DEVELOPMENT OF COGNITIVE ASSESSMENT INSTRUMENTS ON BUFFER SOLUTION MATERIAL BASED ON HOTS (HIGHER ORDER THINKING SKILLS)

Nurmiana Suswanti[1], Febrian Solikhin[2*], Sura Menda Ginting[3]

[123]Program Studi Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Bengkulu

**Corresponding author: febrian.solikhin@unib.ac.id*

Abstract

Abstract. *The goal of this study is to construct a valid and reliable HOTS-based cognitive assessment tool. The HOTS-based assessment instrument was used to measure students' HOTS on the buffer solution material. The stages of this research were the analysis, design, development, implementation, and evaluation phases of the ADDIE model. The research subjects were students of 11th grade at Senior High School in Bengkulu City during the 2021–2022 academic year. The research data came from expert validation sheets, needs analysis questionnaires, empirical test analysis, and examination of students' HOTS skills. The 20 multiple-choice questions and 10 essay questions made up the instrumen's features. The following are the outcomes: (1) The expert's assessment of the practicality of the material led to the creation of 26 valid questions, including 18 multiple-choice questions and 8 essays. (2) The empirical validity study produced 23 valid items, including 6 essay questions and 17 multiple-choice questions. (3) The reliability analysis results for both essays and multiple-choice questions fell into the very good category, with scores of 0.94 and 0.98. (4) The findings of the examination of students' higher-order thinking skills revealed that 1 student was in the very good category, 5 students were in the good category, 5 students were in the fairly good category, 22 students were in the poor category, and 1 student was in the very poor category.*

Keywords: *Development, Cognitive Assessment, HOTS, Buffer Solution*

PENDAHULUAN/INTRODUCTION

Education is a component that has a dynamic state due to competence demands personality changes from time to time. Currently, education is directed at forming changes and improvements related to three things, namely increasing eligibility, quality, and competitiveness. Improvement efforts the quality of an education system can be achieved by increasing the quality of learning and the quality of the assessment system. The two components are interrelated because if the learning system is carried out well, it will produce good quality human resources with integrity. In the world of education, the curriculum is a component that has an important role in determining the sustainability of the education system. The 2013 curriculum mandates that students take a more active role in expanding their knowledge and honing their critical-thinking abilities, among other things [1]. Additionally, the 2013 curriculum mandates that a teacher possess knowledge of creating assessment tools based on higher-order thinking abilities that can foster critical and creative thought processes [2]. Higher order thinking skills (HOTS) is one of the abilities needed in the 21st century. Therefore, education is needed that can develop high-order thinking skills for students. 21st century learning must train students to have 4 competencies (4C) consisting of the ability to creative thinking, critical thinking and problem solving, communication, and collaboration [3].

An educator needs to know the students' understanding level of the topics being taught and to know the extent to which students are trying to construct their knowledge in analyzing the concept. This can be seen by evaluating learning outcomes that emphasize higher-order thinking skills. Evaluation is a cycle for designing,

obtaining, and providing information that is expected to complete several choices in making decisions [4]. Evaluation of learning outcomes is an important part of determining the achievement of a learning process. In carrying out a learning evaluation, an assessment instrument is needed.

In the 2013 curriculum, assessment of learning outcomes is expected to help students improve their thinking skills, one of which is HOTS. The Indonesian National Assessment is directed at an assessment model that requires thinking skills that don't just remember, restate, or refer to without processing [5]. According to Widodo and Wehlage, Higher-order thinking skills will distinguish ideas or ideas clearly, be able to construct descriptions, be able to argue well, be able to hypothesize, be able to solve problems, and be able to understand things that are more complex [6]. Higher-order thinking occurs when students are able to analyze and create the knowledge they get to produce new knowledge. HOTS has a cognitive domain which includes C4 (analysis), C5 (evaluation), and C6 (creation) levels.

The 2018 PISA survey findings, as reported by Schleicher in Agustina indicate that Indonesian students are placed 71 out of 79 nations in terms of their scientific aptitude [7]. These findings suggest that Indonesian students still have a limited capacity for higher-order thought. Because they still do not fully comprehend the assessment process, the majority of teachers are still inadequate when it comes to assessing students' higher-order thinking [8].

An examination of many Senior High Schools in Bengkulu City supports this. Teachers of chemistry were given questionnaires to fill out as part of the study. Additionally, findings about questions that were frequently asked both during exams and the learning process were made. According on survey findings and observations, teachers continue to use tests designed to assess higher order thinking skills very infrequently. The assessment instrument used in schools is in the form of questions that tend to test more on aspects of memory. In this case, the books used already present material that invites students to study actively and have a systematic concept presentation. However, the presentations in the book often end with questions that do not train students' higher-order thinking skills, such as the use of operational verbs in questions that clearly describe problem solving in these questions without involving the student's analysis process. In addition, teachers have problems in compiling higher-order thinking ability test instruments, namely: (1) it takes a long time to prepare, (2) lack of understanding of higher-order thinking, (3) there is no idea for compiling higher-order thinking skills questions. Based on the description above, the researcher is interested in making a higher-order thinking skills-oriented assessment instrument which is valid and reliable which can be used to measure students' cognitive level abilities in meeting the needs of the 21st century.

