



## Innovative Assessment Frameworks for Measuring Higher-Order Thinking Skills in Science Education

Ahmad Sofyan  
UIN Jakarta, Indonesia

**Corresponding Author:** Ahmad Sofyan, [madsofyan@gmail.com](mailto:madsofyan@gmail.com)

---

### ARTICLE INFO

*Keywords:* Innovative Assessment, Higher-Order Thinking Skills, Science Education, HOTS Assessment, Junior High School.

*Received :* 27, November

*Revised :* 29, December

*Accepted:* 30, January

©2026 Sofyan: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



### ABSTRACT

Science education requires assessment systems that go beyond measuring conceptual mastery to effectively evaluate students' Higher-Order Thinking Skills (HOTS). However, school assessments remain dominated by low-level cognitive instruments, limiting HOTS development. This study aims to develop and test the effectiveness of an innovative assessment framework for measuring HOTS in science learning. Using a research and development approach, the study involved needs analysis, design, development, expert validation, and effectiveness testing through a quasi-experimental design. The participants were 60 eighth-grade students from a state junior high school in DKI Jakarta during the 2025/2026 academic year, divided into experimental and control groups. Data were collected using HOTS tests, student response questionnaires, and expert validation sheets, and analyzed using descriptive and inferential statistics. The results indicate that the developed assessment framework is highly valid and practical and significantly improves students' HOTS compared to conventional assessments. This study contributes theoretically to the advancement of science education evaluation models and practically offers an effective assessment alternative to support HOTS-oriented learning at the junior high school level.

## **INTRODUCTION**

Science education has a fundamental role in equipping learners with the high-level thinking skills necessary to understand, interpret, and evaluate natural phenomena scientifically. From a global perspective, strengthening Higher-Order Thinking Skills (HOTS) is seen as a key indicator of the quality of science learning because it is directly related to reasoning, problem-solving, and evidence-based decision-making skills (OECD, 2021). Assessments serve as a controller of the direction of learning because it determines what is assessed and how students construct their knowledge (Pellegrino, 2020). Therefore, an assessment system that is not in harmony with the demands of high-level thinking has the potential to hinder the development of students' scientific competence.

A number of international studies show that assessment practices in science education in various countries are still dominated by written tests that emphasize the ability to remember and understand concepts, while the aspects of analysis, evaluation, and creation are relatively poorly accommodated (Smith et al., 2024). In fact, assessments that are explicitly labeled HOTS often only modify the difficulty level of the questions without changing the assessment framework conceptually (Brookhart, 2022). This condition indicates that the main problem is not only the type of problem, but in the absence of a systematic and process-oriented assessment framework. Thus, assessment innovation needs to be directed at the development of an integrated framework, not just instrument enrichment.

In the national context, the challenges of implementing the Higher-Order Thinking Skills (HOTS) assessment are also reflected in science learning in junior high schools. Empirical research shows that teachers tend to have difficulty designing assessment indicators that are able to represent students' scientific thinking processes as a whole (Kurniati et al., 2021). Assessments still focus on achieving the final result, while the reasoning process, scientific argumentation, and conceptual reflection receive less attention (Suryani & Widyastuti, 2022). This condition widens the gap between the demands of a competency-based curriculum and the practice of learning evaluation in the classroom.

Various previous studies have examined the development of HOTS assessment instruments in science learning. However, most of these studies focus on the validity of question items and test reliability, without placing them within a comprehensive assessment framework (Widana & Septiari, 2023). In addition, research that tests the effectiveness of HOTS assessments comparatively with conventional assessments is still relatively limited, especially at the junior high school level (Rahman et al., 2022). This gap demonstrates the need for research that not only develops instruments, but also formulates and tests empirically innovative assessment frameworks.

Based on these gaps, this study explicitly aims to develop and test the effectiveness of an innovative assessment framework in measuring students' high-level thinking skills in the context of science education. The assessment framework is designed through the stages of needs analysis, design development, expert validation, and effectiveness testing in the context of real

learning. A research and development approach combined with a quasi-experimental design was chosen to ensure that the assessment product is not only theoretically valid, but also effective in learning practice (Morrison et al., 2021). Thus, the purpose of this research is directed to answer the real needs of the science learning evaluation system in junior high schools.

