



## Analyzing Students' Conceptual Understanding in Scientific Writing Using One-Parameter Logistic Item Response Theory

*\*Dian Pertiwi Josua, Riyan Arthur, Dinny Devi Triana, Lussy Dwiutami Wahyuni*

*\*Cosmetology Education Studies, Faculty of Engineering, Universitas Negeri Jakarta, Jakarta, Indonesia*

Correspondence: E-mail: [dianpertiwijosua@unj.ac.id](mailto:dianpertiwijosua@unj.ac.id)

### ABSTRACT

This study investigates the comparative impact of Project-Based Learning (PjBL) and STEM-integrated Project-Based Learning (PjBL-STEM) on students' creativity in continental food processing classes among 12th-grade culinary students at SMK N 1 Bawen. Employing a quasi-experimental design with a non-equivalent control group pretest-posttest structure, the study involved a sample of 72 students selected through simple random sampling from a total population of 108. Data were collected using structured classroom observation instruments, which were validated through expert judgment and demonstrated strong content validity and reliability (Cronbach's Alpha > 0.80). The data analysis was conducted using independent samples t-test, preceded by normality and homogeneity testing. The results revealed a statistically significant difference in creativity between the two groups ( $p = .001$ ), indicating that the PjBL-STEM model led to superior learning outcomes. Students taught using the PjBL-STEM model achieved a higher average post-test score ( $M = 76.03$ ) compared to those taught using PjBL alone ( $M = 69.11$ ). These findings suggest that integrating STEM into project-based learning effectively enhances students' creativity in vocational culinary education, offering promising implications for instructional design in 21st-century skills development.

© 2025 Kantor Jurnal dan Publikasi UPI

### ARTICLE INFO

#### Article History:

Submitted/Received 17 March 2025

First Revised 203 June 2025

Accepted 23 July 2025

First Available Online 28 July 2025

Publication Date 01 August 2025

#### Keyword:

*Conceptual Understanding, Higher Education, Item Response Theory, One-Parameter Logistic Model, Scientific Writing.*



## 1. INTRODUCTION

Scientific writing is a core academic literacy in higher education because universities create, evaluate, and disseminate knowledge largely through written genres such as essays, proposals, theses, and journal articles. Academic discourse is therefore not merely a channel for reporting results, but also a means of constructing disciplinary knowledge, developing critical thinking, and participating in scholarly communities (Flowerdew & Wang, 2015; Zou & Hyland, 2019).

In Indonesia, this expectation is reinforced by Law No. 12 of 2012 on Higher Education, which positions research, publication, and the cultivation of academic culture within Tridharma and explicitly links higher education to scientific reasoning, research, and the search for truth. At the institutional level, scientific publications and citations are also treated as key indicators of research strength in the national WCU analysis aligned with QS-based metrics, making scientific writing competence strategically important for both students and universities.

Despite its importance, academic writing remains a demanding competence for many students. Writing is a recursive cognitive activity involving planning, translating, and reviewing while simultaneously managing linguistic, rhetorical, and contextual constraints rather than simply arranging sentences in the correct format (Cheung, 2016; Silva, 2022). Recent higher education research likewise shows that writing supports learning, reflection, and knowledge construction, yet assessment practices often emphasize the final product more than the conceptual resources and processes behind it (Lam, 2016; Sala-Bubaré & Castelló, 2018; Strobl et al., 2019). Evidence from instructional interventions further indicates that students' writing performance can improve substantially when structured support is provided, suggesting that weaknesses in scientific writing are diagnosable and teachable rather than incidental (Lee & De La Paz, 2021).

In the Indonesian context, students continue to report difficulties with paraphrasing, coherence, cohesion, writer voice, thesis development, structure, grammar, and reference use in academic and scientific texts (Bram & Angelina, 2022; Wicaksono et al., 2025; Subandowo & Sardi, 2023). These recurring difficulties suggest that success in scientific paper writing depends not only on procedural compliance with format, but also on conceptual understanding of how the components of a paper function together. Conceptual understanding can be viewed as an integrated grasp of ideas and their relationships that enables learners to explain, apply, and adapt knowledge rather than merely reproduce isolated rules (Mills, 2016).

In scientific writing, such understanding includes knowing how a research problem justifies a study, how claims must be supported by evidence, how sections of a paper perform distinct rhetorical functions, and how citation practices contribute to credibility and disciplinary participation (Kivunja & Kuyini, 2017; McKinley, 2015). When assessment focuses mainly on holistic scores or surface correctness, it may obscure which concepts students genuinely understand and which aspects remain fragile. More broadly, measurement research has shown that strong assessment depends on alignment between the conceptualization of a construct and the analytical approach used to measure it (Luciano et al., 2018).

This is where Item Response Theory (IRT) becomes particularly relevant. IRT models the relationship between a latent trait and responses to individual items, enabling the estimation of person ability and item difficulty on a common scale rather than relying solely on total scores (Bichi & Talib, 2018; Cai et al., 2016). Among dichotomous IRT models, the one-

parameter logistic model (1PL) is especially suitable for early instrument development because it offers a parsimonious framework, assumes a single dominant construct, and focuses attention on item difficulty when discrimination is held constant across items (Mallinckrodt et al., 2016).

