

Research Article

Cite this article: Ayubi, M., Ikhsan, J., Arthamena, V. D., Chaniago, M. A. & Pradesta, E. (2024). Validity and Reliability of HOTS (High Order Thinking Skills) Instrument for Voltaic Cell Subject Using the Classical Method. *Üniversitepark Bülten*, 13(1): 53-64. <https://doi.org/10.22521/unibulletin.2024.131.4>

Received March 27, 2024

Accepted May 29, 2024

Keywords:

HOTS, Borg and Gall, validity, RnD

Author for correspondence:

Mizzan Ayubi

✉ Mizzan0019fmipa.2022@student.uny.ac.id

✉ Yogyakarta State University, Indonesia.



OPEN ACCESS

© The Author(s), 2024. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Validity and Reliability of HOTS (High Order Thinking Skills) Instrument for Voltaic Cell Subject Using the Classical Method

Mizzan Ayubi^{ID}, Jaslin Ikhsan^{ID}, Vegha Dwi Arthamena^{ID}, Muhammad Akbar Chaniago^{ID}, Endar Pradesta^{ID}

Abstract

Background/purpose. Higher order thinking skills (HOTS) are conceptually qualified critical-thinking skills, based on Bloom's revised taxonomy. Frequently used assessments in schools may not always optimally help learners deal with problems conceptually. The habit of teachers preparing test questions without being based on a test grid, but relying only upon questions sourced from textbooks, can result in students not being adequately trained in higher-level thinking skills. Especially in chemistry subjects, it appears that the ability of students to analyze and solve problems is very low.

Materials/methods. This research and development (R&D) study employed the Borg and Gall development model on the research subject of HOTS instruments. The study population was from the Muara Beliti sub-district of Musi Rawas, South Sumatra, Indonesia, and was conducted with 102 students from seventh-grade IPA (International Program on Science Education) classes 1, 2, and 3. Data analysis included tests for validity, reliability, proportion of questions, and discrimination of questions were analyzed according to the classical method.

Results. The research found 10 questions to be valid and reliable. The developed HOTS instrument may therefore be effectively employed to measure the high-level thinking ability of high school students in learning chemistry, especially on Voltaic cell material. Tests that utilize HOTS instruments have the potential to encourage students to develop higher order thinking skills in the subject area.

Conclusion. This study concluded that: 1) The developed HOTS assessment instrument consisted of 10 essay questions that were found to be valid and appropriate in terms of construct, language, and material aspects; 2) Reliability of the HOTS assessment instrument was categorized as high, with a coefficient of .7420.

1. Introduction

In 2018, PISA (Programme for International Student Assessment) released its latest report on student literacy levels in various countries, including Indonesia. This data provided an overview of the extent of Indonesian students' literacy skills compared to other countries that were part of the PISA assessment at that time (Ayubi et al., 2023). Based on this PISA data, Indonesia was ranked 74th out of 79 member countries in terms of student literacy. The development of critical thinking, creativity, and problem-solving skills have been emphasized for learning in the 21st century (Chalkiadaki, 2018; Conklin, 2011; Quieng et al., 2015; Udompong & Wongwanich, 2014). The importance of developing critical thinking, creativity, and problem-solving skills in learning demands an appropriate assessment concept that can reflect the desired learning objectives. In this context, Indonesia's Curriculum 2013 uses an assessment standard based on an international standardized assessment model (Chinedu et al., 2015; Sepriyanti et al., 2022). Higher order thinking skills, or HOTS, serve to train learners' abilities at a high level in the cognitive domain, including levels C4 (analyze), C5 (evaluate), and C6 (create) in Bloom's revised taxonomy. With a focus on these skills, learners are encouraged not only to understand information, but also to analyze, evaluate, and create new solutions or concepts (Abidin et al., 2019; Kusuma et al., 2017; Wardany et al., 2017).

Education systems play a crucial role in supporting the acquisition of HOTS by learners since these skills can be enhanced through learner-centered teaching (Guskey, 2007; Halpern, 1998; Hmeli & Ferrari, 1997). However, not all learners receive adequate education to develop HOTS skills, and the quality of teachers is not always evenly distributed across schools (Clotfelter et al., 2005). Schools in some areas tend to have less experienced teachers who are new to the profession, and who often face difficulties in terms of basic teaching skills and may lack the necessary skills to develop a learner-centered teaching environment.

