4. Wright Map of overall science learning outcomes

Figure 4 shows that students' ability estimates were generally distributed at a higher level than most item difficulties. The mean person ability was approximately +1.36 logits, while the mean item difficulty was approximately −0.83 logits. This indicates that, overall, the test items were easier than the average ability level of the students. In other words, many students had sufficient ability to answer most of the items correctly.

The item distribution also shows a clear cognitive pattern. Items at the C1 level were located in the lower part of the logit scale, indicating that these items were relatively easy. This

finding is reasonable because C1 items generally require students to recall basic facts or definitions. Meanwhile, items at the C4 level, especially B18, B19, and B20, were located closer to the upper part of the item difficulty scale. This indicates that analytical items were more difficult and provided greater cognitive challenge for students.

The map also reveals a gap between student ability and item difficulty. Most students were located above the average item difficulty line, while only a few items were positioned near the higher range of student ability. This suggests that the instrument was able to measure students' basic and moderate understanding of global warming, but it was less optimal for distinguishing students with higher levels of achievement. Therefore, future test development should include more items with moderate to high difficulty, particularly items that assess analytical reasoning, causal explanation, and evidence-based problem solving related to global warming.

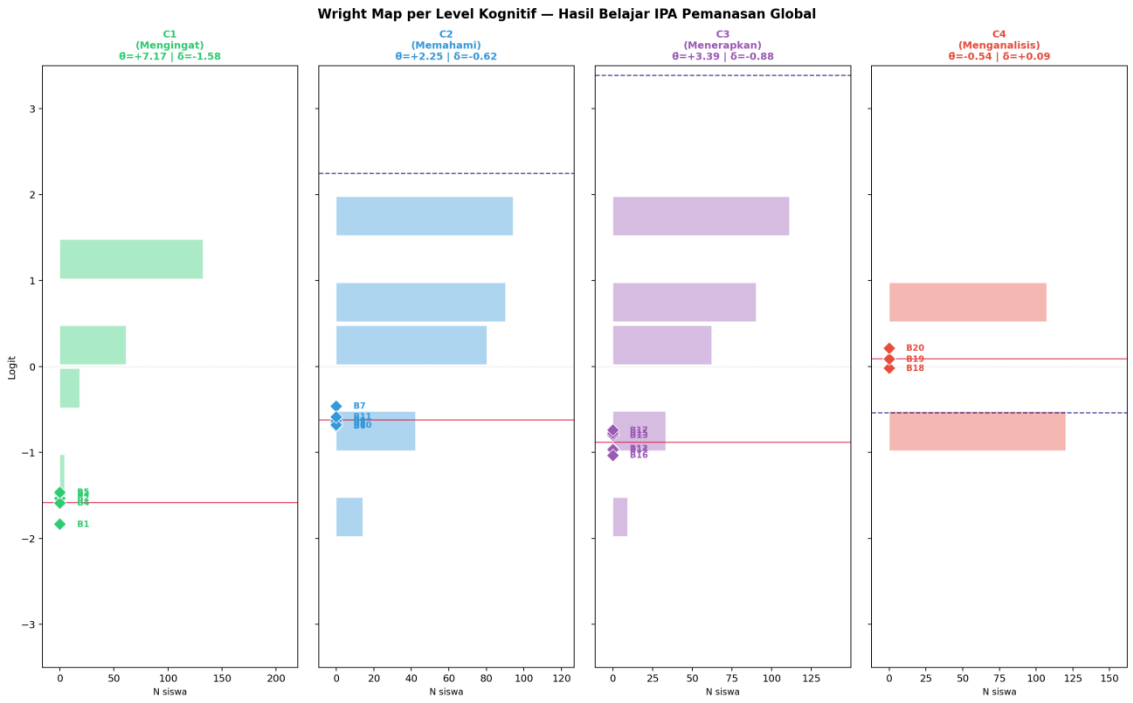

Figure 5. Wright Map by cognitive level

Figure 5 presents the Wright Map of science learning outcomes by cognitive level. The map shows that each cognitive level had a different pattern of item difficulty and student ability distribution. At the C1 remembering level, the items were positioned at the lower part of the logit scale, indicating that remembering items were relatively easy for students. This result is consistent with the high percentage score obtained at C1, suggesting that most students were able to recall basic information and factual concepts related to global warming.

