JPPIPA 12 (7) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of a Three-Tier Higher Order Thinking Skills (HOTS) Assessment Instrument on Acid Base Material Assisted by Gimkit

Ananda Emyliah[1*], Jusniar[1], Ramlawati[1]

[1] Postgraduate Master's Program of Chemistry Education, Faculty of Mathematics and Sains, Makassar University, Indonesia.

Received: March 12, 2026
Revised: June 25, 2026
Accepted: July 25, 2026
Published: July 31, 2026

Corresponding Author:
Ananda Emyliah
nemy3118@gmail.com

DOI: 10.29303/jppipa.v12i7.15993

Abstract: This study aims to describe the development process of a HOTS three-tier assessment instrument on acid-base material assisted by Gimkit and to find out the validity, reliability, and practicality of the HOTS three-tier assessment instrument on acid-base material assisted by Gimkit that was developed. The data source for the validity test came from 2 validators, while the reliability and practicality tests came from 2 subject teachers and 31 11th-grade students of Merdeka 2 at SMA Negeri 1 Campalagian. The reliability and practicality trials were conducted in the even semester of the 2025/2026 academic year. The research method used was a research and development (R&D) with the Plomp model, which includes the preliminary investigation phase through observation and interviews, the design and realization phase covering the design and creation of the HOTS three-tier assessment instrument on acid-base material, and the test, evaluation, and revision phase including validity tests and limited trials to find out reliability and practicality. The developed instruments consist of: (1) instrument grid, (2) HOTS three-tier assessment instrument, (3) answer key, (4) validity sheet, and (5) perception questionnaire. The results show that the HOTS three-tier assessment instrument product is highly valid with an inter-rater reliability value of 1.00, reliable, and very practical. Based on the results, it can be concluded that the instrument can be an important reference for teachers in assessing students' cognitive aspects and misconceptions.

Keywords: Acid and base; Gimkit; HOTS; Instrument development; Three-tier

Introduction

The world of education in the 21st century is at a point where knowledge is advancing and improving incredibly. Learning styles nowadays need to match the current times, where learning materials should offer a more authentic design that challenges students to collaborate, solve problems, and create solutions (Jihannita et al., 2023). Using technology today is needed to develop various skills in students, especially referring to the 21st-century skills outlined in the 4Cs, which are: (1) Critical thinking, where students use logical and rational thinking to solve contextual problems, (2) Creativity, where students create and find solutions, (3) Collaboration, where students improve their ability to work in teams, and (4) Communication, where students are encouraged to communicate, share ideas, and argue effectively (Dwijayanti, 2021).

The results of the 2022 PISA study in the field of science showed an average score of 383, ranking 69th out of 80 countries (OECD, 2023). This indicates that students' performance in TIMSS and PISA is still relatively low (Surayanah et al., 2023). This is of course caused by several factors, one of which is that students in Indonesia are not yet trained to solve TIMSS and PISA-type questions, which have contextual content characteristics and require reasoning, argumentation, and creativity to solve (Ariani et al., 2025).

HOTS is needed to solve factual problems that are also challenges in the 21st century (Andriyatno et al.,

How to Cite:
Emyliah, A., Jusniar, & Ramlawati. (2026). Development of a Three-Tier Higher Order Thinking Skills (HOTS) Assessment Instrument on Acid Base Material Assisted by Gimkit . *Jurnal Penelitian Pendidikan IPA*, *12*(7), 450–456. https://doi.org/10.29303/jppipa.v12i7.15993

2023). This is because HOTS can measure and develop students higher-order thinking skills, which are one of the important competencies in the 21st century (Zebua, 2024). HOTS skills include the ability to analyze, evaluate, and create creatively to solve problems. HOTS assessment instruments based on contextual problems and multiple representations encourage students to explore information and apply concepts in the real world to solve problems (An-Nisa et al., 2023). Therefore, planning, implementing, and assessing learning that can train students to improve their higher-order thinking skills is needed. To achieve this, HOTS assessment instruments are required to enhance higher-order thinking skills.

