SAINTIS
publishing

Research Article

Development of HOTS Assessment Tools to Measure Achievement in Acid-Base Learning Outcomes

Rauzatul Jannah, Isna Rezkia Lukman*, Fiqih Choiruddin, Naila Nurul Izzah and Ulfa Zahra, Halimatus Sakdiah

Faculty of Teachers Training and Education, Universitas Malikussaleh, Aceh Utara, Indonesia, 24351
*Corresponding Author: rezkia.lukman@unimal.ac.id | Phone: +6281220237241

ABSTRACT

This study aims to develop a valid and reliable assessment instrument to measure students' higher-order thinking skills in the context of acid-base chemistry. The research employed a Research and Development (R&D) approach using the 4-D model, which includes the stages of Define, Design, Develop, and Disseminate. The initial product was validated through expert judgment, evaluated for feasibility by chemistry teachers, and subjected to field testing. The study involved 200 twelfth-grade science students from SMAN Bireuen. Data collection techniques included interviews, questionnaires, and tests. Data collection instruments included interview guidelines, item validation sheets, feasibility evaluation forms, and multiple-choice questions with open-ended justifications. The research findings revealed that the material validation assessment obtained an average score of 47.5 ($\bar{X} > 46.194$), categorized as "very good." Similarly, the feasibility assessment achieved an average score of 72.5 ($\bar{X} > 67.188$), also falling within the "very good" category. A total of 30 test items from the trial were analyzed in terms of validity, reliability, difficulty level, discrimination index, and distractor effectiveness. Therefore, it can be concluded that the assessment instrument is appropriate for measuring higher-order thinking skills in the topic of acid-base chemistry.

Keywords: acid-base; assessment; chemistry; instruments; HOTS

1. INTRODUCTION

Learning is defined as a change in behavior among both learners and educators, resulting from the process of delivering and receiving knowledge through various instructional methods. The effectiveness of the learning process can be evaluated based on the achievement of predetermined learning objectives (Fatmawati, 2021; Sirait, 2021; Sutikno, 2023). To assess the achievement of these objectives, an evaluation of students' learning outcomes is necessary. Assessment within the learning system serves as a means to evaluate the effectiveness of both the learning process and its outcomes for students (Mendiknas, 2007; Pemerintah Indonesia, 2005). Assessment plays a pivotal role in the educational process, serving as a key benchmark for evaluating learning outcomes. Broadly speaking, assessment functions as a tool for educational evaluation, enabling stakeholders to understand the current status and effectiveness of the learning process. The degree to which educational objectives and targeted competencies are achieved is measured through systematic evaluation activities. These activities involve the structured collection, analysis, and interpretation of data to determine the extent of learning achievement and to inform future instructional decisions (Agustianti et al., 2022; Calista & Yefterson, 2022; Pitaloka et al., 2021) and has the ability to stimulate students to optimize their potential in the learning process. Accurate assessment plays a crucial role in enhancing students' Higher-Order Thinking Skills (HOTS), as it provides meaningful feedback and encourages deeper cognitive engagement in the learning process (Armanto et al., 2021; Nisa et al., 2022; Rorimpandey et al., 2023)

Higher-order thinking skills (HOTS) occupy the upper levels of Bloom's cognitive taxonomy and are designed to stimulate learners to apply knowledge and skills in novel contexts (Halimah, 2021; Istiyono et al., 2014; Rochman & Hartoyo, 2018). HOTS encompass the abilities to analyze, evaluate, and create, as outlined by Anderson (1983). Higher Order Thinking Skills (HOTS) assessments are designed to stimulate learners to apply their existing knowledge and skills to novel situations (Sari et al., 2023; Setyowati et al., 2023). Rather than relying solely on rote memorization, students are encouraged to connect previously learned concepts with real-world problems encountered in everyday life. This approach fosters meaningful and relevant learning experiences while simultaneously cultivating advanced cognitive abilities that are essential in today's globalized and rapidly evolving world.

Based on interviews conducted with chemistry teachers of grade XI science students at SMAN 1 Bireuen (a top-performing school), it was found that assessment practices have not yet been fully aligned with appropriate pedagogical standards. The teachers continue to rely predominantly on conventional multiple-choice tests, and cognitive process assessments are often based on the teacher's subjective judgment. Specifically, students are awarded high scores if they appear to engage actively in classroom activities, without a comprehensive evaluation of the cognitive dimensions that should be systematically measured.