At the C2 understanding level, the distribution of item difficulty was slightly higher than C1, but most items were still located below the average student ability. This indicates that although understanding items were more demanding than remembering items, they were still generally within students' ability range. Meanwhile, the C3 applying level showed a wider spread of student ability and item difficulty, suggesting that students' ability to apply concepts varied more considerably.

The C4 analysing level showed the highest item difficulty compared with the other cognitive levels. Items in this level were positioned closer to the upper part of the logit scale, indicating that analytical items provided greater cognitive challenge. This finding confirms that items requiring students to analyse causes, effects, and relationships in global warming phenomena were more difficult than items requiring recall or basic understanding. Therefore, the instrument should include more C4-level items in future development to better measure students' higher-order thinking skills.

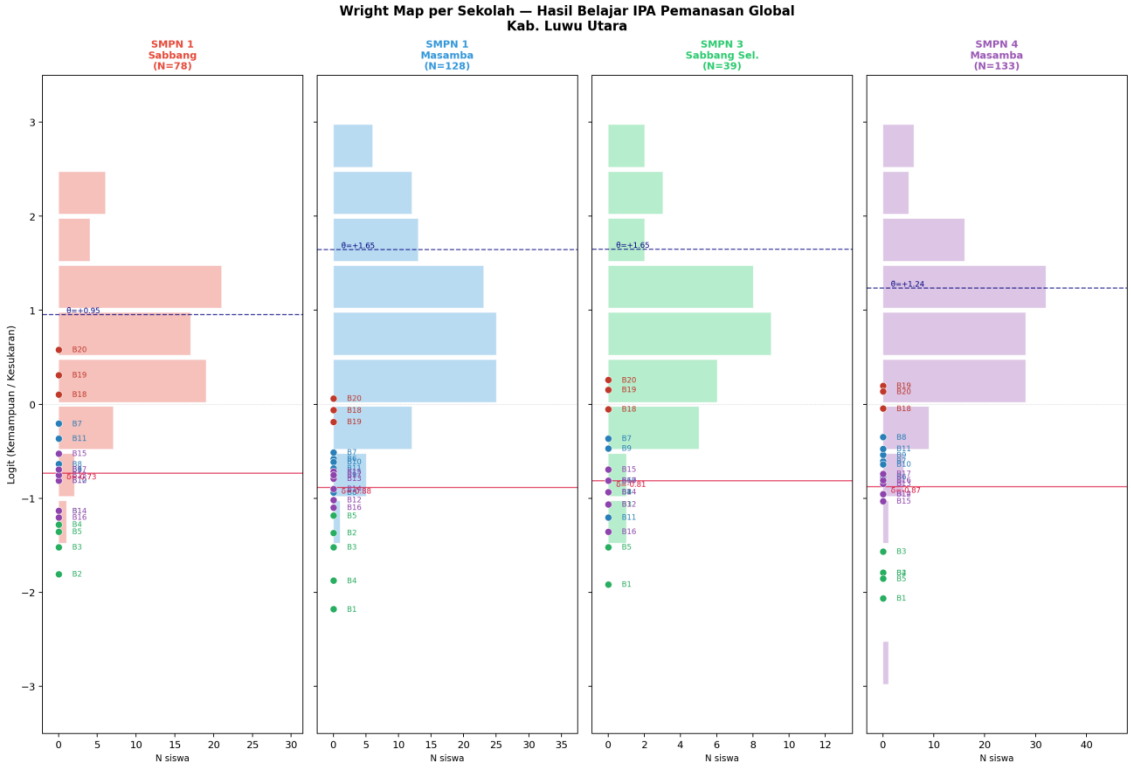

Figure 6. Wright Map by school

Figure 6 presents the Wright Map by school, showing the distribution of student ability and item difficulty across the four schools. In general, all schools showed a similar pattern: student ability tended to be positioned higher than most item difficulty levels. This indicates that the test items were generally easier than the average ability of students in each school.