According to Demirhan and Koklukaya (2014), teachers have an important role in developing and improving students' higher order thinking skills. To equip learners with these abilities is one of the important goals of education and is a critical issue worthy of discussion and development (D'Alessio et al., 2019; Gojkov et al., 2015). Many universities have been criticized for putting insufficient effort into helping learners to acquire problem-solving and higher-order skills (McCormick, 2015; McLaughlin et al., 2014). Studies related to the development of higher-order thinking skills in learners, such as critical and analytical skills, are therefore needed (McDonald, 2012; Qing et al., 2010).

HOTS (higher order thinking skills) are conceptually qualified critical-thinking skills, based on Bloom's revised taxonomy (Madhuri et al., 2012). Learners in the 21st century not only require the ability to remember and recall information, but must also employ higher thinking skills in order to address today's increasingly complex problems. In addition, they need to have the ability to think critically and creatively (Madhuri et al., 2012; Mahoney & Harris-Reeves, 2019). Learning in schools can not only teach students to remember concepts and knowledge, but can also encourage the ability to analyze, evaluate, and create solutions to problems as they are encountered (Barbara & Wendy, 2010; Heong et al., 2011). Today's learners are expected to effectively train themselves by answering HOTS questions that are interesting to solve as a means to improving their higher order thinking skills.

In some Indonesian schools, the syllabus of Curriculum 2013 serves as a guideline for student learning activities to develop their critical-thinking skills (Fischer et al., 2011). However, some high school teachers in Indonesia have assessed that a mismatch exists between the requirements of Curriculum 2013 and the reality practiced in schools. In this context, as many as 86% of teachers in Indonesia still assess students at the remembering level (C1) in the cognitive domain (Abosalem, 2016). Research (Atitullah et al., 2019) has shown that there are 158 questions related to reading comprehension in the textbooks for grade 5 high school students in Indonesia. However, only 24 of

these questions are aimed at higher-order thinking. Research conducted by Febrina et al. (2019) revealed that starting from grade 6 textbooks are used in one semester, and that 66.8% of comprehension questions have a focus on HOTS, while 33.4% of questions are at the LOTS level. In a study conducted by Febriyani et al. (2020), it was stated that most of the questions in grade 7 textbooks were at the HOTS level.

Frequently used assessments in schools may not always optimally help learners deal with problems conceptually (Afflerbach et al., 2015; Boone et al., 2014; Haley et al., 2011). The habit of teachers creating test questions without being based on a test grid, and only relying on questions sourced from textbooks, can cause students not to develop higher-level thinking skills. Especially in chemistry subjects, it appears that the ability of students to analyze and solve problems is very low (Avargil et al., 2012; Drake & Nelson, 2008).

Based on the results of interviews conducted with high school chemistry teachers, it was found that the initial ability of students in the seventh-grade IPA (International Program on Science Education) class two was quite good, although there were still some scores below the average. The teacher worked to develop the students' higher order thinking skills (HOTS) in chemistry learning and tested the questions both as part of the mid-semester assessment (PTS) and the end-of-semester assessment (PAS). An interview with the seventh-grade IPA (International Program on Science Education) chemistry teacher for class one revealed that HOTS questions had been used during the learning process, although not on a regular basis. The teacher reportedly remained committed to preparing HOTS questions to support students' chemistry learning. In support of the current research, the participant chemistry teachers agreed to conduct item (question) trials. They expressed their belief that the right questions could help improve learners' abilities to answer questions correctly and were a means to encouraging the development of students' analytical skills.

Based on several previous studies on HOTS, the researchers in the current study identified a gap in the literature and aimed to develop a HOTS instrument that used essay-based questions. As such, the current study tests the validity and reliability of a HOTS instrument on the Volta cell chemistry topic.

2. Methodology

2.1. Type of Research

This paper adapted the research and development (R&D) study type using the Borg and Gall development model (Gall et al., 2007).

2.2. Research Design

The development model was modified and adapted in several stages in accordance with the objectives and interests of the current study (Kinay & Bagceci, 2016; Weiss, 2011). After applying the revisions, the product was revalidated by the chemistry teacher, and the revised product tested on three classes at SMAN 2 Muara Beliti school located in Musi Rawas, South Sumatra, Indonesia.