Therefore, teachers need to actively introduce and implement various forms of representative assessments, such as problem-solving assessments. Developing HOTS assessment instruments is necessary to make students more capable of tackling problems. The HOTS assessment instruments developed not only serve as measurement tools but also help train students' higher-order thinking skills so that they get used to facing problems and solving them creatively (Muliyadi et al., 2026; Saepuzaman et al., 2025). Higher-order thinking allows students to process information logically, determine which information is important, irrelevant, or useless, and produce clear perceptions, visions, and logical methods (Baihaki et al., 2021; Susilawati et al., 2021). HOTS assessment instruments are made in the form of three-tier multiple-choice questions, because this type of multiple-choice test has high validity and reliability and minimal errors.

Three-tier multiple-choice questions consist of three tiers. The first tier contains questions about the tested concepts, the second tier includes reasons for each answer to the first-tier questions, and the third tier (Sundari et al., 2021) relates to the students' confidence. The first tier is about students' knowledge, the second tier aims to enhance students' higher-order thinking skills, and the third-tier addresses students' confidence (Umroh et al., 2024).

One innovation in education is the use of digital platforms as a more interactive and enjoyable learning experience. With a variety of platforms available, Gimkit stands out as a tool that can turn quizzes into more immersive experiences, motivate actions, and make learning instruments more fun. Gimkit is a learning game platform designed to boost student engagement in the learning process. It combines learning elements with game elements, creating an interactive and enjoyable experience for students (Sugiyati et al., 2025). Besides increasing student motivation and curiosity, Gimkit is also great for providing quick and accurate feedback, making it one of the platforms that can help analyze assessment results. Gimkit also has advantages, such as its features that can directly connect to data analysis tools for use. This makes it even easier to analyze assessment results (Yuanta et al., 2025). Therefore, using Gimkit can improve the efficiency and effectiveness of the overall assessment process.

The HOTS assessment instrument developed in this study takes the form of third-tier multiple-choice questions with the help of Gimkit. The third-tier multiple-choice assessment instrument has several advantages, such as having a high level of objectivity, being effective, and being able to measure higher-order thinking skills. To meet valid, reliable, and practical requirements, the HOTS assessment instrument developed must be validated by experts and tested on a limited scale.

Method

This research uses research and development (R&D) methods with a Plomp development model with four stages are the focus of this research. The first stage is a preliminary investigation the initial investigation phase through observation and interviews, the design and implementation phase includes designing and creating HOTS three-tier assessment instruments on acid-base material, and the testing, evaluation, and revision phase includes validity testing and limited trials to know reliability and practicality. Through these four stages, the essence of development to produce a valid and reliable assessment instrument has already been met.

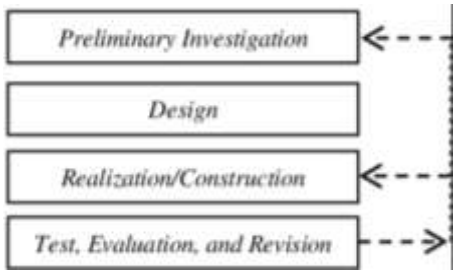

Figure 1. Flowchart of Plomp design development

For the product validity test, 2 expert validators will be used with a validation questionnaire, while for testing the reliability and practicality of the product, 2 teachers and 31 students will be involved using a perception questionnaire. The subjects for the limited trial are the 11th grade Merdeka 2 students at SMA Negeri 1 Campalagian who have already received lessons on acids and bases.

Result and Discussion

The development of the HOTS three-tier assessment instrument on acids and bases using Gimkit

refers to the Plomp development model, which consists of 4 phases: 1) the initial investigation phase, 2) the design phase, 3) the realization phase, and 4) the Test, Evaluation, and Revision phase. The assessment instrument developed refers to three quality criteria: valid, reliable, and practical. The process and results of the developed assessment instrument are as follows.

Preliminary Investigation

The initial investigation phase starts by identifying the problem through analyzing articles related to the HOTS three-tier assessment instrument. Based on the results of initial observations, it was found that the improvement in higher-order thinking skills is still very low. According to the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA), Indonesia is still ranked low (OECD, 2023). This situation shows that Indonesian students still have difficulty in solving problems that require reasoning, analysis, evaluation, and relatively simple problem-solving (Baihaki et al., 2021).