This feature is useful for mapping which concepts in scientific paper writing are relatively easy or difficult for students to master. In writing assessment research, Rasch-based approaches have been used to examine essay scoring and rating behavior, including automated essay evaluation, but these applications have largely addressed performance measurement rather than students' conceptual understanding of scientific writing content itself (Kim, 2020; Aryadoust et al., 2021).

Accordingly, a clear research gap remains in the development of item-level, theory-informed instruments that can diagnose students' conceptual understanding of scientific paper writing in higher education. Existing writing studies have richly documented students' difficulties and support strategies, but fewer studies have translated those insights into psychometrically robust tools capable of distinguishing student ability from item difficulty within this specific domain.

The present study addresses this gap by developing and validating an instrument to measure students' conceptual understanding of scientific paper writing using the 1PL IRT model. Specifically, the study seeks to estimate students' ability levels and identify the relative difficulty of key concepts such as problem formulation, reference management, academic language use, and paper organization. By doing so, the study offers an objective basis for diagnosing learning needs and for improving scientific writing instruction through evidence-based course design and assessment in higher education.

## **2. METHOD**

### **2.1 Research Design**

This study employed a quantitative research design to examine students' conceptual understanding of scientific paper writing using the one-parameter logistic (1PL) model within the Item Response Theory (IRT) framework. The 1PL model was selected because it is appropriate for dichotomously scored items and enables the estimation of student ability and item difficulty on a common latent scale. In this study, students' responses to multiple-choice items were analyzed to determine the probability of a correct response as a function of individual ability and item difficulty. This design was intended not only to measure the overall level of students' conceptual understanding but also to identify response patterns that may indicate specific conceptual difficulties. In particular, the analysis was used to detect students with relatively high ability who responded incorrectly to certain items, as well as students with lower estimated ability who answered more difficult items correctly, thereby providing a more diagnostic view of performance in scientific paper writing materials.

The population of the study consisted of undergraduate students at Universitas Negeri Jakarta who had taken courses related to scientific paper writing. The sample was selected using stratified random sampling to ensure representation across different academic levels. A total of 135 students participated in the study. This sample size was considered adequate for preliminary IRT calibration and for estimating item difficulty under the 1PL model. The instrument used in this study was a 30-item multiple-choice test developed on the basis of learning indicators in scientific paper writing. The items covered several dimensions of conceptual understanding, including comprehension of scientific paper structure, formulation of research problems, use of academic language, reference management, and

organization of arguments. Prior to administration, the instrument underwent content validation through expert judgment to ensure the relevance, clarity, and representativeness of the items. Internal consistency was also examined before the main analysis to confirm that the instrument was sufficiently reliable for use in the study.

## **2.2 Data Collection Technique**

Data were collected through a test-based procedure administered in a controlled academic setting. The selected students completed the conceptual understanding test within 60 minutes under supervised conditions to reduce potential distractions and ensure consistent testing procedures across participants. Before the test was administered, participants were informed about the purpose of the study and were assured that their responses would be treated confidentially and used solely for research purposes. Participation was voluntary, and students provided their consent prior to taking part in the study.

The response data obtained from the test were coded dichotomously, with correct answers scored as 1 and incorrect answers scored as 0. After the administration process was completed, all answer sheets were checked for completeness and accuracy. The cleaned dataset was then prepared for IRT analysis using Winsteps software. This procedure ensured that the data entered into the analysis accurately reflected students' response patterns and could be used to estimate both person ability and item difficulty parameters with greater precision.

## **2.3 Data Analysis Technique**

The data were analyzed using the 1PL IRT model with the assistance of Winsteps software. The analysis began with preliminary data screening to verify the completeness and consistency of students' responses. Afterward, the assumption of unidimensionality was examined to confirm that the test measured a single dominant latent trait, namely conceptual understanding of scientific paper writing. This step was important because the 1PL model assumes that responses are driven primarily by one underlying ability dimension.

Following the assumption check, item calibration was conducted to estimate the difficulty parameter of each item and the ability level of each student. The primary output of the analysis consisted of item difficulty estimates, person ability estimates, and model-based information regarding the functioning of the instrument. Item difficulty parameters were interpreted to determine which concepts were relatively easy or difficult for students to understand. In addition, the test information generated by the model was used to evaluate how effectively the instrument measured student ability across different levels of the latent trait. The analysis also considered the extent to which individual items functioned appropriately in distinguishing students with different levels of conceptual understanding.

The results were interpreted diagnostically to describe the distribution of student abilities and to identify items that functioned effectively or ineffectively in measuring conceptual understanding of scientific paper writing. Items showing inappropriate functioning were considered for revision in future instrument development. Overall, the analysis was intended to provide empirical evidence for improving both the quality of the assessment instrument and the teaching of scientific paper writing in higher education.