2.3. Research Subject

The study researched higher order thinking skill instruments in chemistry learning; specifically, Voltaic cell material, with a focused population in the Muara Beliti sub-district of Musi Rawas, South Sumatra, Indonesia. The sample of the study were three seventh-grade IPA (International Program on Science Education) classes, with a combined total of 102 students. The research was conducted during February and March of 2022.

2.4. Data Collection Techniques and Instruments

The data collection techniques applied in this study involved the use of questionnaires and tests (Aman, 2019). Questionnaires were used to assess the expert variables, specifically higher order thinking skills (HOTS). The data collection instrument involved an essay or polytomous test (Arifin, 2016; Boopathiraj & Chellamani, 2013) consisting of 10 questions. Validation of the questionnaire was conducted by a test instrument expert in order to assess the validity of the developed HOTS instrument (Brown et al., 2018; Pommerich, 2004; Szilagyi et al., 2013).

2.5. Data Analysis Techniques

Data analysis techniques were applied so as to obtain validity and reliability values (Cooper, 2023). The following subsections outline the data analysis processes employed.

2.5.1 Qualitative Data Analysis of Validation Sheet Results

Qualitative analysis of the validation sheets used was conducted by experts to assess the developed HOTS instrument based on material, construction, and language aspects (Cavallo et al., 2010; Polit & Beck, 2006; Setiawan et al., 2020). The assessment was conducted using a Likert-type scale, with items each given a certain value. The analysis of the HOTS instrument was calculated according to V'aiken's coefficient.

2.5.2 Quantitative Data Analysis of HOTS Test Questions

Quantitative analysis was conducted using Microsoft Excel program by considering the characteristics of the HOTS items.

a. Validity Test

A validity test was conducted using the correlation formula, with the calculation results compared to the r -table value, with a significance level of 5%. Items were considered valid where the calculated r -count value was found to be greater than or equal to that of the r -table value. Conversely, if the calculated r -count value was smaller than the r -table value, then the item was considered to be invalid (Azwar, 2016). The validity of the description items were calculated using Equation 1.

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (1)$$

b. Reliability Test

The reliability test was conducted according to the KR-21 or Cronbach's alpha formula, since the questions in this study were in essay form. The reliability criteria was set with a coefficient interval value of $\geq .70$ (Mardapi, 2008; Retnawati, 2016). Table 1 presents the reliability test criteria, with the formula used shown as Equation 2.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_t^2}{\sigma_t^2} \right) \quad (2)$$

Table 1. Test Reliability Coefficient Criteria (Miterianifa, 2016)

No	Range	Criteria
1	$\alpha \leq .20$	Very low
2	$.20 < \alpha \leq .40$	Low
3	$.40 < \alpha \leq .60$	Simply
4	$.60 < \alpha \leq .80$	High
5	$.80 < \alpha \leq 1.00$	Very high

c. Proportion Test of Questions

Analysis of the difficulty level of the test questions was needed in order to categorize them as difficult, moderate, or easy questions. The level of difficulty was calculated according to Equation 3.

$$P = \frac{\sum A + \sum B - (2N \text{ skor}_{\min})}{2N (\text{skor}_{\max} - \text{skor}_{\min})} \quad (3)$$

Regarding how to interpret the item difficulty index number, the book entitled “Measurement and Evaluation in Psychology and Education” by Robert L. Thorndike and Elizabeth Hagen mentioned criteria that can be used to assess level of difficulty, which was interpreted as shown in Table 2 (as cited in Sudijono, 2013).

Table 2. Range of Problem Proportions (Sudijono, 2012)

Score range	Difficulty level
$.00 \leq x < .30$	Difficult
$.31 \leq x < .70$	Moderate
$.71 \leq x \leq 1.030$	Easy

d. Question Item Discrimination

Item discrimination concerns the ability of the test to distinguish between high-ability students and low-ability students. The index of the differentiating power of the items uses standard deviation and product moment correlation, and the formula used is the *t*-test, as shown in Equation 4. Items are said to have differentiating power if $t_{count} > t_{table}$, with a significance level of 5%. Conversely, items are said not to have differentiating power if $t_{count} < t_{table}$, with a significance level of 5%.

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n_1(n_2 - 1)}}} \quad (4)$$

3. Results and Discussion

3.1. Expert Validation

Validation was conducted by six experts who assessed the construct, language, and material aspects of the test items. The process involved validation sheets used by the construct and content experts, which were then analyzed according to the V'Aiken formula in order to calculate the content validity coefficient. Results of the validation process conducted by the construct experts is shown in Table 3, whilst the content experts' validation results are shown in Table 4 for language and in Table 5 for material aspects.