Curriculum analysis is done to identify and determine the cognitive aspects that need to be developed in students based on the school's applicable curriculum. This analysis becomes an important basis for developing assessment instruments, especially instruments designed to measure students' higher-order thinking skills.

Based on an interview with one of the teachers involved with the curriculum used at SMA Negeri 1 Campalagian, it was found that SMA Negeri 1 Campalagian uses the Merdeka curriculum, which is integrated with various policies and learning needs at the school. The implementation of this Merdeka curriculum focuses on developing students' competencies in cognitive aspects, such as problem-solving, creative thinking, and critical thinking.

This instrument was prepared based on the levels of cognitive taxonomy from Bloom, which were later revised by Anderson et al. (2001) which includes the ability to analyze (C4), evaluate (C5), and create (C6), as well as integrating them with three-tier multiple choice as the basis for developing the three-tier HOTS instrument assessment on acids and bases material framework. The three-tier HOTS assessment instrument developed consists of three main components: the first tier is a multiple-choice answer to the question, the second tier is a reason for the chosen answer, and the third tier is the students' confidence in the answer given. Through this structure, the developed instrument is expected to provide information and train students' higher-order thinking skills, especially in the acids and bases material, and serve as a basis for improving the quality of chemistry learning in schools.

Design

The second step is designing and preparing the three-tier HOTS assessment instrument for the acids and bases material that was developed. This consists of a three-tier HOTS assessment instrument blueprint created by looking at the learning objectives formulated from learning outcomes and question indicators. Once the blueprint is complete, a three-tier HOTS assessment instrument with 30 questions is made, with cognitive levels ranging from C4 to C6 according to the HOTS cognitive levels stated by Anderson-Krathwohl (2001), which have been adjusted according to the blueprint. The instrument includes answer choices, reasons for choosing answers, and the confidence in selecting the reasons, along with an answer key. Before and After of Product Three-Tier HOTS Instrument Assessment can be seen in the Table 1.

Table 1. Revision Product

Before Revision

1. Seorang peserta didik kelas XII IPA melakukan percoaan terhadap tiga buah sampel A, B, dan C menggunakan beberapa indicator asam dan basa di laboratorium sekolah. Data hasil percobaan yang diperoleh peserta didik tersebut dapat dilihat pada tabel

Trayek Ph Indikator	Perubahan Warna	Warna Larutan		
		Sampel A	Sampel B	Sampel C
Metil Merah (4,2-6,3)	Merah-Kuning	Kuning	Kuning	Merah
Bromtimol Biru (6,0-7,6)	Kuning-Biru	Biru	Hijau	Kuning
Phenolphthalein (8,3-10,0)	Tidak Berwarna-Merah	Tidak Berwarna	Tidak Berwarna	Tidak Berwarna

Kesimpulan yang paling tepat dapat diputuskan oleh peserta didik tentang larutan sampel percobaannya adalah ….

A. sampel A adalah larutan NaOH dan sampel B adalah H_2SO_4.
B. sampel A adalah adalah larutan NaOH dan sampel C adalah H_2SO_4

C. sampel B adalah adalah larutan NaOH dan sampel A adalah H_2SO_4
D. sampel C adalah adalah larutan NaOH dan sampel A adalah H_2SO_4
E. sampel B adalah adalah larutan NaOH dan sampel C adalah H_2SO_4

Apa alasan anda memilih jawaban tersebut?

A. pH sampel B berada di sekitar basa dan pH sampel A berada di sekitar asam
B. pH sampel A netral dan pH sampel C berada di sekitar asam
C. pH sampel A berada di sekitar basa dan pH sampel C berada di sekitar asam
D. pH sampel B netral dan pH sampel C berasa di sekitar basa
E. pH sampel C netral dan pH samperl B berada disekitar asam

Apakah anda yakin dengan pilihan jawaban dan alasan jawaban diatas?