### 3. RESULT AND DISCUSSION

#### 3.1 Result

The results of this study (Table 1) analyze students' conceptual understanding of scientific writing materials using the 1 Parameter Logistic (1PL) Model of Item Response Theory (IRT). The research data were obtained from 135 students with 25 questions evaluated using a two-category scale (true/false). The analysis was carried out using WINSTEPS software to evaluate the quality of the instrument and data consistency.

**Table 1.** Data Analysis Results

Mark	Total score	Measure	INFIT		OUTFITS	
			MNSQ	ZSTD	MNSQ	ZSTD
Mean	12.9	.11	.99	.1	1.02	.1
SD	7.3	1.98	.18	.7	.64	.8
Max.	24.0	3.83	1.8	1.8	4.01	2.2
Min.	1.0	-3.75	-2.0	-2.0	.23	-1.5
Real RMSE .63						
Separation 2.96						
Person reliability .90						
Maximum extreme score: 2 people						
Minimum extreme score: 1 person						

The average total score is 12.9, with a maximum score of 24 and a minimum of 1. The Mean Measure of 0.11 indicates that students' abilities are slightly above the logit average (0), with significant variations in ability (SD Measure = 1.98). The average model error is 0.59, indicating moderate uncertainty in estimating individual abilities. The average student ability (person measure) of 0.11 logit indicates that overall, the level of student understanding is slightly above the average instrument ability. This means that most students understand the material well when writing scientific papers.

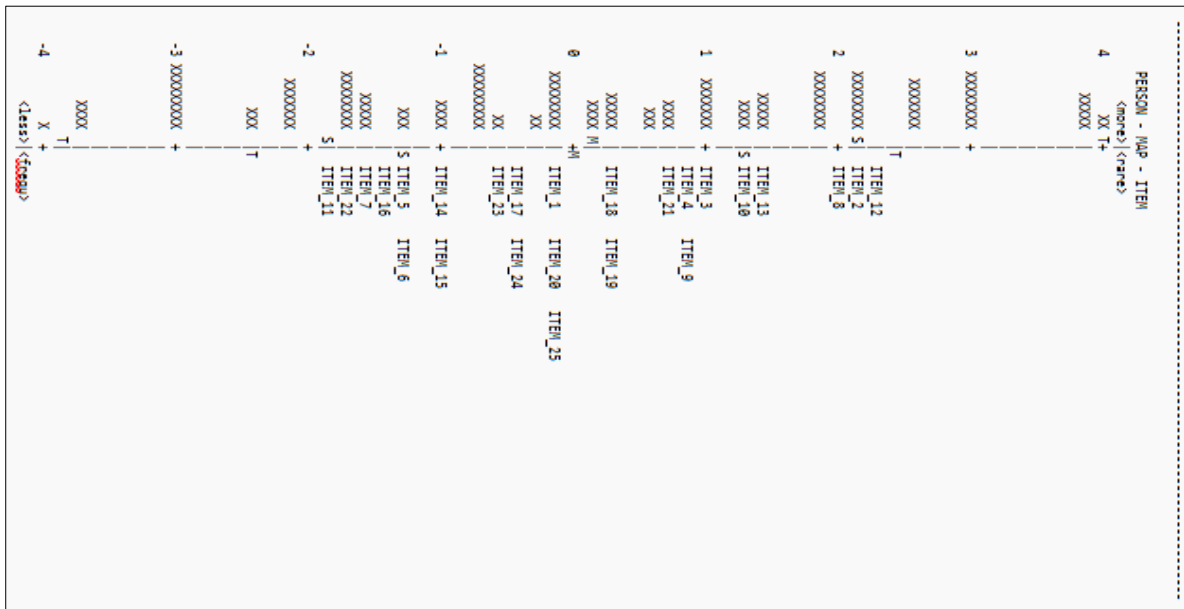
The standard deviation (SD) measure = 1.98 indicates significant variation in student ability. Some students have an excellent understanding, while others show difficulty understanding the material. Two respondents had maximum scores (24), indicating they answered correctly and could not be measured further. One respondent had a minimum score (1), possibly answering all incorrectly, requiring additional verification. A total of 2 students with maximum scores and one with a minimum score reflect the presence of students who answered correctly (excellent understanding) or all incorrectly (very low understanding). This is relevant to evaluate the effectiveness of learning or the distribution of the difficulty level of the questions. The analysis results show a Real RMSE of 0.63, indicating a fairly small average error rate. Person reliability of 0.90 indicates that this instrument is highly reliable in distinguishing individual abilities. Person separation of 2.96 indicates that student abilities can be grouped into three different levels: low, medium, and high.

The analysis results show that student data are consistent with infit and outfit statistics that are within expected limits. Variation in student abilities (SD Measure = 1.98) and person reliability of 0.90 indicates that this instrument can accurately measure student abilities.

However, the presence of individuals with extreme scores requires further analysis to understand response patterns.