Among the four schools, SMPN 3 Sabbang Selatan showed the strongest ability distribution. This finding is consistent with the descriptive statistics, where this school obtained the highest mean score and was categorized as very high. The Wright Map also indicates that students in this school were more concentrated at higher ability levels, suggesting stronger mastery of the global warming topic.

In contrast, SMPN 1 Sabbang showed a lower and more varied student ability distribution compared with the other schools. This pattern is consistent with its lower mean score and larger standard deviation. However, the school still remained in the high category, meaning that students generally achieved adequate learning outcomes.

The school-level Wright Map confirms that the instrument was able to describe differences in learning outcomes among schools. However, the gap between student ability and

item difficulty suggests that the test may not have been sufficiently challenging for students with higher achievement. Future instruments should therefore include more moderate and difficult items, especially at the C3 and C4 levels, to improve measurement precision across schools.

The Wright Map figures support the item-level interpretation by visually locating both student ability and item difficulty on a shared logit scale (Boone et al., 2014). In general, the maps indicate that the majority of items were located below the average student ability range, confirming that the instrument was generally too easy for this sample. This means the test provided adequate information for identifying low-performing students but was less sensitive in distinguishing between students at the high end of the ability range. The maps also show that C4 items (Items 6, 16, 20) and several C2 items (Items 9, 18) clustered closest to the average ability level, making them the most diagnostically valuable items in the current instrument. An improved instrument should be strengthened with additional moderate and difficult items, particularly at C2 and C4 levels, to better capture students' analytical understanding of global warming (Osborne, 2010; National Research Council, 2012).

Discussion

The overall finding that Grade IX students in North Luwu Regency achieved high science learning outcomes on the topic of global warming is pedagogically meaningful. Global warming is not an isolated science topic; rather, it is closely connected to environmental risks that directly affect students' local communities, particularly in areas that have experienced climate-related disasters. In this context, students' high achievement suggests that science instruction has been reasonably effective in helping students acquire essential knowledge about the causes, mechanisms, impacts, and mitigation of global warming. This finding is in line with previous studies which emphasize that science learning about environmental issues can strengthen students' conceptual awareness and support their understanding of real-world ecological problems (Hifjir & Agustizar, 2020; Prabawati, 2021; Wildan et al., 2020).

The mean score of 14.77 out of 20 indicates that, in general, students were able to answer most of the test items correctly. This result confirms that the learning outcomes were in the high category. From a classroom assessment perspective, this achievement reflects that students had mastered many of the basic concepts related to global warming. However, the standard deviation of 3.07 also indicates variation in students' mastery levels. Although the majority of students performed well, not all students reached the same level of understanding. This variation is important because it suggests that science teachers still need to provide differentiated learning support, especially for students whose scores remain in the lower range.

The analysis by school showed that all schools achieved high to very high categories, but the level of achievement differed across schools. SMPN 3 Sabbang Selatan obtained the highest mean score and was the only school categorized as very high. This result may indicate stronger learning support, better classroom engagement, or more effective instructional delivery in that school. In contrast, SMPN 1 Sabbang obtained the lowest mean score, although it still remained in the high category. The difference among schools suggests that learning outcomes are not only influenced by students' individual abilities, but also by school-level factors such as learning environment, teacher strategy, availability of learning resources, classroom atmosphere, and students' learning motivation. This is consistent with the view that learning

achievement is shaped by both internal and external factors, including teaching quality, learning facilities, and students' readiness to learn (Arikunto, 2018; Siregar, 2024).