To address the identified issues, a two-tier multiple-choice diagnostic test was developed. This type of test consists of five answer options accompanied by a rationale for each selected response. A two-tier multiple-choice item includes two levels: the first level presents a set of options comprising the correct answer and distractors, while the second level requires respondents to select a justification that reflects their conceptual understanding or reasoning behind the initial choice (Jamhari, 2021; Rintayati et al., 2021). This format has been recognized as a more effective assessment tool compared to conventional multiple-choice questions. Previous studies Samaduri (2022) have reported positive outcomes using this type of instrument to identify students' alternative conceptions. Previous research conducted by Desiriah & Setyarsih (2021) concluded that the development of Higher-Order Thinking Skills (HOTS) assessment instruments in the form of reasoned multiple-choice items effectively measures students' abilities in analysis, evaluation, and creation. The instrument was found to meet the necessary criteria for evaluating HOTS competencies. The primary objective of this study is to evaluate the feasibility of a Higher-Order Thinking Skills (HOTS) assessment instrument designed to measure learning outcomes of eleventh-grade science students (Class XI IPA) on acid–base chemistry topics. Accordingly, the study seeks to answer the following research question: How feasible is the HOTS-based assessment instrument in effectively measuring students' learning achievement.

2. RESEARCH METHOD

This study employed a qualitative approach with the primary aim of outlining the procedural steps necessary for developing an educational product. The research was conducted at SMA Negeri 1 Bireuen, located in Bireuen Regency, Aceh, involving a total of 200 student participants. The study adopted a research and development (R&D) methodology. The development of the HOTS (Higher-Order Thinking Skills) assessment instrument was carried out using the 4D model of instructional development. The product development process consisted of four key phases, namely:

1. Define

    This stage involves a comprehensive initial-to-final analysis, including learner analysis, formulation of learning objectives, conceptual analysis, and task analysis.

2. Design

    At this stage, the development process focuses on determining the type of instrument to be used, constructing a blueprint (test specification table), and designing the instrument accordingly.

3. Develop

    This phase encompasses expert validation and field testing. The process begins with the initial product being validated by subject matter experts, followed by revisions based on the feedback received. The revised version is then reviewed by external reviewers to finalize the product. After this, pilot testing is conducted to evaluate its practicality and effectiveness.

4. Disseminate

    In this final phase, the completed product is disseminated for broader use by relevant stakeholders. The dissemination process aims to facilitate large-scale implementation and adoption.

The instruments employed in this study included a small-scale exploratory interview to gather initial data on educators' needs. A content expert validation sheet was designed to enhance the initial development of the instrument. The validation technique functioned as a tool for evaluation, feedback, and recommendations from experts, which were then used to revise the instrument accordingly. A product feasibility sheet was assessed by reviewers to evaluate the overall quality of the product, encompassing aspects such as content, construction, language clarity, validity, and practicality. Additionally, a multiple-choice reasoning test was specifically developed to assess students' learning outcomes in chemistry. A field trial was conducted to determine the quality of the test items, evaluating their validity, reliability, difficulty level, discrimination index, and distractor effectiveness.

3. RESULTS AND DISCUSSION

3.1 Results

The research data were obtained using the 4-D model, comprising both questionnaire responses and test results. The results of the material expert validation and product feasibility assessments are presented in **Table 1** and **Table 2**.

Table 1. Assessment of Material Aspects

Aspects	Average Score	Category
Learning	14,5	Excellent
Content	17,0	Excellent
Language	8,5	Excellent
Presentation	7,5	Good
Total Score	47,5	Excellent

Based on **Table 1**, the average score from the assessment by subject matter experts across all aspects was 47.5, which falls within the "very good" category. This indicates that the quality of the HOTS assessment instrument for acid-base topics is rated as very good. Based on **Table 1**, the average score from the assessment by subject matter experts across all aspects was 47.5, which falls within the "very good" category. This indicates that the quality of the HOTS assessment instrument for acid-base topics is rated as very good.