The logit scale shows the student's ability and difficulty level of the questions (Figure 1), where according to the analysis results, most students are in the logit range of -2 to +2, reflecting average to relatively high abilities. Most items are in the logit range of -2 to +2, reflecting moderate to slightly above or below average difficulty levels. There are some students with very high abilities (logit > 3) and very low (logit < -3), but the number is small. The most difficult items are ITEM\_12, ITEM\_2, and ITEM\_8 (logit around +2 to +3). The easiest items are ITEM\_11, ITEM\_22, and ITEM\_7 (logit around -2). The map shows a relatively even distribution of students in terms of the difficulty level of the questions, meaning that the questions are designed to cover students' abilities well.

Difficult items (e.g., ITEM\_12, ITEM\_2, ITEM\_8) are suitable for students with high ability (logit > +2). High difficulty levels are irrelevant for students with average ability. Easy items (e.g., ITEM\_11, ITEM\_22, ITEM\_7) are more suitable for students with low ability (logit < -2). Low difficulty levels need to be more challenging for students with high ability.



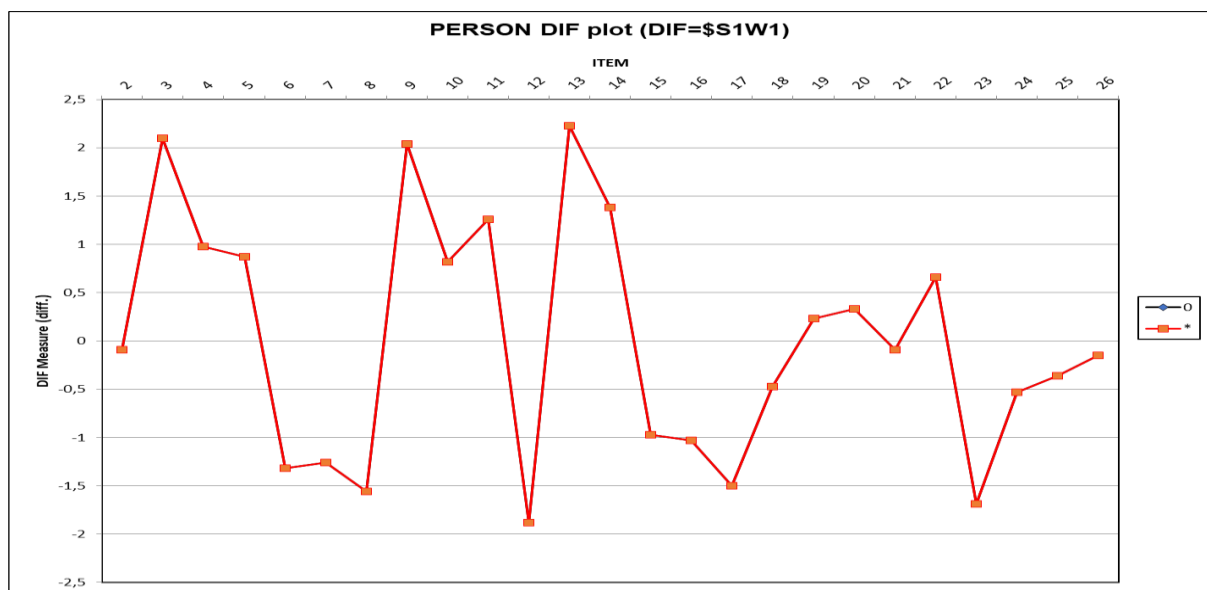
**Figure 1.** Visualization of the relationship between student ability and item difficulty level.

The extremely difficult items (logit > 3) underrepresent very low-ability students, while very high-ability students may find them too easy. The items cover various difficulty levels, providing a comprehensive measure of student understanding. Most students are moderate to fairly high-ability (-2 to +2 logits), indicating a good overall understanding of the concepts. Students with low ability (< - two logits) need special attention, such as remedial learning. This map supports the analysis that the questions can differentiate the student's understanding level well. This study uses the 1 Parameter Logistic (1PL) model in item response theory (IRT) to identify the difficulty level of questions related to the concept.

The research analysis results are depicted graphically (Figure 2) so that items requiring revision can be identified because they show bias or disproportionate difficulty between

student groups. Positive DIF Measure (above 0) indicates that the item is easier for the focus group than for the reference group. The items on selecting research keywords and the structure of writing abstracts are more familiar to students with previous experience writing or reading scientific articles.

A negative DIF Measure (below 0) indicates that the item is more difficult for the focus group. This can occur if the reference group has an advantage in certain materials, such as correct citation methods and ethical standards in publication. The graph shows a significant DIF fluctuation pattern in several items, such as items 3, 9, 14, 22, and 26, indicating that students' understanding of certain concepts in the scientific writing material still varies between groups.



**Figure 2.** Variative Gap Between Respondent Groups and Question Items

Item 3 has a high positive DIF level (easy for the focus group) and refers to questions related to the fundamental understanding of scientific work. Item 9 has a high negative DIF value (difficult for the focus group) and is related to statistical data analysis that requires a more technical understanding. The graph identifies items that require revision because they show bias or disproportionate difficulty among student groups. Items with significant differences indicate that certain material needs to be reviewed, especially if the concept is crucial to the paper's writing, such as in the discussion and conclusion sections.