The cognitive-level analysis provides a more detailed picture of students' achievement. The highest achievement was found at the C1 remembering level, indicating that students were generally able to recall factual information and basic definitions related to global warming. This result is understandable because C1 items usually require lower cognitive demand, such as recognizing terms, identifying causes, or recalling examples. Strong performance at this level shows that students had acquired foundational knowledge. However, science learning should not stop at remembering facts. According to the revised Bloom's taxonomy, meaningful learning requires students to move beyond recall toward understanding, applying, analysing, evaluating, and creating knowledge (Anderson & Krathwohl, 2015; Nafiati, 2021).

Interestingly, the lowest mean value was found at the C2 understanding level. This suggests that some students may still experience difficulty in explaining concepts, interpreting scientific information, or connecting one concept with another. In science learning, understanding is a crucial bridge between remembering and applying. Students may be able to memorize terms such as greenhouse gases, global warming, or climate change, but they may not fully understand the causal mechanisms behind these phenomena. This finding indicates the need for teachers to strengthen conceptual explanation through visual representations, contextual examples, discussion, and inquiry-based activities. Without strong conceptual understanding, students may struggle when they are required to apply or analyse scientific ideas in more complex situations.

The relatively good achievement at the C3 applying level indicates that students were able to use global warming concepts in familiar contexts. This may be because the topic is closely related to daily life, such as the use of vehicles, electricity consumption, deforestation, plastic waste, and environmental conservation. Contextual science learning can help students connect abstract concepts with real phenomena, making learning more meaningful (Umar et al., 2024). This supports the argument that science instruction should be linked to students' lived experiences so that scientific knowledge becomes relevant and applicable (Prabawati, 2021; National Research Council, 2012).

The C4 analysing level showed a relatively strong result, but this finding should be interpreted carefully. Although the percentage score at C4 was in the high category, there were only three items measuring this level. A small number of C4 items limits the depth of interpretation because it may not fully represent students' analytical ability. Analytical thinking in science requires students to identify relationships, compare evidence, explain causal mechanisms, interpret data, and construct logical arguments. Therefore, future instruments should include more C4 items with varied contexts and more complex stimuli (Maryani et al., 2021). This is important because higher-order thinking skills are central to science education and are needed for students to understand environmental issues critically (Ramdani et al., 2021; Osborne, 2010; National Research Council, 2012).

The item difficulty analysis showed that the test items were generally easy to moderate for the participants. None of the items fell into the difficult category. This indicates that the instrument was suitable for describing students' general mastery of global warming concepts, but it was less optimal for distinguishing students with high and very high ability. In educational

measurement, a good achievement test should contain a balanced distribution of easy, moderate, and difficult items. Easy items are useful for identifying basic mastery, moderate items are useful for differentiating average ability, and difficult items are necessary to challenge and identify students with higher levels of understanding. If most items are easy, many students may obtain high scores, but the test becomes less sensitive in detecting deeper conceptual understanding (Bond & Fox, 2015; Boone et al., 2014).

The Rasch-logit item difficulty analysis strengthens this interpretation. Items with negative logit values indicate easier items, while positive logit values indicate more difficult items. In this study, only a small number of items had positive logit values, while most items had negative values. This pattern shows that most items were located below the average ability level of students. The most difficult items were those with the lowest proportion of correct responses, especially Item 18 and Item 9 (Soeharto & Csapó, 2021). These items likely required more complex reasoning or deeper conceptual understanding than the other items. Conversely, the easiest items, such as Item 11 and several others with high correct-response proportions, may have measured basic factual knowledge or familiar applications. This pattern suggests that the instrument needs further refinement so that it can better represent a wider range of cognitive demands.

The KR-20 coefficient of 0.674 indicates moderate internal consistency. This value suggests that the instrument was acceptable for descriptive classroom-based research, but it still requires improvement for stronger measurement quality. Moderate reliability may be caused by several factors, including variation in item difficulty, limited number of items, weak distractors, or uneven representation of cognitive levels. In multiple-choice tests, distractor quality is particularly important because weak distractors make items too easy and reduce the ability of the test to discriminate among students. Therefore, improving the instrument should involve revising item stems, strengthening distractors, adding more moderate and difficult items, and increasing the number of items that assess higher-order thinking skills (Ernawati et al., 2022; Arikunto, 2018; Boone et al., 2014).