Table 2. Assessment of Feasibility Aspects

Aspects	Average Score	Category
Substance	12,6	Excellent
Construction	13,6	Excellent
Language	18,2	Excellent
Validity	8,9	Excellent
Practicality	19,2	Excellent
Total Score	47,5	Excellent

Based on **Table 2**, the average score from the reviewers' assessments across all aspects was 72.5, which falls under the "very good" category, indicating that the HOTS assessment instrument is deemed suitable for use. Subsequently, the HOTS assessment instrument will proceed to the trial phase to evaluate the quality of the test items, including their validity, reliability, difficulty level, discriminating power, and distractor analysis.

1. Validity

   An item is considered valid if the correlation coefficient (r_{xy}) is greater than or equal to the critical value of the r-table ($r_{xy} \geq r_{table}$); otherwise, it is deemed invalid ($r_{xy} \leq r_{table}$) Based on this criterion, 29 items were found to be valid, while 1 item was categorized as invalid.

2. Reliability

   The reliability analysis was conducted using Cronbach's Alpha (α) formula. The obtained reliability value was 0.84, which falls within the range of very high reliability.

3. Difficulty Level

   Based on the analysis, it was found that one item was categorized as easy, while the remaining 29 items were classified as having a moderate level of difficulty.

4. Discriminatory Power

   Based on the analysis, it was found that 11 items fell into the "good" category, 12 items were categorized as "fair," and 7 items were classified as "poor."

5. Distractor Analysis

   Based on the analysis, it was found that 23 test items were of good quality and deemed suitable for use, while 7 items required revision.

3.2 DISCUSSION

Higher-order thinking skills (HOTS) refer to students' ability to apply and evaluate knowledge in complex contexts (Hamzah et al., 2022; Kim How et al., 2022) . As such, the development of HOTS is strongly emphasized to foster students' cognitive growth. These skills can be cultivated through instructional programs that focus on cognitive reasoning processes, particularly by creating learning conditions that promote practice in solving problems aligned with HOTS (Indriyana & Kuswandono, 2019). One effective approach to assessing HOTS is through the use of reasoned multiple-choice tests. This form of assessment consists of two-tiered questions: the first tier includes a correct answer along with distractors, while the

second tier presents alternative conceptions or justifications that reflect the students' reasoning behind their initial choices (Lengkong et al., 2021; Sesli & Kara, 2012).

The evaluation results from both content experts and reviewers indicate that the instruments are not only appropriate for use but also fall within the "excellent" category. This is evidenced by the overall mean score from the content expert assessments, which reached 47.5 exceeding the threshold of $X > 46.194$, thus qualifying as excellent. Similarly, the reviewers' overall mean score was 72.5, placing it within the $X > 67.188$ range, also categorized as excellent. Furthermore, the results of the instrument trials analyzing aspects such as validity, reliability, item difficulty, discrimination power, and distractor effectiveness demonstrated that 23 items were considered good and suitable for use, while 7 items required revision. The trial test consisted of 30 multiple-choice items, each accompanied by a rationale. Every item was constructed based on specific learning indicators, Higher-Order Thinking Skills (HOTS) indicators, knowledge dimensions, and item criteria. For example, in an item related to soil pH, students were expected to differentiate between acidic, alkaline, and neutral soil conditions. The HOTS level was categorized as C4, emphasizing analytical skills. The associated knowledge dimension was procedural, focusing on understanding how to perform tasks based on the concept of acid-base neutralization in soil. Procedural knowledge includes familiarity with techniques and methods, as supported by previous studies (Alexander et al., 1991; De Jong & and Ferguson-Hessler, 1996; Dochy & Alexander, 1995) .The item criteria required that each question be valid and demonstrate good discriminative power. A higher number of correct responses indicated greater student mastery of specific items, which was attributed to appropriate item difficulty levels.

4. CONCLUSION

Based on the research objectives and findings, it can be concluded that the assessment instrument developed—consisting of multiple-choice tests with open-ended justifications—is of high quality and deemed highly appropriate for use, as validated by subject matter experts and independent reviewers. The average evaluation scores indicate that the developed HOTS (Higher Order Thinking Skills) assessment instrument falls within the "excellent" category.

RECOMMENDATIONS

The results of this research and study are expected to provide information, insight, and knowledge that will benefit readers and institutions.