Items with high DIF (items 3, 9, 14, 22, and 26) must be reviewed to ensure that they measure actual ability and are not influenced by external factors such as writing experience or prior knowledge. Material on the structure of scientific articles (abstract, introduction, methods) and technical aspects such as citations need to be improved to reduce the gap in understanding. Graphics are important in improving the quality of learning and assessment. Based on more inclusive learning materials and fairer questions for all students, further analysis data can be generated. The results indicate that re-simulation or testing can be done on other data for further validation.

### 3.2 Discussion

The findings indicate that students' conceptual understanding of scientific paper writing was generally concentrated in the moderate range, as reflected by the mean person measure being slightly above the mean item difficulty. This pattern suggests that the cohort had not yet achieved a uniformly high level of mastery, although their performance was not indicative of minimal understanding either. Such a profile is plausible in higher education, where students are often exposed to the formal conventions of scientific papers but still struggle to integrate rhetorical structure, evidence use, argument development, and citation practices into a coherent academic performance (Khedri, 2018; Phillips Galloway et al., 2020).

The moderate average score should therefore not be interpreted as evidence of strong scientific writing competence in a comprehensive sense. Rather, it suggests that many students remain in a transitional stage between procedural familiarity and deeper conceptual control over scientific writing conventions (Gere et al., 2019; Teng et al., 2022). Academic writing is widely recognized as a complex activity because it requires students to coordinate planning, organization, source integration, formal conventions, and audience awareness simultaneously rather than in isolation (Deane, 2018; Tusting et al., 2019).

The relatively large standard deviation of person measures further indicates considerable heterogeneity within the cohort. This means that students in the same class may require different forms of instructional support depending on their level of conceptual understanding. Some students may still need assistance with foundational concepts such as text structure and citation conventions, whereas others may require more advanced support in strengthening argument logic and improving the alignment between research problems, methods, and conclusions (Hudson et al., 2021; Yang & Sang, 2023). Differences of this kind are also consistent with studies showing that writing development is shaped by prior literacy experience, disciplinary socialization, reading proficiency, and access to structured feedback (Goldman et al., 2016; Grabe & Zhang, 2016; Hall et al., 2015).

Overall, the present findings suggest that conceptual understanding of scientific paper writing should not be viewed as a binary condition in which students either understand or do not understand the material. Instead, it appears to develop along a continuum, with students occupying different levels of conceptual readiness and demonstrating different instructional needs across the domain of scientific writing (Bell et al., 2016; Gere et al., 2019).

The item difficulty range from -1.88 to 2.23 logits indicates that the instrument captured multiple levels of conceptual demand. This is an important strength because an instrument designed to diagnose conceptual understanding should include both relatively accessible items and more demanding items rather than concentrating only on one difficulty level (Pellegrino et al., 2016; Sands et al., 2018). A broad spread of item difficulty makes it possible to locate both respondents and concepts on the same latent continuum, thereby identifying which aspects of scientific paper writing are relatively easier and which remain more challenging for students (Van Vo & Csapó, 2021).

The presence of highly difficult items may suggest that some concepts in scientific writing require more than simple declarative recall. In the context of scientific paper writing, difficult items are likely to involve integrated forms of reasoning, such as connecting the research problem to its rationale, selecting relevant evidence, and maintaining coherence across sections of a paper (Stiller et al., 2018). This interpretation is aligned with recent research showing that many students continue to experience difficulties when they move beyond sentence-level correctness and are required to coordinate structure, evidence, and argumentation at a broader rhetorical level (Knecht et al., 2023).

At the same time, the relatively easy items remain equally valuable because they help determine whether students have mastered the foundational conventions of scientific paper writing. Basic competencies such as understanding article structure, using academic terminology appropriately, and applying citation conventions correctly are essential because weaknesses in these areas may limit more advanced writing performance (Suleimanova & Lykova, 2023). Previous studies also show that problems related to structure, coherence, referencing, and formal academic language remain persistent obstacles in higher education writing contexts (Alsariera & Yunus, 2023; Rahman et al., 2024).

The close alignment between the mean person measure and the mean item difficulty further indicates that the instrument was reasonably well targeted to the present sample. In measurement terms, this means that the test was appropriate for estimating the abilities of most respondents within the cohort. However, the existence of extreme-score cases and the lower measurement precision at the upper and lower ends of the continuum suggest that future versions of the instrument should include a wider spread of item difficulty (Immekus et al., 2019). More difficult items would help differentiate high-performing students more effectively, whereas easier items would allow for more stable ability estimates among lower-performing students (Woods et al., 2020).

From an instructional perspective, the findings reinforce the view that scientific writing should be taught as a conceptual and disciplinary practice rather than merely as a formal exercise. When students demonstrate moderate ability and the items vary meaningfully in difficulty, the implication is that instruction should move beyond surface correction and template compliance (Doyle, 2023). Instead, students need to understand how each section of a scientific paper serves a distinct communicative purpose and how these sections work together to construct an evidence-based argument (Mercer-Mapstone & Kuchel, 2015).