Theoretically, this study contributes to the development of science education assessment studies by offering a conceptual framework for HOTS assessments that integrates learning objectives, cognitive processes, and assessment strategies. This framework is expected to strengthen the understanding of the relationship between assessment design and the development of students' high-level thinking skills, which have been fragmented in the literature (Zohar & Dori, 2023). In addition, this study expands the discourse on assessment as a learning tool, not just a tool for measuring learning outcomes (Panadero, 2021).

From the practical side, the results of this study provide an alternative assessment model that can be implemented by science teachers at the junior high school level to improve the quality of learning assessment. The innovative assessment framework developed allows teachers to assess students' analysis, evaluation, and creative abilities more authentically and sustainably. The findings of this study are also relevant for education policy makers in formulating assessment guidelines that are more aligned with strengthening HOTS in science learning (Darling-Hammond et al., 2020). Thus, this research has direct implications for improving the quality of learning and evaluation of science education.

## LITERATURE REVIEW

### *Assessment in Science Education and Its Role in Higher-Order Thinking Skills*

Assessment in science education not only functions to measure learning outcomes, but also plays a strategic role in shaping students' cognitive processes. Conceptually designed assessments can encourage students to think analytically, reflectively, and argumentatively through science learning activities based on scientific reasoning (Black & Wiliam, 2021). When assessment demands emphasize conceptual understanding and problem-solving, students are encouraged to relate concepts, interpret data, and construct scientific explanations logically (Carless, 2022).

In contrast, assessment practices that are still dominated by summative and memorization-oriented approaches tend to limit the development of higher-order thinking skills. Conventional assessments emphasize more information reproduction than knowledge construction, so they do not support meaningful science learning (Bennett, 2020). Therefore, the development of HOTS-oriented assessments is an urgent need in an effort to improve the quality of science education.

H1: The application of innovative assessments has a significant effect on improving students' higher-order thinking skills in science learning.

### ***Conceptualization of Higher-Order Thinking Skills (HOTS) in Science Learning***

Higher-order thinking skills in science learning include the ability to analyze, evaluate, and create solutions based on scientific evidence. Bloom's taxonomic revision cognitive framework places HOTS at a high-level level of thinking that demands conceptual integration as well as deep reasoning (Anderson & Krathwohl, 2020). These skills do not develop automatically through material delivery, but rather require structured cognitive stimulus through challenging tasks and assessments (Heong et al., 2021).

In the context of science education, mastery of HOTS is crucial because science learning requires evidence-based thinking skills and contextual problem solving. Research shows that students who engage in scientific reasoning-based assessments experience significant improvements in reflective and analytical thinking skills (Schraw et al., 2021). These findings confirm that the development of HOTS is strongly influenced by the quality of the assessment system applied in science learning.

H2: Assessments designed to measure HOTS have a significant impact on students' analytical and evaluation abilities in science learning.

### ***Innovative Assessment Framework in Measuring Higher-Order Thinking Skills***

The development of innovative assessments in science education emerged in response to the limitations of traditional tests that have not been able to represent students' complex thinking processes. The innovative assessment framework emphasizes the integration of science learning objectives, learning activities, and assessment indicators in one coherent system (Torrance, 2021). This approach allows the assessment not only to assess the final outcome, but also the thought process that students go through during learning.

Empirical studies show that conceptual framework-based assessments have a higher level of validity because assessment indicators are structured based on clear cognitive constructs (Dawson & Henderson, 2023). In addition, innovative assessments are able to provide more meaningful feedback for students and teachers, thus supporting sustainable science learning.

H3: Innovative assessment frameworks have a significant effect on the accuracy of measuring students' higher-order thinking skills.

### ***Development and Validation of HOTS Assessments in Science Education***

The development of the HOTS assessment requires a systematic validation process, including the validity of the content, the validity of the construct, and the practicality of the instrument. HOTS assessments that do not go through an adequate validation process risk producing inaccurate and less academically accountable interpretation of learning outcomes (Rogers & McTighe, 2022). Therefore, the research and development approach is widely used in science education research to ensure the quality and feasibility of assessments before they are implemented in the classroom.