ACKNOWLEDGEMENTS

AUTHOR'S CONTRIBUTIONS

All authors discussed the results and contributed to from the start to final manuscript.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

Agustianti, R., Abyadati, S., Nussifera, L., Irvani, A. I., Handayani, D. Y., Hamdani, D., & Amarulloh, R. R. (2022). *Asesmen Dan Evaluasi Pembelajaran*. Tohar Media.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to Terms: How Researchers in Learning and Literacy Talk about Knowledge. *Review of Educational Research*, *61*(3), 315–343. https://doi.org/10.2307/1170635

Anderson, L. W. (1983). *The Architecture of Cognition*. Harvard University Press. .

Armanto, D., Suprayetno, E., Sinaga, K., & Sugiarto, A. (2021). Pelatihan Penyusunan Instrumen Penilaian Berbasis HOTS bagi Guru SD IT Taman Cahaya Siantar. *RESWARA: Jurnal Pengabdian Kepada Masyarakat*, *2*(2), 379–386. https://doi.org/https://doi.org/10.46576/rjpkm.v2i2.1225

Calista, R. D. A., & Yefterson, R. B. (2022). Pelaksanaan Pembelajaran Sejarah Menggunakan Model Pembelajaran Kooperatif Tipe Jigsaw di SMA Pembangunan Bukittinggi. *Jurnal Kronologi*, *4*(4), 152–162. https://doi.org/https://doi.org/10.24036/JK.V4I4.561

De Jong, T., & and Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2

Desiriah, E., & Setyarsih, W. (2021). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Fisika diSMA. *ORBITA: Jurnal Pendidikan Dan Ilmu Fisika*, *7*(1), 79–89. https://doi.org/https://doi.org/10.31764/orbita.v7i1.4436

Dochy, F., & Alexander, P. (1995). Mapping Prior Knowledge: A Framework for Discussion Among Researchers. *European Journal of Psychology of Education*, *10*, 225–242. https://doi.org/10.1007/BF03172918

Fatmawati, I. (2021). Peran Guru dalam Pengembangan Kurikulum dan Pembelajaran. *Revorma: Jurnal Pendidikan Dan Pemikiran*, *1*(1), 20–37. https://doi.org/https://doi.org/10.62825/revorma.v1i1.4

Halimah, S. (2021). Implementasi Pendekatan HOTs (Higher Order Thinking Skills) dalam Pembelajaran PAI. *Evaluasi: Jurnal Manajemen Pendidikan Islam*, *5*(2), 342–362. https://doi.org/http://doi.org/10.32478/evaluasi.v5i2.762

Hamzah, H., Hamzah, M. I., & Zulkifli, H. (2022). Systematic Literature Review on the Elements of Metacognition-Based Higher Order Thinking Skills (HOTS) Teaching and Learning modules. *Sustainability*, *14*(2), 813. https://doi.org/https://doi.org/10.3390/su14020813

Indriyana, B. S., & Kuswandono, P. (2019). Developing Students' Higher Order Thinking Skills (HOTS) in Reading: English Teachers' Strategies in Selected Junior High Schools. *Journal of English Teaching*, *5*(3), 204–216. https://doi.org/https://doi.org/10.33541/jet.v5i3.1313

Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (pysthots) Peserta Didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Jamhari, M. (2021). Developing The Two-Tier Multiple Choice Tests In Enhancing Students Higher-Order Thinking Skills On Human Digestive System. *Eduproxima: Jurnal Ilmiah Pendidikan IPA*, *3*(1), 50–64. https://doi.org/https://doi.org/10.29100/eduproxima.v3i1.1853

Kim How, R., Zulnaidi, H., & Rahim, S. S. B. (2022). Development of Higher-Order Thinking Skills Test Instrument on Quadratic Equations (HOTS-QE) for Secondary School Students. *Pegem Journal of Education and Instruction*, *13*(1), 379–394. https://doi.org/10.47750/pegegog.13.01.41

Lengkong, M., Istiyono, E., Rampean, B., Tumanggor, A., & Nirmala, M. (2021, January 1). Development of Two-Tier Test Instruments to Detect Student's Physics Misconception. *Conference: 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences*. https://doi.org/10.2991/assehr.k.210305.082

Mendiknas. (2007). *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 41 Tahun 2007 Tentang Standar Proses untuk Satuan Pendidikan Dasar dan Menengah*.