This point is especially important because source-based academic writing is cognitively demanding. Students are expected not only to understand disciplinary content but also to interpret sources, select relevant evidence, paraphrase accurately, and integrate citations into their own reasoning (Wette, 2017). Studies on source use and academic writing consistently show that these processes remain challenging even for university students who have already had prior exposure to academic writing tasks (Cumming et al., 2016; Gebril & Plakans, 2016). This helps explain why students may achieve moderate overall performance while still showing difficulty on particular items that require integrated conceptual understanding.

The results therefore support more explicit, scaffolded, and formative approaches to teaching scientific writing. Teaching strategies that break writing into analyzable components, provide structured support, and gradually increase conceptual demand are more likely to strengthen students' understanding than one-time, product-oriented writing assignments (Karimnia, 2023). Recent evidence likewise suggests that students benefit from guided support, online scaffolding, structured training, and feedback-rich learning environments when developing academic writing competence (Wang et al., 2024).

The high person reliability obtained in this study also has direct pedagogical implications. Because the instrument was able to distinguish among different levels of conceptual understanding, it can be used not only for summative assessment but also for diagnostic decision-making in instruction. Lecturers may use such evidence to identify concepts that are consistently difficult across cohorts and redesign teaching accordingly. In addition, item-level analysis can help distinguish between actual student misconceptions and weaknesses caused

by ambiguous item construction, which is particularly important in evidence-based assessment practice (Federer et al., 2015).

The present findings also have important implications for assessment development. In many higher education contexts, scientific writing is evaluated through broad rubrics or holistic judgments of written products. Although such approaches are useful for grading, they are often less effective for diagnosing specific conceptual misconceptions in students' understanding of scientific writing. By contrast, the 1PL IRT approach used in this study makes it possible to estimate which concepts are relatively difficult and whether the instrument is appropriately targeted to the sample (Yalçın, 2018). This makes Rasch-oriented measurement particularly useful when the goal is to develop interpretable instruments for instructional improvement rather than merely to generate total scores.

Nevertheless, the current findings also indicate that the instrument still requires refinement. Because the Test Characteristic Curve showed that the test was most informative in the middle of the ability continuum, future development should aim to improve precision at both extremes of the scale. One way to do this is by adding easier items that assess basic conceptual recognition and more difficult items that tap higher-order integration, such as evaluating argument quality or identifying weaknesses in methodological justification. Such revisions would allow the instrument to better capture both lower-performing and higher-performing students, whose abilities were less precisely estimated in the current version of the test.

Another important implication concerns the relationship between psychometric evidence and writing theory. Assessment quality does not improve simply because a more sophisticated statistical model is applied. The interpretability of the results still depends on a clear construct definition and strong alignment between item content and the conceptual dimensions of scientific writing. Research on academic literacy and writing development repeatedly shows that writing competence is multidimensional, socially situated, and context-sensitive, which means that instrument development should remain grounded in both theory and classroom realities.

#### 4. CONCLUSION

Investigating logical writing impacts students' aptitudes and the quality of their research. Understudies who understand the standards of logical writing have a more noteworthy potential to deliver work that can be distributed in legitimate diaries, hence contributing to the accomplishment of execution markers of higher instruction. Through this research, higher education institutions are expected to be more aware of the importance of evaluating and developing students' writing competencies as an integral part of learning. This research contributes to developing IRT-based evaluation instruments, especially in higher education. Lecturers can be more effective in designing assessment instruments that measure learning outcomes and provide meaningful feedback to improve the quality of learning by using this approach.

This opens up openings to encourage ponders to investigate other variables that impact students' capacity in logical composing. For this case, an approach that combines the IRT show with subjective investigation can provide more profound insights into students' discernments of the logical composing learning preparation. In expansion, comparative investigation can be connected to distinctive organization settings to give a broader understanding and generalization of the investigations.

## REFERENCES

- Alsariera, A. H., & Yunus, K. (2023). Issues and challenges in the use of coherence among Jordanian EFL students in writing academic essays. *Journal of Nusantara Studies (JONUS)*, 8(2), 141-156. <https://doi.org/10.24200/jonus.vol8iss2pp141-156>
- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language testing*, 38(1), 6-40. <https://doi.org/10.1177/0265532220927487>
- Bell, R. L., Mulvey, B. K., & Maeng, J. L. (2016). Outcomes of nature of science instruction along a context continuum: Preservice secondary science teachers' conceptions and instructional intentions. *International Journal of Science Education*, 38(3), 493-520. <https://doi.org/10.1080/09500693.2016.1151960>
- Bichi, A. A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in Education*, 7(2), 142-151. <http://doi.org/10.11591/ijere.v7i2.12900>
- Bram, B., & Angelina, P. (2022). Indonesian Tertiary Education Students' Academic Writing Setbacks and Solutions. *International Journal of Language Education*, 6(3), 267-280. <https://doi.org/10.26858/ijole.v6i3.22043>
- Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item response theory. *Annual Review of Statistics and Its Application*, 3(1), 297-321. <https://doi.org/10.1146/annurev-statistics-041715-033702>
- Cheung, Y. L. (2016). Teaching writing. In *English language teaching today: Linking theory and practice* (pp. 179-194). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-38834-2\\_13](https://doi.org/10.1007/978-3-319-38834-2_13)
- Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic purposes*, 23, 47-58. <https://doi.org/10.1016/j.jeap.2016.06.002>
- Deane, P. (2018). The challenges of writing in school: Conceptualizing writing development within a sociocognitive framework. *Educational Psychologist*, 53(4), 280-300. <https://doi.org/10.1080/00461520.2018.1513844>
- Doyle, T. (2023). *Helping students learn in a learner-centered environment: A guide to facilitating learning in higher education*. Taylor & Francis. <https://doi.org/10.4324/9781003445067>
- Federer, M. R., Nehm, R. H., Opfer, J. E., & Pearl, D. (2015). Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. *Research in Science Education*, 45(4), 527-553. <https://doi.org/10.1007/s11165-014-9435-9>
- Flowerdew, J., & Wang, S. H. (2015). Identity in academic discourse. *Annual Review of Applied Linguistics*, 35, 81-99. <https://doi.org/10.1017/S026719051400021X>

- Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for academic purposes*, 24, 78-88. <https://doi.org/10.1016/j.jeap.2016.10.001>
- Gere, A. R., Limlamai, N., Wilson, E., MacDougall Saylor, K., & Pugh, R. (2019). Writing and conceptual learning in science: An analysis of assignments. *Written Communication*, 36(1), 99-135. <https://doi.org/10.1177/0741088318804820>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., ... & Project READi. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219-246. <https://doi.org/10.1080/00461520.2016.1168741>
- Grabe, W., & Zhang, C. (2016). Reading-writing relationships in first and second language academic literacy development. *Language Teaching*, 49(3), 339-355. <https://doi.org/10.1017/S0261444816000082>
- Hall, A. H., Simpson, A., Guo, Y., & Wang, S. (2015). Examining the effects of preschool writing instruction on emergent literacy skills: A systematic review of the literature. *Literacy Research and Instruction*, 54(2), 115-134. <https://doi.org/10.1080/19388071.2014.991883>
- Hudson, A. K., Owens, J., Moore, K. A., Lambright, K., & Wijekumar, K. (2021). "What's the Main Idea?": Using Text Structure to Build Comprehension. *The Reading Teacher*, 75(1), 113-118. <https://doi.org/10.1002/trtr.2016>
- Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019, May). Multidimensional item response theory for factor structure assessment in educational psychology research. In *Frontiers in Education* (Vol. 4, p. 45). Frontiers Media SA. <https://doi.org/10.3389/feduc.2019.00045>
- Karimnia, A. (2023). Exploring TEFL professors' perceptions of the effectiveness of academic writing courses in MA-level TEFL curricula. *Journal of Studies in Language Learning and Teaching*, 1(1), 183-194.
- Khedri, M. (2018). Evidentials in research articles: A marker of discipline. *Pertanika Journal of Social Sciences and Humanities*, 26(T), 145-158.
- Kim, H. (2020). Effects of rating criteria order on the halo effect in L2 writing assessment: a many-facet Rasch measurement analysis. *Language Testing in Asia*, 10(1), 1-23. <https://doi.org/10.1186/s40468-020-00115-0>
- Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of higher education*, 6(5), 26-41.
- Knecht, R., Larson, L., & Townsend, D. (2023). Exploring teacher and student knowledge of sentence-level language features. *Journal of Adolescent & Adult Literacy*, 66(6), 344-354. <https://doi.org/10.1002/jaal.1290>

- Lam, R. (2016). Assessment as learning: examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in higher education*, 41(11), 1900-1917. <https://doi.org/10.1080/03075079.2014.999317>
- Lee, Y., & De La Paz, S. (2021). Science writing intervention research for students with and at risk for learning disabilities, and English learners: A systematic review. *Learning Disability Quarterly*, 44(4), 261-274. <https://doi.org/10.1177/073194872111018213>
- Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic theories. *Organizational Research Methods*, 21(3), 592-632. <https://psycnet.apa.org/doi/10.1177/1094428117728372>
- Mallinckrodt, B., Miles, J. R., & Recabarren, D. A. (2016). Using focus groups and Rasch item response theory to improve instrument development. *The Counseling Psychologist*, 44(2), 146-194. <https://psycnet.apa.org/doi/10.1177/0011000015596437>
- McKinley, J. (2015). Critical argument and writer identity: Social constructivism as a theoretical framework for EFL academic writing. *Critical inquiry in language studies*, 12(3), 184-207. <https://doi.org/10.1080/15427587.2015.1060558>
- Mercer-Mapstone, L., & Kuchel, L. (2015). Teaching scientists to communicate: Evidence-based assessment for undergraduate science education. *International Journal of Science Education*, 37(10), 1613-1638. <https://doi.org/10.1080/09500693.2015.1045959>
- Mills, S. (2016). Conceptual understanding: A concept analysis. *The Qualitative Report*, 21(3), 546-557. <https://doi.org/10.46743/2160-3715/2016.2308>
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51(1), 59-81. <https://psycnet.apa.org/doi/10.1080/00461520.2016.1145550>
- Phillips Galloway, E., Qin, W., Uccelli, P., & Barr, C. D. (2020). The role of cross-disciplinary academic language skills in disciplinary, source-based writing: Investigating the role of core academic language skills in science summarization for middle grade writers. *Reading and Writing*, 33(1), 13-44. <https://psycnet.apa.org/doi/10.1007/s11145-019-09942-x>
- Rahman, M. M., Hossain, M. A., & Uddin, M. M. (2024). Challenges in Academic Writing for Undergraduate Students in Bangladesh. *European Journal of Teaching and Education*, 6(4), 49-77. <https://doi.org/10.33422/ejte.v6i4.1418>
- Sala-Bubaré, A., & Castelló, M. (2018). Writing regulation processes in higher education: a review of two decades of empirical research. *Reading and Writing*, 31(4), 757-777. <https://doi.org/10.1007/s11145-017-9808-3>

- Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies*, 3(1), 173-182. <https://doi.org/10.1080/23752696.2018.1433546>
- Silva, B. (2022). The writing cycle and cognitive processes that may affect learning. In *Writing to Learn Academic Words: Assessment, Cognition, and Learning* (pp. 45-62). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-06505-7\\_3](https://doi.org/10.1007/978-3-031-06505-7_3)
- Stiller, J., Hartmann, S., Mathesius, S., Straube, P., Tiemann, R., Nordmeier, V., ... & zu Belzen, A. U. (2018). Assessing scientific reasoning: A comprehensive evaluation of item features that affect item difficulty. In *Assessing Student Learning Outcomes in Higher Education* (pp. 67-78). Routledge. <http://dx.doi.org/10.1080/02602938.2016.1164830>
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & education*, 131, 33-48. <https://doi.org/10.1016/j.compedu.2018.12.005>
- Subandowo, D., & Sardi, C. (2023). Academic essay writing in an English medium instruction environment: Indonesian graduate students' experiences at Hungarian universities. *Ampersand*, 11, 100158. <https://doi.org/10.1016/j.amper.2023.100158>
- Suleimanova, O. A., & Lykova, T. A. (2023). Academic writing details in critical perspective. *Training, Language and Culture*, 7(3), 94-102. <https://doi.org/10.22363/2521-442X-2023-7-3-94-102>
- Teng, M. F., Qin, C., & Wang, C. (2022). Validation of metacognitive academic writing strategies and the predictive effects on academic writing performance in a foreign language context. *Metacognition and learning*, 17(1), 167-190. <https://doi.org/10.1007/s11409-021-09278-4>
- Tusting, K., McCulloch, S., Bhatt, I., Hamilton, M., & Barton, D. (2019). *Academics writing: The dynamics of knowledge creation*. Routledge. <https://doi.org/10.4324/9780429197994>
- Van Vo, D., & Csapó, B. (2021). Development of scientific reasoning test measuring control of variables strategy in physics for high school students: Evidence of validity and latent predictors of item difficulty. *International Journal of Science Education*, 43(13), 2185-2205. <https://psycnet.apa.org/doi/10.1080/09500693.2021.1957515>
- Wang, Y., Kasuma, S. A. B. A., Lah, S. B. C., & Zhang, Q. (2024). Exploring the relationship between EFL students' writing performance and activity theory related influencing factors in the blended learning context. *Plos one*, 19(6), e0305668. <https://doi.org/10.1371/journal.pone.0305668>
- Wette, R. (2017). Source text use by undergraduate post-novice L2 writers in disciplinary assignments: Progress and ongoing challenges. *Journal of Second Language Writing*, 37, 46-58. <https://doi.org/10.1016/j.jslw.2017.05.015>

- Wicaksono, R. A., Taufik, M., & Faqih, A. (2025). Cracking the Code of Academic Writing: Student Insights on Academic Writing Hurdles & Breakthroughs. *International Journal of Science, Management, Engineering and Information Technology*, 1(1), 36-46.
- Woods, A. D., Ahmed, S. F., Katz, B. D., & Morrison, F. J. (2020). How stable is early academic performance? Using cluster analysis to classify low achievement and EF. *Journal of learning disabilities*, 53(1), 19-35.  
<https://psycnet.apa.org/doi/10.1177/0022219419881632>
- Yalçın, S. (2018). Data fit comparison of mixture item response theory models and traditional models. *International Journal of Assessment Tools in Education*, 5(2), 301-313.  
<https://doi.org/10.21449/ijate.402806>
- Yang, R. Y., & Sang, Y. R. (2023). Concept-based language instruction and the teaching of citation in English academic writing. *The Modern Language Journal*, 107(S1), 46-67.  
<https://doi.org/10.1111/modl.12822>
- Zou, H., & Hyland, K. (2019). Reworking research: Interactions in academic articles and blogs. *Discourse studies*, 21(6), 713-733. <https://doi.org/10.1177/1461445619866983>