Table 3. Validator's Assessment of the Construct Aspect

Indicator statement	V'Aiken coefficient	Criteria
The subject matter should be formulated clearly and unequivocally.	.71	Instrument can be used
The main formulation of the question must be a statement related to the material being measured.	.82	Instrument can be used

Indicator statement	V'Aiken coefficient	Criteria
The subject matter does not provide clues to the answer.	.78	Instrument can be used
Stimuli in the form of graphs, figures, tables, diagrams, and the like contained in the question must be clear, explicit and functional.	.80	Instrument can be used
The sentence formulation of the question must use an interrogative word or command that demands an elaborated answer.	.79	Instrument can be used

Table 4. Validator Assessment of Language Aspects

Indicator statement	V'Aiken coefficient	Criteria
The subject matter should be formulated clearly and unequivocally.	.74	Instrument can be used
The main formulation of the question must be a statement related to the material being measured.	.72	Instrument can be used
The subject matter does not give clues to the answer.	.90	Instrument can be used

Table 5. Validator Assessment of Material Aspects

Indicator Statement	V'Aiken coefficient	Criteria
The subject matter must be in accordance with the indicators.	.74	Instrument can be used
The subject matter must be logical in terms of material.	.77	Instrument can be used
The boundaries of questions and answers must be clear.	.73	Instrument can be used
Images, sentences, or slogans must not contain elements of any commercial product or agency promotion or advertisement.	.93	Instrument can be used
No names of living characters can be used in questions since it could be interpreted as the character.	.86	Instrument can be used
Illustrations must not contain elements of racism, violence, pornography, politics, or content that may cause negative impacts.	.87	Instrument can be used
Images, text, or quotations must be attributed to its source. The layout of images should be consistent.	.74	Instrument can be used

Based on Tables 3, 4, and 5, all questions items were declared as valid, meaning that the 10-item instrument may be used. Interpretation of validity is based on the validity coefficient value. If the validity coefficient value is less than .60, then the validity is considered low, whereas if it is in the range of .60 to .80 it is considered moderate, and if it is greater than .80 it is considered high.

3.2. Empirical Trial Data

After the theoretical test, the question items were tested empirically on sixth-grade high school students. This empirical trial was conducted at the SMA Negeri 2 Muara Beliti school, in Musi Rawas, South Sumatra, Indonesia, involving 102 students from three seventh-grade IPA (International Program on Science Education) classes. The quality of the test instrument was assessed according to several characteristics; validity, reliability, item proportion, and item discriminating power.

3.2.1 Validity Test

The test consisted of 10 items in the form of descriptions which were calculated according to the classical method using Microsoft Excel. The results were then interpreted using *r*-table at a significance level of 5%, $N = 102$ and $df = 100$. The *r*-table obtained was .1946 and all items were shown to be valid, as can be seen in Table 6.

Table 6. Results of Validity Test of Essay Question Items

Question item	r_{count}	r_{table}	Description
1	.4983	.1946	Valid
2	.4611	.1946	Valid
3	.6569	.1946	Valid
4	.6428	.1946	Valid
5	.6684	.1946	Valid
6	.3018	.1946	Valid
7	.4333	.1946	Valid
8	.3578	.1946	Valid
9	.6652	.1946	Valid
10	.3747	.1946	Valid

Based on Table 6, it can be concluded that the HOTS question item instrument for chemistry Volta cell material was of good quality in terms of its validity. The items were found to be able to measure what was intended since the number of valid items accounted for more than 40%.

3.2.2 Reliability Test

The HOTS instrument reliability test was based on Cronbach's alpha formula using Microsoft Excel. The results obtained from the reliability test were .7420, meaning that the instrument's reliability was categorized as high.

3.2.3 Proportion Test of Questions

The proportion of questions or measure of their difficulty level was tested using Microsoft Excel. The obtained results showed that three items were found to be in the difficult category, whilst seven were in the moderate category. Test results for the proportion of questions and their level of difficulty is presented as shown in Table 7.