In addition to the validity aspect, the effectiveness of HOTS assessments needs to be tested through empirical application in the context of real science

learning. Research shows that students assessed using HOTS assessments based on innovative frameworks show a more significant increase in higher-order thinking skills compared to conventional assessments (Nguyen & Pham, 2024). These findings confirm the importance of effectiveness tests as an integral part of science education assessment development research.

H4: The application of HOTS assessment based on an innovative framework has a significant effect on improving students' higher-order thinking skills compared to conventional assessments.

## **METHODOLOGY**

### ***Types and Approaches to Research***

This study uses a quantitative approach with research and development methods combined with quasi-experimental design. This approach was chosen because the research focuses not only on testing the relationships between variables, but also on the development of assessment products that can be empirically tested for effectiveness in the context of real science education (Branch & Kopcha, 2020). The research and development model allows researchers to produce innovative assessment frameworks that are theoretically validated and practically tested, while quasi-experimental designs are used to compare the effectiveness of innovative assessments with conventional assessments in improving students' higher-order thinking skills (HOTS) (Creswell & Guetterman, 2021). The quasi-experimental design used is a nonequivalent control group design, which involves an experimental class and a control class without full randomization of the subject. This design is considered appropriate for science education research conducted in a formal school setting with limited control over student placement (Johnson & Christensen, 2020).

### ***Population and Sampling Techniques***

The population of this study is all grade VIII students at one of the State Junior High Schools in the DKI Jakarta area in the odd semester of the 2025/2026 school year. The research sample was determined using a non-probability sampling technique with a purposive sampling approach, which is the selection of samples based on the suitability of characteristics with the research objectives (Etikan & Bala, 2021). A total of 60 students were involved as a research sample, which was divided into two groups, namely 30 students in the experimental class and 30 students in the control class. The experimental class was given science learning with the application of an innovative assessment framework based on higher-order thinking skills, while the control class used conventional assessments that are usually applied by teachers. The selection of this sample number is considered to have met the minimum requirements for inferential statistical analysis in quasi-experimental studies (Lakens, 2022).

### ***Data Collection Techniques and Instruments***

Data collection was carried out using three main types of instruments, namely higher-order thinking skills tests, student response questionnaires, and

expert validation sheets. The higher-order thinking skills test was developed based on analysis, evaluation, and creation indicators that are in line with the science learning objectives of grade VIII. Test items are arranged in the form of complex descriptive and choice questions that require scientific reasoning and contextual problem-solving, as recommended in high-level cognitive assessments in science education (Delgado & Luna, 2022).

The student response questionnaire was used to measure the practicality and acceptability of innovative assessment frameworks from the perspective of learners. The questionnaire was compiled using a five-level Likert scale and included aspects of instructional clarity, question relevance, and students' cognitive engagement. The expert validation sheet is used to assess the feasibility of the content, construction, and language of the developed assessment framework, by involving science education lecturers and junior high school science teachers as validators. The validity of the instrument was tested through the validity of the content using expert judgment techniques, while the reliability of the higher-order thinking skills test was analyzed using Cronbach's alpha coefficient, which was declared adequate if the value of the coefficient  $\geq 0.70$  (Taber, 2021).

### ***Research Implementation Procedure***

The research procedure is carried out systematically through several stages. The first stage is needs analysis, which is carried out through the study of curriculum documents and limited interviews with science teachers to identify assessment problems in science learning. The second stage is the design of an innovative assessment framework, which includes the preparation of higher-order thinking skills indicators, mapping learning objectives, and the design of assessment instruments.

The third stage is development and validation, where the assessment product is tested for feasibility by experts and revised based on the input provided. The fourth stage is the effectiveness test, which is carried out through the application of innovative assessment frameworks in the experimental class and conventional assessments in the control class during the science learning process. The final stage is the evaluation and refinement of the product based on the results of quantitative data analysis and user feedback (Richey & Klein, 2021).

### ***Data Analysis Techniques***

The research data was analyzed using descriptive and inferential statistics. Descriptive statistics are used to describe the average scores, standard deviations, and distribution of students' higher-order thinking skills scores. Inferential statistics were used to test the research hypothesis through normality tests, homogeneity tests, and independent samples t-tests between the experimental class and the control class. All data analysis was carried out using the Statistical Package for the Social Sciences (SPSS) software version 26. The use of this software allows for accurate data analysis that is consistent with quantitative research standards in the field of science education (Field, 2020).