The Wright Map analysis provides further evidence that the instrument was easier than students' average ability. The map showed that student ability was generally positioned higher than most item difficulty levels. This indicates a person-item gap, meaning that the items did not fully match the students' ability distribution. From the perspective of Rasch measurement, an ideal test should show good alignment between person ability and item difficulty. When many students are positioned above most items, the instrument becomes less precise for measuring students at the upper ability range. In this study, the Wright Map suggests that the test was able to measure basic and moderate mastery, but it was less effective in differentiating students with stronger conceptual and analytical ability (Olsen et al., 2023; Bond & Fox, 2015).

The Wright Map by cognitive level also showed that C1 items were located at the lower part of the logit scale, confirming that remembering items were relatively easy. C2 and C3 items were generally located in the middle-to-lower range, while C4 items were closer to the higher difficulty range. This pattern is theoretically consistent with Bloom's taxonomy, where remembering represents lower-order thinking and analysing represents higher-order thinking. However, even the C4 items were not sufficiently difficult to fully challenge high-performing students. This finding indicates that the instrument should be improved by developing C4 items

that require interpretation of data, analysis of cause-and-effect relationships, evaluation of environmental cases, and evidence-based explanation.

The Wright Map by school confirmed the descriptive findings. SMPN 3 Sabbang Selatan showed the strongest ability distribution, consistent with its highest mean score. Meanwhile, SMPN 1 Sabbang showed a lower and more varied distribution of student ability. However, across all schools, the general pattern remained similar: most item difficulties were below the average student ability. This consistency indicates that the issue is not limited to one school, but relates to the overall construction of the test instrument. Therefore, the revision of the instrument should be conducted at the item level, particularly by increasing difficulty and improving cognitive complexity.

Pedagogically, the findings imply that science learning on global warming should move beyond factual delivery, supported by appropriate science learning media and tools (Samputri et al., 2024). Students appear to have acquired basic knowledge, but the instrument and learning process need to provide more opportunities for deeper reasoning. Teachers can strengthen learning by using problem-based learning, inquiry-based learning, argumentation activities, data interpretation tasks, and local environmental case studies. For example, students can be asked to analyse the relationship between deforestation, rainfall intensity, flooding, and climate change in their own region. Such activities can help students develop not only knowledge, but also scientific reasoning and environmental awareness.

In addition, the local context of North Luwu Regency provides a strong foundation for contextual science learning. Because students live in an area affected by environmental and climate-related problems, global warming can be taught through real local phenomena rather than abstract textbook explanations. Local disaster cases, changes in rainfall patterns, land-use change, and community adaptation can be used as learning resources. This approach may help students understand that science is not separated from everyday life, but is directly related to social, environmental, and community resilience issues.

Overall, the findings indicate that Grade IX students in North Luwu Regency demonstrated good science learning outcomes on global warming. However, the assessment results also reveal important limitations in the instrument. The dominance of easy and moderate items, the moderate KR-20 reliability, and the person-item gap in the Wright Map suggest that the instrument needs to be strengthened to better measure higher-order thinking. Future test development should increase the number of C3 and C4 items, improve distractor quality, and include more contextual problem-solving tasks. These improvements will make the assessment more sensitive, more diagnostically useful, and more aligned with the goals of science education in preparing students to understand and respond to environmental challenges.

CONCLUSION

Based on the reanalysis of the science learning outcome data of 378 Grade IX students on the topic of global warming, this study concludes that students generally demonstrated a high level of achievement. The recalculated mean score was 14.77 out of a maximum score of 20, closely matching the mean score reported in the original thesis (14.76), with a total of 5,583 correct responses across all items. These findings indicate that students possessed a relatively strong understanding of global warming concepts. At the school level, SMPN 3 Sabbang Selatan