Table 7. Result of Proportion Test of Question

Question item	Difficulty index	Description
1	.47	Moderate
2	.47	Moderate
3	.42	Moderate
4	.36	Moderate
5	.25	Difficult
6	.36	Moderate
7	.34	Moderate

Question item	Difficulty index	Description
8	.18	Difficult
9	.19	Difficult
10	.40	Moderate

3.2.4 Discriminant Test of Questions

Discriminant test of questions or differentiation of items was performed in order to measure the level of the upper and lower groups. The question discrimination test used the *t*-test calculation, and the results of the question discrimination test or differentiation power are presented in Table 8.

Table 8. Result Discriminant Test of Questions

Question item	T _{measure}	t _{table}	Description
1	5.46	.1984	Good
2	5.39	.1984	Good
3	6.46	.1984	Good
4	5.73	.1984	Good
5	4.23	.1984	Good
6	6.09	.1984	Good
7	2.28	.1984	Good
8	2.28	.1984	Good
9	6.96	.1984	Good
10	0.47	.1984	Good

4. Discussion

This research aimed to produce a HOTS instrument for use in chemistry learning on the subject of Voltaic cell material. The results of the study were 10 items in the form of descriptions with abilities at the C4 (analyze), C5 (evaluate), and C6 (create) levels, referring to higher order thinking skills in Bloom's revised taxonomy.

The HOTS instrument validation process involved several criteria, including assessment of its content, construction, and language aspects. Validity analysis was conducted by referring to the validators' assessment using V'Aiken's formula to calculate validity coefficients, and this theoretical validation performed by a group of six validators revealed the HOTS instrument to be valid and feasible for use in the context of chemistry learning.

Empirical validity was conducted with 102 students using Microsoft Excel. The results showed that all 10 items were valid, with validity greater than 40%. This result indicates that the HOTS instrument was found fit to measure the aspects it was intended to. In addition, the reliability test results showed a coefficient of .7420, meaning that it has a high level of reliability. In terms of discriminant power, all items in this study demonstrated good discriminant power, with three items categorized as difficult questions and seven items categorized as being of moderate difficulty. This shows that the HOTS instrument is able to differentiate between learners of high and low ability.

The developed HOTS instrument can therefore be effectively used to measure the high-level thinking ability of high school students in learning chemistry, especially on Voltaic cell material. Tests that utilize HOTS instruments have the potential to encourage students to develop higher order thinking skills in the subject (Barnett & Francis, 2011).

5. Conclusion

In forming conclusions about the current study, the following can be stated: 1) This research produced a HOTS instrument for chemistry learning on Voltaic cell material, which consists of 10 items in the form of descriptions. Validation of the HOTS instrument involved six validators, and the analysis revealed that the instrument was suitable for use after revision; 2) The characteristics of the developed HOTS instrument revealed that all 10 items were valid, with a reliability test coefficient value of .7420, which indicates a high level of reliability. The test of the proportion of questions in terms of their level of difficulty revealed that three questions were categorized as difficult and seven questions as moderately difficulty. In addition, the discriminant test of questions or the differentiation of questions showed that all items had good differentiation. Overall, the HOTS instrument developed in the current study can be considered as a valid, reliable, and differentiated evaluation tool for use in measuring students' higher order thinking skills on Voltaic cell material in chemistry learning.

The HOTS instrument developed in the current study in the form of description questions was shown to have met the required standards of item eligibility, according to its validity, reliability, proportion of questions, and discriminant questions. Application of these items can be expected to provide benefits for students as practice materials used to help improve their higher order thinking skills. In addition, these items can also be used as an alternative instrument for teachers in preparing HOTS-based questions in chemistry learning.

One limitation of the current study was that the trial was limited to a sample of three classes and did not involve a field trial. Therefore, it is recommended that future research conducts field trials involving a wider range of respondents. In addition, the developed HOTS instrument was limited to a single basic competency; hence, future studies are recommended that develop HOTS-based instruments aimed at other basic competencies.

Declarations

Author Contributions. M.A, V.D.A.: Literature review, conceptualization, methodology, data analysis. M.A., V.D.A., J.I., M.A.C., E.P.: Review-editing and writing, original manuscript preparation. All authors have read and approved the published final version of the article.

Conflicts of Interest. The authors declared no potential conflicts of interest.

Ethical Approval: The author declares that this work is original and has not been previously published.

Funding. None

Data Availability Statement: Data is available upon request.