A. Yakin
B. Tidak yakin

After Revision

Kesimpulan yang paling tepat dapat diputuskan oleh peserta didik tentang larutan sampel percobaannya adalah ….

A. sampel A adalah larutan basa dan sampel B adalah asam.
B. sampel A adalah adalah larutan basa dan sampel C adalah asam
C. sampel B adalah adalah larutan basa dan sampel A adalah asam
D. sampel C adalah adalah larutan basa dan sampel A adalah asam
E. sampel B adalah adalah larutan basa dan sampel C adalah asam

Apa alasan anda memilih jawaban tersebut?

A. Sampel B memiliki pH < 7 karena uji dengan phenolphthalein menghasilkan larutan tidak berwarna yang menunjukkan basa lemah
B. Sampel C memiliki pH > 7 karena uji dengan phenolphthalein menghasilkan larutan tidak berwarna yang menunjukkan basa lemah
C. Sampel A memiliki pH > 7 karena hasil uji dengan bromtimol biru menghasilkan larutan berwarna biru yang menunjukkan basa lemah
D. Sampel B memiliki pH < 7 karena hasil uji dengan bromtimol biru menghasilkan larutan berwarna hijau yang menunjukkan asam kuat
E. Sampel A memiliki pH < 7 karena hasil uji dengan metil merah menghasilkan laruan berwarna kuning yang menunjukkan asam lemah

Realization

At this phase, the HOTS three-tier assessment instrument that has been created and declared valid by both experts can be tried out on students. The validity results can be seen in the Table 2.

Table 2. Result of Validity Analysis

Aspects being assessed	Average	Category
Material Aspect	3,58	Very Valid
Construction Aspect	3,56	Very Valid
Language Aspect	3,5	Very Valid
Additional Rules	4	Very Valid

The next test is the inter-rater test, which examines the level of agreement between the raters given by the two validators, then it's analyzed using the percentages of agreements formula (Borich et al). The assessment given by both validators in Table showed that the number of agreements (A) = 18 and the number of disagreements (D) = 0, so the result of reliability coefficient (degree) of the instrument obtained is 1.00 or r = 100%, which means the HOTS three-tier assessment instrument on the topic of acids and bases falls into the 'very high' agreement level category, with the inter-rater agreement criteria ranging from 0.81 – 1.00. Therefore, it can be stated that the HOTS three-tier assessment instrument on acids and bases is reliable.

Test, Evaluation, and Revision

Difficulty Index (DI)

Difficulty testing is very important to know the level of difficulty of each question on the instrument. The difficulty index is used to categorize the questions into easy, medium, and hard. The results of the difficulty index analysis are presented in a table, and the comparison of the difficulty index between Tier 1 and 2 can be seen in the Table 3.

Table 3. Result of Difficulty Index

Criteria	Tier 1		Tier 2	
	No. Question	Total	No. Question	Total
Easy	2	1	1,2,3,4,15	5
Medium	1,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22, 23,24,25,26,27,28,29,30	29	5,6,7,8,9,11,12,13,14,16,17,19,20,21,22,23,24,25,27,28,29	21
Difficult	-	-	10,18,26,30	4

Table shows that the questions at the second tier have a higher DI compared to the first tier, which means the difficulty in the reasoning choices is greater than in the answer choices at the first tier. This is because explaining reasons at the second tier requires the ability to explain cause-and-effect relationships, whereas at the first tier it only involves stating the answer (Rohim et al., 2025). The table shows that the developed product has questions at a moderate level, with only a few at easy and difficult levels, which means that the product can measure cognitive levels well. According to Kwangmuang et al. (2021), if a question is neither too difficult nor too easy, it can be said that the question is good. Therefore, if all participants are unable to answer correctly because it is too difficult, or if all participants can answer correctly because it is too easy, the question can still be considered good.

Item Discrimination (ID)

The discriminating power of a question shows how well the question can tell the difference between high-performing and low-performing students who answer each question correctly. A high discriminating power value indicates that the item is better at identifying students with high and low abilities (Handayani et al., 2023).