Nisa, H., Junus, M., & Komariyah, L. (2022). Penerapan Model Problem Based Learning Berbantuan Simulasi PhET Berbasis Instrumen HOTS Terhadap Hasil Belajar Siswa. *Jurnal Ilmiah Pendidikan Fisika*, *6*(3), 560–567. https://doi.org/https://doi.org/10.31949/jee.v6i2.5376

Pemerintah Indonesia. (2005). *Peraturan Pemerintah RI Nomor 19 Tahun 2005 Tentang Standar Nasional Pendidikan*. https://peraturan.bpk.go.id/Details/49369/pp-no-19-tahun-2005

Pitaloka, D. L., Dimyati, D., & Purwanta, E. (2021). Peran Guru dalam Menanamkan Nilai Toleransi pada Anak Usia Dini di Indonesia. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, *5*(2), 1696–1705. https://doi.org/10.31004/obsesi.v5i2.972

Rintayati, P., Lukitasari, H., & Syawaludin, A. (2021). Development of Two-Tier Multiple Choice Test to Assess Indonesian Elementary Students' Higher-Order Thinking Skills. *International Journal of Instruction*, *14*(1), 555–566. https://doi.org/https://doi.org/10.29333/iji.2021.14133a

Rochman, S., & Hartoyo, Z. (2018). Analisis High Order Thinking Skills (HOTS) Taksonomi Menganalisis Permasalahan

Fisika. *Science and Physics Education Journal (SPEJ)*, *1*(2), 78–88.

Rorimpandey, W., Lumintang, P., & Tuerah, P. (2023). Pengaruh Model PBL Dan Evaluasi Berbasis Hots Terhadap Hasil Belajar Bilangan Bulat Kelas VI SD Negeri Desa Dodap. *Jurnal Elementaria Edukasia*, *6*(2), 858–873. https://doi.org/https://doi.org/10.31949/jee.v6i2.5376

Samaduri, A. (2022). Analisis Pemahaman Konsep Siswa yang Diukur Menggunakan Tes Pilihan Ganda Beralasan pada Mata Pelajaran Beralasan Pada Mata Pelajaran Biologi. *Jurnal Pendidikan Glasser*, *6*, 109–119. https://doi.org/10.32529/glasser.v6i1.1466

Sari, N., Ratu, T., Rittianti, R., & Erfan, M. (2023). Pengaruh Model Pembelajaran Poe (Predict-Observe-Explain) Terhadap Kemampuan Berpikir Tingkat Tinggi (Hots) Muatan Ipa Siswa Kelas V Disekolah Dasar. *BIOCHEPHY: Journal of Science Education*, *3*(1), 83–88. https://doi.org/https://doi.org/10.52562/biochephy.v3i1.776

Sesli, E., & Kara, Y. (2012). Development and Application of a Two-Tier Multiple-Choice Diagnostic Test for High School Students' Understanding of Cell Division and Reproduction. *Journal of Biological Education*, *46*(4), 214–225. https://doi.org/https://doi.org/10.1080/00219266.2012.688849

Setyowati, Y., Priambudi, S., & Dewanto, D. (2023). Supervision of the Development of Higher-Order Thinking Skills (HOTS)-Based Assessment of Learning in Wijaya Putra School. *Mattawang: Jurnal Pengabdian Masyarakat*, *4*(4), 367–376. https://doi.org/https://doi.org/10.35877/454RI.mattawang2240

Sirait, J. E. (2021). Analisis Pengaruh Kompetensi Guru Terhadap Keberhasilan Pembelajaran di Sekolah Dasar Bethel Tanjung Priok Jakarta Utara. *Diegesis: Jurnal Teologi*, *6*(1), 49–69. https://doi.org/https://doi.org/10.46933/DGS.vol6i149-69

Sutikno, Y. (2023). Peran Guru dalam Evaluasi Pembelajaran di Kelas. *Jurnal Maitreyawira*, *4*(1), 36–44. https://doi.org/https://doi.org/10.69607/jm.v4i1.73