References

- Abidin, A., Istiyono, E., Fadilah, N., & Dwandaru, W. (2019). A computerized adaptive test for measuring the physics critical thinking skills. *International Journal of Evaluation and Research in Education*, 8(3), 376-383. <http://doi.org/10.11591/ijere.v8i3.19642>
- Abosalem, Y. (2016). Assessment Techniques and Students' Higher-Order Thinking Skills. *International Journal of Secondary Education*, 4(1), 1-11. <http://doi.org/10.11648/j.ijsedu.20160401.11>
- Afflerbach, P., Cho, B.-Y., & Kim, J.-Y. (2015). Conceptualizing and Assessing Higher-Order Thinking in Reading Order Thinking in Reading. *Theory Into Practice*, 54(3), 203-212. <http://dx.doi.org/10.1080/00405841.2015.1044367>
- Aman. (2019). Final Examination Test Instruments for History Subject in Yogyakarta, Indonesia: A Quality Analysis. *Universal Journal of Educational Research*, 7(12), 2857-2866. <http://doi.org/10.13189/ujer.2019.071237>
- Arifin, Z. (2016). *Evaluasi Pembelajaran* [Learning Evaluation]. Remaja Rosdakarya.

- Atitullah, K., Fitriati, S., & Rukmini, D. (2019). Using Revised Bloom's Taxonomy to Evaluate Higher Order Thinking Skills (Hots) in Reading Comprehension Questions of English Textbook for Year X of High School. *English Education Journal*, 9(4), 428-436. <https://doi.org/10.15294/eej.v9i4.31794>.
- Avargil, S., Herscovitz, O., & Dori, Y. (2012). Teaching Thinking Skills in Context-Based Learning: Teachers' Challenges and Assessment Knowledge. *Journal of Science Education and Technology*, 21, 207-225. <http://dx.doi.org/10.1007/s10956-011-9302-7>
- Ayubi, M., Arthamena, V., & Aznam, N. (2023). Development of HOTS (High Order Thinking Skill) Question Instrument on Thermochemistry Material. *Jurnal Penelitian Pendidikan IPA*, 9(11), 9945-9954. <https://doi.org/10.29303/jppipa.v9i11.5162>
- Azwar, S. (2016). Reliabilitas dan Validitas Aitem [Reliability and Validity of Aitems]. *Buletin Psikologi*, 3(1), 19-26. <https://jurnal.ugm.ac.id/buletinpsikologi/article/view/13381>
- Barbara, L., & Wendy, W. (2010). Developing Higher Level Thinking. *Journal of Instructional Pedagogies*, 3(1), 1-9. <https://www.aabri.com/manuscripts/09423.pdf>
- Barnett, J., & Francis, A. (2011). Using higher order thinking questions to foster critical thinking: a classroom study. *Educational Psychology*, 32(2), 201-211. <https://doi.org/10.1080/01443410.2011.638619>
- Boone, W., Staver, J., & Yale, M. (2014). *Rasch Analysis in the Human Sciences*. SpringerLink.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189-193.
- Brown, G., Irving, S., & Keegen, P. (2018). *An Introduction to Educational Assessment, Measurement, and Evaluation: Improving the Quality of Teacher-Based Assessment*. Pearson Education.
- Cavallo, M., Potter, W., & Rozman, M. (2010). Gender Differences in Learning Constructs, Shifts in Learning Constructs, and Their Relationship to Course Achievement in a Structured Inquiry, Yearlong College Physics Course for Life Science Majors. *School Science and Mathematics*, 104(6), 288-300. <https://doi.org/10.1111/j.1949-8594.2004.tb18000.x>
- Chalkiadaki, A. (2018). A Systematic Literature Review of 21st Century Skills and Competencies in Primary Education. *International Journal of Instruction*, 11(3), 1-16. <https://doi.org/10.12973/iji.2018.1131a>
- Chinedu, C., Kamin, Y., & Olabiyi, O. (2015). Strategies for Improving Higher Order Thinking Skills in Teaching and Learning of Design and Technology Education. *Journal of Technical Education and Training*, 7(2), 35-43. <https://publisher.uthm.edu.my/ojs/index.php/JTET/article/view/1081>
- Clotfelter, C., Ladd, H., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377-392. <https://doi.org/10.1016/j.econedurev.2004.06.008>
- Conklin, W. (2011). *Higher-Order Thinking Skills to Develop 21st Century Learners*. Shell Education.
- Cooper, C. (2023). *An Introduction to Psychometrics and Psychological Assessment*. Routledge.
- D'Alessio, F., Avolio, B., & Charles, V. (2019). Studying the impact of critical thinking on the academic performance of executive MBA students. *Thinking Skills and Creativity*, 31, 275-283. <https://doi.org/10.1016/j.tsc.2019.02.002>
- Demirhan, E., & Koklukaya, A. (2014). The Critical Thinking Dispositions of Prospective Science Teachers *Procedia-Social and Behavioral Sciences*, 116, 1551-1555. <https://doi.org/10.1016/j.sbspro.2014.01.433>
- Drake, F., & Nelson, L. (2008). *Engagement in Teaching History: Theory and Practices for Middle and Secondary Teachers* (2nd ed.). Pearson.
- Febrina, F., Usman, B., & Muslem, A. (2019). Analysis of Reading Comprehension Questions by Using Revised Bloom's Taxonomy on Higher Order Thinking Skill (HOTS). *English Education Journal*, 10(1), 1-15. <https://jurnal.usk.ac.id/EEJ/article/view/13253/10123>
- Febriyani, R., Yunita, W., & Damayanti, I. (2020). An Analysis on Higher Order Thinking Skill (HOTS) in Compulsory English Textbook for the Twelfth Grade of Indonesian Senior High Schools. *Journal of English Education and Teaching*, 4(2), 170-183. <https://doi.org/10.33369/jeet.4.2.170-183>