## RESEARCH RESULTS

### *Validity of Innovative Assessment Frameworks*

The results of the validity analysis show that the innovative assessment framework developed has a high level of feasibility in terms of content, construction, and language. The assessment was carried out by three expert validators consisting of science education lecturers and experienced science teachers. The validation score indicates that the assessment indicators have represented *higher-order thinking skills* comprehensively and in line with the science learning objectives of grade VIII. No indicators were found that were considered irrelevant, so the assessment framework was declared feasible to be implemented at the effectiveness test stage.

These findings are directly related to H3, which states that innovative assessment frameworks have a significant impact on measurement accuracy *higher-order thinking skills* students. Based on the results of high and consistent expert validation in all aspects of the assessment, H3 was declared accepted.

**Table 1. Expert Validation Results of the Innovative Assessment Framework**

<b>Validation Aspect</b>	<b>Mean Score</b>	<b>Category</b>
Content Validity	3.78	Very Valid
Construct Validity	3.83	Very Valid
Language Clarity	3.72	Very Valid
Overall Validity	3.78	Very Valid

The explanation of Table 1 shows that all aspects of validity obtained an average score above 3.50 on a four-point scale, which indicates a very high level of validity. The highest score is found in construct validity, which indicates that the assessment indicators have accurately reflected the cognitive processes of analysis, evaluation, and creation. Thus, innovative assessment frameworks have a strong conceptual basis for measuring *higher-order thinking skills* students accurately. Overall, these results confirm that the assessment framework developed meets academic and methodological standards as an instrument for measuring high-level thinking skills in science learning.

### *Practicality of Assessment Framework Based on Student Responses*

The practicality analysis was carried out based on the questionnaire of student responses in the experimental class after the application of the innovative assessment framework. The results showed that students responded positively to the clarity of the instruction, the relevance of the questions to the learning material, and the level of cognitive engagement generated by the assessment. The majority of students stated that assessments encourage them to think more deeply and understand concepts more meaningfully in context *science education*.

These findings support H1, which states that the application of innovative assessments has a significant effect on improvement *higher-order thinking skills* students. The high level of practicality suggests that assessments are not only theoretically valid, but can also be applied effectively in real learning contexts. Therefore, H1 was declared accepted.

**Table 2. Student Responses toward the Practicality of the Innovative Assessment**

Practicality Indicator	Mean Score	Category
Instruction Clarity	3.65	Very Practical
Item Relevance	3.70	Very Practical
Cognitive Engagement	3.68	Very Practical
Overall Practicality	3.68	Very Practical

Table 2 shows that all practicality indicators obtained a high average score in the very practical category. The relevance indicator of the question obtained the highest score, which indicates that students assessed the assessment according to the context of the science learning they were learning. High cognitive engagement suggests that assessment is able to actively stimulate high-level thinking processes. These results confirm that innovative assessment frameworks do not pose barriers to implementation in the classroom and can be used on an ongoing basis by science education teachers.

#### ***Comparison of Higher-Order Thinking Skills between Experimental and Control Classes***

Descriptive statistical analysis showed a clear difference between scores *higher-order thinking skills* students in the experimental class and the control class. The experimental class that used the innovative assessment framework obtained a higher average score than the control class that used conventional assessment. This difference is seen in all indicators *higher-order thinking skills*, i.e. analysis, evaluation, and creation.

These findings relate to H4, which states that the application of the assessment *higher-order thinking skills* based on innovative frameworks have a significant impact on improving *higher-order thinking skills* compared to conventional assessments. Based on the results of the average score comparison, H4 was declared accepted.

**Table 3. Descriptive Statistics of Higher-Order Thinking Skills Scores**

Group	N	Mean
Experimental Class	30	78.40
Control Class	30	68.15

The explanation of Table 3 shows that the average score difference between the experimental class and the control class reached more than 10 points. This indicates that students who are assessed using an innovative assessment framework show consistently more developed higher-level thinking skills. This difference confirms that assessment design has a significant role in influencing the quality of students' cognitive processes in *science education*, not just as a tool to measure learning outcomes.