recorded the highest mean score (17.20), placing it in the very high category, whereas SMPN 4 Masamba (15.16), SMPN 1 Masamba (14.17), and SMPN 1 Sabbang (13.85) were all classified in the high category, reflecting consistently positive learning outcomes across the participating schools. Analysis of achievement across cognitive domains revealed that students performed best at the C1 (remembering) level, achieving an average score of 83.33%, while the lowest performance was observed at the C2 (understanding) level, with an average score of 67.72%. Achievement at the C3 (applying) and C4 (analysing) levels reached 78.61% and 72.75%, respectively. These results suggest that students were more successful in recalling factual knowledge than in demonstrating deeper conceptual understanding and higher-order cognitive skills. The psychometric evaluation of the instrument showed a KR-20 reliability coefficient of 0.674, indicating moderate internal consistency. Item difficulty analysis further revealed that 13 items were categorized as easy and 7 items as moderate, with no items classified as difficult. Item 18 emerged as the most challenging item ($p = 0.415$), whereas Item 11 was the easiest ($p = 0.902$). Overall, the instrument was effective in describing students' general mastery of global warming concepts; however, its predominance of easy items limited its ability to discriminate among students with higher levels of achievement. Consequently, future assessment development should include a greater proportion of moderately difficult and difficult items, particularly those targeting the C2, C3, and C4 cognitive levels. In addition, future studies are encouraged to employ complete student-by-item response datasets to facilitate more comprehensive Rasch analyses, including examinations of item fit, person reliability, item reliability, and separation indices, thereby providing a more robust evaluation of instrument quality and student learning outcomes (Olsen et al., 2023; Boone et al., 2014).

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2015). A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives. Longman.

Annisa, L. W., & Habibi, A. A. (2020). Hubungan keterampilan berpikir kritis dengan hasil belajar peserta didik. Edubiologica, 8(1), 35–37. https://journal.uniku.ac.id/index.php/edubiologica/article/view/2337

Arif, R. N. H., Arif, R. M., & Samputri, S. (2024). Description ability literacy science numeracy in Class VIII students at SMP IT Wahdah Islamiyah Makassar. ARRUS Journal of Social Sciences and Humanities, 4(3), 338–343. https://doi.org/10.35877/soshum2601

Arif, R. N. H., Samputri, S., & Yanto, N. (2023). Differences in student motivation by using Quizizz media and printed books in Class VIII SMPN 27 Makassar. ARRUS Journal of Social Sciences and Humanities, 3(5), 718–723. https://doi.org/10.35877/soshum2193

Arikunto, S. (2018). Prosedur penelitian: Suatu pendekatan praktik (Ed. revisi). Rineka Cipta.

Azizah, N., Suseno, M., & Hayat, B. (2021). Item analysis of the Rasch model items in the final semester exam Indonesian language lesson. World Journal of English Language, 12(1), 15–23. https://doi.org/10.5430/wjel.v12n1p15

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage.

Ernawati, E., Manik, F. Y., Trisnawati, R. D., Emiliana, E., & Yuliawati, S. (2022). Understanding and quality of minimum competency assessment (AKM) questions made by integrated science teachers in junior high schools. Jurnal Penelitian dan Evaluasi Pendidikan, 26(2), 125–138. https://doi.org/10.21831/pep.v26i2.48670

Haw, L. H., Sharif, S. B., & Han, C. G. K. (2022). Analyzing the science achievement test: Perspective of classical test theory and Rasch analysis. International Journal of Evaluation and Research in Education (IJERE), 11(4), 1714–1724. https://doi.org/10.11591/ijere.v11i4.22304

Hifjir, M., & Agustizar, M. (2020). Tingkat pengetahuan lingkungan hidup dengan sikap peduli lingkungan hidup pada siswa SMP Negeri se-Kluet Raya. GeoRafflesia: Jurnal Pendidikan Geografi, 5(2), 65–74.