- Fischer, C., Bol, L., & Pribesh, S. (2011). An Investigation of Higher-Order Thinking Skills in Smaller Learning Community Social Studies Classrooms. *American Secondary Education*, 39(2), 5-26.
- Gall, M. D., Gall J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Allyn, & Bacon/Pearson.
- Gojkov, G., Stojanovic, A., & Rajic, A. (2015). Critical Thinking Of Students – Indicator Of Quality In Higher Education. *Procedia-Social and Behavioral Sciences*, 191, 591-596. <https://doi.org/10.1016/j.sbspro.2015.04.501>
- Guskey, T. (2007). Closing Achievement Gaps: Revisiting Benjamin S. Bloom’s “Learning for Mastery”. *Journal of Advanced Academics*, 19(1), 8-31. <https://doi.org/10.4219/jaa-2007-704>
- Haley, S., Coster, W., Dumas, H., Fragala-Pinkham, M., Kramer, J., Ni, P., Tian, F., Kao, Y.-C., Moed, R., & Hudlow, L. (2011). Accuracy and precision of the Pediatric Evaluation of Disability Inventory computer-adaptive tests (PEDI-CAT). *Developmental Medicine and Child Neurology*, 53(12), 1100-1106. <http://doi.org/10.1111/j.1469-8749.2011.04107.x>
- Halpern, D. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449-455. <https://psycnet.apa.org/doi/10.1037/0003-066X.53.4.449>
- Heong, Y., Othman, W., Yunos, J., Kiong, T., Hassan, R., & Mohamad, M. (2011). The Level of Marzano Higher Order Thinking Skills among Technical Education Students. *International Journal of Social Science and Humanity*, 1(2), 121-125. <http://dx.doi.org/10.7763/IJSSH.2011.V1.20>
- Hmeli, C., & Ferrari, M. (1997). The Problem-Based Learning Tutorial: Cultivating Higher Order Thinking Skills. *Journal for the Education of the Gifted*, 20(4), 401-422. <https://doi.org/10.1177/016235329702000405>
- Kinay, I., & Bagceci, B. (2016). The Investigation of the Effects of Authentic Assessment Approach on Prospective Teachers’ Problem-Solving Skills. *International Education Studies*, 9(8), 51-59. <http://dx.doi.org/10.5539/ies.v9n8p51>
- Kusuma, M., Rosidin, U., Abdurrahman, & Suyanta, A. (2017). The Development of Higher Order Thinking Skill (Hots) Instrument Assessment In Physics Study. *IOSR Journal of Research & Method in Education*, 7(1), 26-32. <http://doi.org/10.9790/7388-0701052632>
- Madhuri, G., Kantamreddi, V., & Goteti, L. (2012). Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education*, 37(2), 117-123. <https://doi.org/10.1080/03043797.2012.661701>
- Mahoney, J., & Harris-Reeves, B. (2019). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education*, 20(1), 25-37. <https://doi.org/10.1177/1469787417723243>
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes* [Techniques for Preparing Test and Non-Test Instruments]. Mitra Cendekia Press.
- McCormick, N. (2015). Engaging Students in Critical Thinking and Problem Solving: A Brief Review of the Literature. *Journal of Studies in Education*, 5(4), 100-113. <http://dx.doi.org/10.5296/jse.v5i4.8249>
- McDonald, G. (2012). Teaching Critical & Analytical Thinking in High School Biology? *The American Biology Teacher*, 74(3), 178-181. <https://doi.org/10.1525/abt.2012.74.3.9>
- McLaughlin, J., Roth, M., Glatt, D., Gharkholonarehe, N., Davidson, C., Griffin, L., & Mumper, R. (2014). The flipped classroom: a course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2), 236-243. <https://doi.org/10.1097/acm.0000000000000086>
- Miterianifa, M. (2016). *Evaluasi Pembelajaran Kimia* [Evaluation of Chemistry Learning]. Cahaya Firdaus.
- Polit, D., & Beck, C. (2006). The content validity index: Are you sure you know what’s being reported? critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497. <https://doi.org/10.1002/nur.20147>
- Pommerich, M. (2004). Developing Computerized Versions of Paper-and-Pencil Tests. *The Journal of Technology, Learning and Assessment*, 2(6), 3-44. <https://ejournals.bc.edu/index.php/jtla/article/view/1666>