Table 4. Result of Item Discrimination

Criteria	Tier 1		Tier 2	
	No. Question	Total	No. Question	Total
VG	3,4,6,7,8,9,10,11,13,14,15,16,17,19,20,21,22,24,27,29	20	1,2,3,4,5,6,7,8,9,10,11,13,14,15,16,19,20,21,22,24,27,28,29	23
RG	1,2,12,30	4	12,18,23	3
MI	5,26	2	17,25	2
PI	18,23,25,28	4	26,30	2

The table shows that the products have an average ID with a very good category. Questions that have an ID in the PI (poor item) category cannot be used in the research because they are rejected. Based on the table above, there are several questions in the PI category that cannot be used or are rejected.

Reliability

Reliability testing was carried out to measure the consistency of the HOTS three-tier assessment instrument, which was analyzed using the Cronbach's Alpha equation, the most commonly used method to estimate the consistency or reliability that is a function of the correlation of an item and the number of statements (Suseno et al., 2021). Reliability testing was also conducted at tiers 1 and 2. The test results can be seen in the Table 5.

Table 5. Result of Realibility

Aspect	Alpha Cornbach Coefficient	Category
Answer (Tier 1)	0.882	Very High
Reason (Tier 2)	0.925	Very High

Based on the table above, it was found that the product has reliability in the very high category. This means that the assessment instrument product has a very good level of measurement consistency.

Validity

The validity of test items refers to whether the information obtained from the test truly represents the students' understanding. This empirical validity test is carried out at tiers 1 and 2. The results of the empirical validity test analysis are presented in the table 6.

Table 6. Result of Validity

Criteria	Tier 1		Tier 2	
	No. Question	Total	No. Question	Total
Valid	1,2,3,4,6,7,8,9,10,11,12,13,14,15,14,17,18,19,20,21,22,23,24,27,29,30	26	1,2,3,4,5,6,7,8,9,10,12,14,15,16,17,19,20,21,22,23,24,27,28,29	24
Invalid	5,25,26,28	4	11,13,18,25,26,30	6

Table 6 shows that the average questions in the category of valid products, indicating that the product of this three-tier HOTS assessment instrument is able to measure HOTS accurately.

Distractor Functioning (DF)

The purpose of putting in a decoy is so that students taking the test with many learning outcomes are more interested in choosing that answer because they think it is correct (Sudijono, 2012).

In this study, the analysis of the distractor test at tiers 1 and 2 on 30 multiple-choice questions found 1 question with a good distractor at both tiers 1 and 2, and 4 questions with poor distractors at tier 1 and 7 questions with poor distractors at tier 2. The rest of the distractor

analysis results can be categorized as having bad distractors.

From the results of the quality test of the HOTS three-tier assessment instrument questions on acid-base material, it was found that 2 out of 30 questions were not suitable for use because in terms of validity, ID, DI, and DF they did not meet the criteria, specifically questions number 25 and 26. Meanwhile, the other 28 questions still meet the criteria in terms of validity, ID, DI, and DF, so they can still be used.

Table 7. Result of Distractor Functioning

Criteria	Tier 1		Tier 2	
	No. Question	Total	No. Question	Total
Very Good	-	-	-	-
Good	23	1	10	1
Medium	1,21,25,30	4	17,18,20,25,26,29,30	7
Bad	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20, 22,24,26,27,28,29	25	1,2,3,4,5,6,7,8,9,11,12,13,14,15,16,19,21,22,2 3,24,27,28	27
Very Bad	-	-	-	-

In addition to providing the HOTS three-tier assessment instruments, a teacher response questionnaire was also given to two chemistry teachers and a student response questionnaire was given to 31 students in a limited trial. The analysis of the teacher response questionnaire showed a percentage of 96.25% with the criteria of very practical, and the analysis of the student response questionnaire showed a percentage of 84.91% with the criteria of very practical.

The HOTS three-tier assessment instrument on the topic of acids and bases was trialed with the help of the Gimkit platform. Therefore, the Gimkit platform itself also needs to be evaluated for its practicality. The practicality analysis according to media experts shows a percentage of 72.5%, categorized as practical. This is because there are still many shortcomings in the Gimkit platform, such as links to questions that sometimes change and the limited number of questions in a single learning room.