Irmayanti, R., Rusdi, M., & Yusnaidar. (2023). The Rasch model: Implementation of physics learning evaluation instrument based on higher order thinking skills. Integrated Science Education Journal, 4(2), 62–68. https://doi.org/10.37251/isej.v4i2.325

Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitrianawati, M. (2021). HOTs multiple choice and essay questions: A validated instrument to measure higher-order thinking skills of prospective teachers. Turkish Journal of Science Education, 18(4), 674–690. https://doi.org/10.36681/tused.2021.97

Nafiati, D. A. (2021). Revisi taksonomi Bloom: Kognitif, afektif, dan psikomotorik. Humanika: Jurnal Ilmiah Kajian Pendidikan, 21(1), 45–56. https://jurnal.uny.ac.id/index.php/humanika/article/view/29252

National Research Council. (2012). A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. The National Academies Press.

Olsen, A. A., Mostacedo-Marasovic, S. J., & Forbes, C. T. (2023). Applying Rasch modeling to a global climate change concept knowledge assessment for secondary students. In X. Liu & W. J. Boone (Eds.), Advances in Applications of Rasch Measurement in Science Education (pp. 169–192). Springer. https://doi.org/10.1007/978-3-031-28776-3_8

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. Science, 328(5977), 463–466. https://doi.org/10.1126/science.1183944

Prabawati, L. (2021). Analisis kepekaan lingkungan siswa SMP pada pembelajaran IPA topik pemanasan global. IMPROVEMENT: Jurnal Ilmiah untuk Peningkatan Mutu Manajemen Pendidikan, 8(2), 36–46.

Purnomo, A. B., Sudibyo, E., Budiyanto, M., Sari, D. A. P., & Sabtiawan, W. B. (2021). Scientific literacy of COVID-19 in the first year pandemic: Design and development test. Jurnal Penelitian Pendidikan IPA, 7(1), 48–55. https://doi.org/10.26740/jppipa.v7n1.p48-55

Ramdani, A., Jufri, A. W., Gunawan, G., Fahrurrozi, M., & Yustiqvar, M. (2021). Analysis of students' critical thinking skills in terms of gender using science teaching materials based on the 5E learning cycle integrated with local wisdom. Jurnal Pendidikan IPA Indonesia, 10(2), 187–199. https://doi.org/10.15294/jpii.v10i2.29956

Samputri, S., Ramlawati, Arif, R. N. H., Arif, R. M., & Muhiddin, S. M. A. (2024). Design alat peraga IPA dan kebermanfaatannya berdasarkan hasil workshop terhadap alumni mahasiswa Program Studi Pendidikan IPA. Vokatek: Jurnal Pengabdian Masyarakat, 2(2). https://doi.org/10.61255/vokatekjpm.v2i2.376

Siregar, H. T. (2024). Faktor-faktor yang mempengaruhi hasil belajar dalam pembelajaran PAI. Jurnal Ilmu Tarbiyah dan Keguruan (JITK), 2(2), 215–226. https://ejournal.edutechjaya.com/index.php/jitk/article/view/791

Soeharto, S., & Cspó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. Heliyon, 7(11), e08352. https://doi.org/10.1016/j.heliyon.2021.e08352

Sufinasa, S. A., Hasanuddin, H., Saenab, S., Arif, R. N. H., Samputri, S., & Hasan, N. R. (2023). Description of science process skills of Class VIII students of SMPN in Makassar City. Jurnal Eduscience (JES), 10(2), 398–406. https://doi.org/10.36987/jes.v10i2.3891

Sugiyono. (2018). Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R&D. Alfabeta.

Umar, R. A., Husain, R. I., & Arif, R. M. (2024). Penerapan model pembelajaran TGT berbantuan media question card untuk meningkatkan hasil belajar siswa pada pembelajaran IPA di kelas IV SDN No. 17 Dungingi. Eduproxima: Jurnal Ilmiah Pendidikan IPA, 6(4). https://doi.org/10.29100/.v6i4.5367

Wildan, W., Hakim, A., & Supriadi, N. (2020). Pengetahuan, sikap, dan perilaku remaja di Pulau Lombok terhadap isu pemanasan global. Jurnal Penelitian Pendidikan IPA, 6(1), 84–90. https://jppipa.unram.ac.id/index.php/jppipa/article/view/319