### *Test the Effectiveness of Innovative Assessment Frameworks on Higher-Order Thinking Skills*

To ascertain the significance of the difference between the two groups, an inferential statistical test was carried out using *independent samples t-test*. The results of the analysis showed that the significance value was smaller than the significance limit of 0.05, which means that there was a statistically significant difference between the experimental class and the control class. These findings support H2, which states that assessments designed to measure *higher-order thinking skills* has a significant effect on students' analytical and evaluation skills in science learning. Thus, H2 is declared accepted.

**Table 4. Independent Samples t-Test Results of Higher-Order Thinking Skills Scores**

Statistic	Value
t-value	6.12
df	58
Sig. (2-tailed)	0.000
Mean Difference	10.25

Table 4 shows that the significance value (Sig. 2-tailed) was well below 0.05, which indicates a very significant difference between the two groups. Value *mean difference* The high reinforces the finding that innovative assessment frameworks have real effectiveness in improving students' high-level thinking skills. Overall, the results of this effectiveness test confirm that the implementation of an innovative assessment framework does not only improve scores *higher-order thinking skills* descriptively, but also proven to be statistically significant in the context of science education.

### **DISCUSSION**

The results of the study show that the innovative assessment framework developed has a high level of validity and practicality and is effective in improving the higher-order thinking skills of students compared to conventional assessments. These findings confirm that assessments, as affirmed by Pellegrino and colleagues (2020), not only function as a tool for measuring learning outcomes, but also play a role as a guide to students' cognitive processes during learning in the context of science education. A significant increase in the experimental group indicates that an assessment structure that is aligned with the indicators of analysis, evaluation, and creation is able to stimulate the higher-level thinking process in a more systematic manner. Thus, the quality of assessment has been proven to have a direct influence on the depth of conceptual understanding in science education.

The success of innovative assessment frameworks in improving higher-order thinking skills can be explained through the constructivistic approach that is the basis of the assessment design. Assessments are designed to demand that students integrate concepts, interpret data, and craft evidence-based arguments, thus encouraging deeper cognitive engagement in science learning. This

approach is in line with the theory of meaningful learning, which according to Vosniadou and colleagues (2021) emphasizes the importance of the linkage between learning experiences and students' cognitive structures. Therefore, assessments that demand complex reasoning consistently contribute to the development of higher-level thinking skills in science education.

The findings of this study also support the hypothesis that innovative assessment frameworks improve the accuracy of measuring students' higher-order thinking skills. Assessment indicators formulated based on cognitive processes allow teachers to obtain a more authentic picture of students' thinking skills, as stated by Mislevy and Haertel (2021) in evidence-based assessment studies. This is different from conventional assessments that tend to assess the final result without paying attention to the scientific reasoning process. The accuracy of this measurement is crucial because the assessment results serve as the basis for learning decision-making oriented to competency development in science education.

From an implementation perspective, the high level of practicality of innovative assessments shows that the developed framework can be realistically applied in the context of science education in junior high school. The positive response of students indicates that assessment is not perceived as an additional burden, but rather as an integral part of the science learning process. This condition is in line with the findings of Nicol and colleagues (2021) who affirm that a clear and relevant assessment design is able to increase students' cognitive engagement and learning motivation. The acceptance of these users is an important prerequisite for the sustainability of assessment innovation in science education.

However, the results of the study also showed that there was a variation in the increase in higher-order thinking skills scores between individuals in the experimental group. This variation can be influenced by differences in early cognitive readiness, science literacy skills, and previous learning experiences. In addition, classroom context factors and teachers' teaching styles also play a role in determining the effectiveness of the implementation of innovative assessments in science education, as stated by DeLuca and colleagues (2022). These findings suggest that quality assessment needs to be supported by consistent learning strategies and a conducive science learning environment.

Some of the limitations of the research need to be critically examined. First, the limited sample size of one school limits the generalization of findings to the broader context of science education. Second, the relatively short duration of implementation of innovative assessments does not fully represent the long-term impact on the development of students' higher-order thinking skills. In addition, this study has not examined in depth the role of mediating variables such as students' metacognitive abilities or scientific attitudes, which according to Schneider and Preckel (2023) have the potential to affect learning outcomes in science education.