The product developed is a HOTS three-tier assessment instrument in multiple-choice form that can identify students' higher-order thinking skills on acid and base materials based on students' answer patterns: understood, less understood, guessing, or misconception. The developed HOTS three-tier assessment instrument consists of a test grid for the HOTS three-tier assessment instrument, already equipped with answer keys and tested with the help of the Gimkit platform.

The developed HOTS three-tier assessment instrument functions as a tool that can identify students' higher-order thinking skills and misconceptions on acid and base materials by considering students' answers at each question level. Since this test can measure students' higher-order thinking skills and understanding, it can also be used as a tool to measure the success of a learning process, whether students really understand the concepts well, whether they have developed their higher-order thinking skills, or if they still lack understanding of the concepts.

Conclusion

The HOTS three-tier assessment instrument that was developed includes validity analysis, reliability analysis, and practical analysis. The content validity analysis is based on expert evaluation in the content validity test, with a value of 91.5%, which falls into the very valid category, with an inter-rater value of 100%. Empirical validity test results showed an average significance value of 0.05 at tier 1 and 0.05 at tier 2, which is considered valid. Consistency is based on reliability test results with a Cronbach's Alpha coefficient of 0.882 at tier 1 and 0.925 at tier 2, categorized as high reliability. The HOTS three-tier assessment instrument is considered practical based on a teacher response questionnaire analysis result of 96.25%, categorized as very practical, and a student response questionnaire result of 84.91%, also categorized as very practical.

Acknowledgments
We would like to express our sencere gratitude to all parties who have contributed to support the implementation of this research.

Author Contributions
All authors contributed to writing this article.

Funding
No external funding.

Conflicts of Interest
No conflict interest.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New

York: Longman Group Limited.

Andriyatno, I., Zulfiani, Z., & Mardiati, Y. (2023). Higher Order Thinking Skills: Student Profile Using Two-Tier Multiple-Choice Instrument. *International Journal of STEM Education for Sustainability*, *3*, 1. https://doi.org/10.52889/ijses.v3i1.79.

Ariani, W., Nahadi, & Munawaroh, H. S. H. (2025). Pengembangan Instrumen Tes Kimia Framework PISA untuk Mengukur Kemampuan Literasi Sains Siswa Madrasah Aliyah. *Jurnal Penelitian Pendidikan IPA*, *11*(11), 1291–1302. https://doi.org/10.29303/jppipa.v11i11.12808

Baihaki, B., Danaryanti, A., & Kamaliyah, K. (2021). Pengembangan LKPD Elektronik Berbasis HOTS Menggunakan Quizizz. *Journal of Mathematics Science and Computer Education*, *1*(1), 36. https://doi.org/10.20527/jmscedu.v1i1.3352

Dwijayanti, N. (2021). Pembelajaran Berbasis HOTS sebagai Bekal Generasi Abad 21 di Masa Pandemi. *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, *9*(1). https://doi.org/10.20961/jkc.v9i1.53837

Handayani, Y., Asia, E., & Hidayat, S. (2023). Peningkatan Kemampuan High Order Thinking Skills (HOTS) melalui Project-Based Learning (PjBL) dalam Implementasi Kurikulum Merdeka. *PTK: Jurnal Tindakan Kelas*, *4*(1), 48–60. https://doi.org/10.53624/ptk.v4i1.236

Jihannita, J., Prasetyo, Z. K., & Wilujeng, I. (2023). How to Prepare HOTS to Face the 21st Century? *Jurnal Penelitian Pendidikan IPA*, *9*(8), 486–492. https://doi.org/10.29303/jppipa.v9i8.2847

Kwangmuang, P., Jarutkamolpong, S., Sangboonraung, W., & Daungtod, S. (2021). The development of learning innovation to enhance higher order thinking skills for students in Thailand junior high schools. *Heliyon*, *7*(6), e07309. https://doi.org/10.1016/j.heliyon.2021.e07309