- Qing, Z., Jing, G., & Yan, W. (2010). Promoting preservice teachers' critical thinking skills by inquiry-based chemical experiment. *Procedia-Social and Behavioral Sciences*, 2(2), 4597-4603. <https://doi.org/10.1016/j.sbspro.2010.03.737>
- Quieng, M., Lim, P., & Lucas, M. (2015). 21st Century-based Soft Skills: Spotlight on Non-cognitive Skills in a Cognitive-laden Dentistry Program. *European Journal of Contemporary Education*, 11(1), 72-81. <http://doi.org/10.13187/ejced.2015.11.72>
- Retnawati, H. (2016). *Validitas, Reliabilitas dan Karakteristik item* [Validity, Reliability and Characteristics of items]. Parama Publikasi.
- Sepriyanti, N., Nelwati, S., Kustati, M., & Afriadi, J. (2022). The Effect of 21st-Century Learning on Higher-Order Thinking Skills (Hots) and Numerical Literacy of Science Students in Indonesia Based on Gender. *Jurnal Pendidikan IPA Indonesia*, 11(2), 314-321. <http://doi.org/10.15294/jpii.v11i2.36384>
- Setiawan, J., Aman, A., & Wulandari, T. (2020). Understanding Indonesian history, interest in learning history and national insight with nationalism attitude. *International Journal of Evaluation and Research in Education*, 9(2), 364-373. <http://doi.org/10.11591/ijere.v9i2.20474>
- Sudijono, A. (2012). *Pengantar evaluasi pendidikan* [Introduction to educational evaluation]. Rajawali Pers.
- Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan* [Introduction to educational evaluation]. Raja Grafindo Persada.
- Szilagyi, J., Clements, D., & Sarama, J. (2013). Young Children's Understandings of Length Measurement: Evaluating a Learning Trajectory. *Journal for Research in Mathematics Education*, 44(3), 581-620. <https://doi.org/10.5951/jresmetheduc.44.3.0581>
- Udompong, L., & Wongwanich, S. (2014). Diagnosis of the Scientific Literacy Characteristics of Primary Students. *Procedia-Social and Behavioral Sciences*, 116, 5091-5096. <https://doi.org/10.1016/j.sbspro.2014.01.1079>
- Wardany, K., Sajidan, & Ramli, M. (2017). Pengembangan Penilaian Untuk Mengukur Higher Order Thinking Skills Siswa [Development of Assessments to Measure Students' Higher Order Thinking Skills]. *Jurnal Inkuiri*, 6(2), 1-16. <https://jurnal.uns.ac.id/inkuiri/article/view/16214>
- Weiss, D. (2011). Better Data From Better Measurements Using Computerized Adaptive Testing. *Journal of Methods and Measurement in the Social Sciences*, 2(1), 1-27. <https://doi.org/10.2458/v2i1.12351>

Publisher's Note: Universitepark Limited remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