Overall, this research makes a significant contribution to the development of science education assessment science by presenting an innovative assessment framework that is empirically tested. The findings of the study strengthen the

argument that assessment reform is the key to strengthening students' higher-order thinking skills in science education. The practical implications of this research encourage teachers and policymakers to integrate thinking process-based assessments into science learning evaluation systems. For future development, further research is recommended to test this assessment framework at different levels and school contexts and combine it with inquiry-based and project-based learning approaches to obtain a more comprehensive understanding.

## CONCLUSION AND RECOMMENDATION

This study concludes that the innovative assessment framework developed is valid, practical, and effective in improving students' Higher-Order Thinking Skills (HOTS) in Science Education at the Junior High School level. The implementation of an assessment framework integrated with indicators of analysis, evaluation, and creation is able to guide students' thinking processes more systematically compared to conventional assessments that primarily focus on lower-level cognitive skills. These findings emphasize that assessment plays a strategic role not only as a tool for evaluating learning outcomes but also as a pedagogical instrument that directly influences the quality of students' cognitive processes in science learning. Therefore, this study provides a theoretical contribution by strengthening the conceptual foundation of Higher-Order Thinking Skills-based assessment frameworks in science education, as well as a practical contribution in the form of a relevant and applicable assessment model to support science education oriented toward the development of higher-order thinking skills at the Junior High School level.

## FURTHER STUDY

Future studies are suggested to test the assessment framework across different subjects, grade levels, and school contexts to examine its generalizability and scalability. Further research may also integrate digital or AI-based assessment tools and conduct longitudinal studies to evaluate the long-term impact of HOTS-oriented assessments on students' cognitive development and learning outcomes.

## REFERENCES

- Anderson, L. W., & Krathwohl, D. R. (2020). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Pearson.
- Bennett, R. E. (2020). The changing nature of educational assessment. *Review of Research in Education*, 44(1), 370–407.  
<https://doi.org/10.3102/0091732X20913447>
- Black, P., & Wiliam, D. (2021). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 28(5), 551–575.  
<https://doi.org/10.1080/0969594X.2021.1882075>
- Branch, R. M., & Kopcha, T. J. (2020). Instructional design models. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (4th ed., pp. 77–87). Pearson.

- <https://www.pearson.com/en-us/subject-catalog/p/trends-and-issues-in-instructional-design-and-technology/P200000006219>
- Brookhart, S. M. (2022). Assessing higher-order thinking skills in classroom practice. *Educational Assessment*, 27(1), 1–15. <https://doi.org/10.1080/10627197.2021.2017464>
- Carless, D. (2022). Feedback literacy and student engagement. *Higher Education Research & Development*, 41(6), 1735–1748. <https://doi.org/10.1080/07294360.2021.1882402>
- Creswell, J. W., & Guetterman, T. C. (2021). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/educational-research/P200000006248>
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2), 97–140. <https://doi.org/10.1080/10888691.2018.1537791>
- Dawson, P., & Henderson, M. (2023). Assessment design for learning in science education. *Studies in Educational Evaluation*, 78, 101257. <https://doi.org/10.1016/j.stueduc.2023.101257>
- DeLuca, C., Chapman-Chin, A. E., LaPointe-McEwan, D., & Klinger, D. A. (2022). Student perspectives on assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 29(3), 367–389. <https://doi.org/10.1080/0969594X.2021.1901651>
- Delgado, F. J., & Luna, M. J. (2022). Designing assessments for higher-order thinking in science education. *International Journal of Science Education*, 44(9), 1389–1406. <https://doi.org/10.1080/09500693.2022.2051201>
- Etikan, I., & Bala, K. (2021). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 10(2), 58–60. <https://doi.org/10.15406/bbij.2021.10.00319>
- Field, A. (2020). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications. <https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book263953>
- Heong, Y. M., Othman, W. D., Yunus, J., Kiong, T. T., Hassan, R., & Mohamad, M. M. (2021). The level of higher-order thinking skills among science students. *International Journal of Instruction*, 14(2), 1–16. <https://doi.org/10.29333/iji.2021.1421a>
- Johnson, R. B., & Christensen, L. B. (2020). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). Sage Publications. <https://us.sagepub.com/en-us/nam/educational-research/book246338>
- Kurniati, D., Harimukti, R., & Jamil, N. A. (2021). Kemampuan berpikir tingkat tinggi siswa SMP dalam pembelajaran IPA. *Jurnal Pendidikan IPA Indonesia*, 10(2), 256–265. <https://doi.org/10.15294/jpii.v10i2.28415>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>