Muliyadi, L., Hamidi, H., Islami, A. V., Rizaldi, D. R., Hasanah, R., Yadaeni, A., Hariadi, M. H., Hanafi, M., & Wathoni, H. (2026). Analysis Validation of Physics Learning Devices Based on Problem Based Learning Assisted by PhET Simulation to Improve Students' Creativity and Critical Thinking. *Indonesian Journal of Science Education and Applied Research*, *1*(1), 27–31. Retrieved from https://journals.balaipublikasi.id/index.php/ijosear/article/view/743

OECD. (2023). *PISA 2022 Results: The State of Learning and Equity in Education*. Paris: OECD Publishing.

Rohim, A. A., Suharti, S., & Herunata. (2025). Enhancing Higher Order Thinking Skills (HOTS) through a Scientific Approach Oriented in Teaching Salt Hydrolysis. *Journal of Physics: Conference Series*, *3148*(1). https://doi.org/10.1088/1742-6596/3148/1/012027

Saepuzaman, D., Istiyono, E., Haryanto, & Retnawati, H. (2025). Characteristics of Two-Tier Multiple Choice (TTMC) High Order Thinking Skill (HOTS) Instruments and the Estimation of HOTS Abilities of Vocational Education Students Using the Item Response Theory Approach. *Journal of Engineering Education Transformations*, *38*(Special Issue), 13–22. https://doi.org/10.16920/jeet/2025/v38is3/25086

Sugiyati, & Syafryadin. (2025). The effectiveness of Gimkit in enhancing student's reading skill. *Indonesian Journal of Educational Development (IJED)*, *6*(1), 148–159. https://doi.org/10.59672/ijed.v6i1.4569

Sundari, S., Kahar, M. S., & Erwinda, E. G. (2021). Analisis Kemampuan Berpikir Tingkat Tinggi Menggunakan Instrumen Hots Berbasis Two Tier Diagnostic Test. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *10*(4), 2726–2735. https://doi.org/10.24127/ajpm.v10i4.4260

Surayanah, S., & Karma, L. (2023). Mengembangkan Higher Order Thinking Skills dan Prestasi Belajar melalui Elaborasi Nilai-Nilai Kearifan Lokal dalam Pembelajaran Sains. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, *6*(5), 3469–3481. https://doi.org/10.54371/jiip.v6i5.2054

Suseno, E., & Susongko, P. (2021). *Mengukur Validitas Tes*. Pemeral Edukreatif.

Susilawati, S., Doyan, A., Hardjono, A., & Muliyadi, L. (2021). Development of Physics Learning Media based on Guided Inquiry Model to Improve Students' Concepts Mastery and Creativity. *Journal of Science and Science Education*, *2*(2), 68–71. https://doi.org/10.29303/jossed.v2i2.711

Umroh, H., Rijal, S., & Yunus, F. M. (2024). Mereformasi Pendidikan : Mengkaji Rendahnya Kemampuan Berpikir Kritis Siswa melalui Pendekatan Pendidikan Kritis Ivan Illich. *ASPIRASI : Publikasi Hasil Pengabdian Dan Kegiatan Masyarakat*, *3*(1), 18–32. https://doi.org/10.61132/aspirasi.v3i1.1306

Yuanta, F., Suroso, N. L. D. H., Wulandari, D. A., Anggraini, D. N., Hermawan, N. S., & Yosepin Betssy Tomasoa. (2025). Pengembangan Media Pembelajaran Berbasis Gimkit Materi Perkalian Kelas V Sekolah Dasar. *Indo-MathEdu Intellectuals Journal*, *6*(1), 920–926. https://doi.org/10.54373/imeij.v6i1.2575

Zebua, N. (2024). Pengembangan Media Pembelajaran Biologi Berbasis HOTS melalui Platform Blog dengan Model PLOMP untuk Meningkatkan Motivasi dan Hasil Belajar. *GEN BIONIX: Jurnal Pendidikan Biologi*, *2*(2), 2988–716. https://doi.org/10.56207/gb.v2i2.771