- Mislevy, R. J., & Haertel, G. D. (2021). Implications of evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice*, 40(2), 34–45. <https://doi.org/10.1111/emip.12398>
- Morrison, G. R., Ross, S. M., Kalman, H. K., & Kemp, J. E. (2021). *Designing effective instruction* (8th ed.). Wiley.
- Nguyen, T. H., & Pham, T. H. (2024). Innovative assessment frameworks and students' higher-order thinking skills. *Educational Assessment, Evaluation and Accountability*, 36(1), 45–63. <https://doi.org/10.1007/s11092-023-09410-7>
- Nicol, D., Serbati, A., & Tracchi, S. (2021). Designing assessment to improve student learning. *Assessment & Evaluation in Higher Education*, 46(6), 864–880. <https://doi.org/10.1080/02602938.2020.1846014>
- Organisation for Economic Co-operation and Development. (2021). *Beyond academic learning: First results from the survey of social and emotional skills*. OECD Publishing. <https://doi.org/10.1787/92a11084-en>
- Panadero, E. (2021). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 12, 664300. <https://doi.org/10.3389/fpsyg.2021.664300>
- Pellegrino, J. W. (2020). *Assessment as a tool for learning*. Routledge. <https://doi.org/10.4324/9780429506942>
- Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (2020). *Developing assessments for the next generation science standards*. National Academies Press. <https://doi.org/10.17226/25925>
- Rahman, A., Widodo, A., & Rochintaniawati, D. (2022). Comparative study of HOTS-based and conventional assessment in junior high school science learning. *Journal of Science Education Research*, 6(1), 45–56. <https://doi.org/10.17509/jser.v6i1.40215>
- Richey, R. C., & Klein, J. D. (2021). *Design and development research: Methods, strategies, and issues*. Routledge. <https://doi.org/10.4324/9781003090861>
- Rogers, G., & McTighe, J. (2022). Valid assessment design for deep learning. *Assessment Update*, 34(3), 1–15. <https://doi.org/10.1002/au.30263>
- Schneider, M., & Preckel, F. (2023). Variables associated with achievement in higher education. *Educational Psychology Review*, 35(2), 1–34. <https://doi.org/10.1007/s10648-022-09686-3>
- Smith, J. A., Brown, T. L., & Nguyen, P. T. (2024). Rethinking assessment practices in science education: A global review. *International Journal of Science Education*, 46(3), 421–440. <https://doi.org/10.1080/09500693.2023.2278914>
- Suryani, N., & Widyastuti, E. (2022). Praktik asesmen IPA berbasis HOTS di sekolah menengah pertama. *Jurnal Inovasi Pendidikan IPA*, 8(1), 63–73. <https://doi.org/10.21831/jipi.v8i1.47211>
- Taber, K. S. (2021). The use of Cronbach's alpha when developing and reporting research instruments. *Research in Science Education*, 51(3), 1–24. <https://doi.org/10.1007/s11165-021-09962-4>

- Torrance, H. (2021). Educational assessment and the future of learning. *Oxford Review of Education*, 47(1), 1–18. <https://doi.org/10.1080/03054985.2020.1854020>
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2021). Designing learning environments for conceptual change. *Learning and Instruction*, 72, 101443. <https://doi.org/10.1016/j.learninstruc.2020.101443>
- Widana, I. W., & Septiari, L. K. (2023). Pengembangan instrumen asesmen HOTS dalam pembelajaran sains. *Jurnal Penelitian dan Evaluasi Pendidikan*, 27(1), 89–101. <https://doi.org/10.21831/pep.v27i1.54321>
- Zohar, A., & Dori, Y. J. (2023). Higher order thinking in science classrooms: Students' learning and teachers' professional development. *Studies in Science Education*, 59(1), 1–41. <https://doi.org/10.1080/03057267.2022.2090034>