MATERIALS AND METHODS

The development model that is used as a reference in this study is the ADDIE model from Branch [9], which is a model with 5 stages, including: Analysis, Design, Development, Implementation, and Evaluation. The subjects in this study include the subject of empirical validation from 3 high school students in Bengkulu City, as well as the subject of implementation carried out in one of 11^{th} grade among the high schools as many as 34 students to measure thinking skills at the level of tall.

The development procedure includes analysis of problems The research instruments used were needs analysis questionnaires, expert validation instruments, and product testing instruments. Data analysis techniques included expert validation test analysis, normality test, homogeneity test, empirical validation analysis, reliability test analysis, and HOTS-based test result analysis.

The data analysis technique used in the validation of HOTS questions with the assessment aspect is carried out on material, construction, and language in the form of an assessment scale. The type of scale used is a *Likert* with a score of 1 to 3. The validity of the

HOTS questions performs calculations using the V-Aiken formula:

$$V = \frac{S}{[n(c-1)]}$$

Keterangan:

Information:

V = rater/evaluator agreement index

S = score set by each rater minus the lowest score in the category

n = number of raters

c = number of categories that raters can choose

The level of validity can be determined by matching the calculation results with the criteria in the Aiken validity table [10].

Empirical validation analysis technique was carried out using Rasch modeling with *Winstep 3.73 software* using *item fit* for each item. Examination of items that can be said to be valid using Rasch modeling follows 3 criteria, namely *outfit means-square* (MNSQ) with a coefficient of 0.5 < MNSQ < 1.5, *outfit* Zs*ZSTD*) with a coefficient of -2 < ZSTD < +2, and *point measure criteria corr* with a coefficient of 0.4 < Pt Measure Corr < 0.8 [11].

Analysis of the reliability of the test instrument was carried out with the help of the Winsteps 3.73 program. The criteria *item reliability* shown in Table 1.

Table 1. Test Instrument Reliability Criteria

Score	Category
0,94 ≤ r ≤ 1,00	Special
0,91 ≤ r < 0,94	Very Good
0,80 ≤ r < 0,90	Good
0,67 ≤ r < 0,80	Enough
r < 0,67	Weak

Analysis of higher order thinking skills is carried out by calculating student test results with scoring guidelines that have been designed which are then interpreted according to Table 2

Table 2. Categories of Students' HOTS Ability

Guidelines	Category
x ≥ (μ+1,5σ)	Very good
(μ+0,5σ) ≤ x < (μ+1,5σ)	Good
(μ-0,5σ) ≤ x < (μ+0,5σ)	Fairly Good
(μ-1,5σ) ≤ x < (μ-0,5σ)	Not Good
x < (μ-1,5σ)	Very Poor

Information

μ = ideal average

μ = (Highest ideal score + ideal lowest score)

σ = Deviation

σ = (Ideal highest score – ideal lowest score)

RESULTS AND DISCUSSION

Analysis

The first analysis carried out is needs analysis. This was carried out by distributing questionnaires to 6 teachers from different high schools and grade levels in Bengkulu City. The results of the analysis show that teachers tend to use questions that are similar to the examples of questions found in the textbooks used during the learning process. However, the questions in these textbooks tend to be more dominated by LOTS (*Low Order Thinking Skills*) questions or the cognitive level of remembering (facts and concepts) and MOTS (*Middle Order Thinking Skills*) or the cognitive level of understanding and applying [12]. Teachers tend to use LOTS and MOTS questions due to the unavailability of valid higher-order thinking questions used by teachers to support the learning process.

The second analysis carried out was an analysis of indicators of achievement of HOTS competencies based on the 2013 curriculum. The results of the analysis of achievement indicators were associated with the context of preparing HOTS questions and operational verbs used as shown in Table 3.

Table 3. Analysis of HOTS Competency Achievement Indicators

Basic Competency	Indicators for Competency Achievement
Explaining the working principle, calculation, and the role of buffer solutions in	Analyzing buffer solutions and non-buffer solutions based on their constituent components
	Analyzing the properties and workings of buffer solutions in maintaining pH

Basic Competency	Indicators for Competency Achievement
living organisms	Analyzing and proving calculations buffer solution
	Analyze the role of buffer solution in the body

Design

At the design stage, the outcomes include choosing the multiple-choice and essay format for the instrument with a total of 20 multiple-choice questions and 10 essays., creating a grid based on the GPA analysis of the chosen material, and designing the instrument that will be created.

Development

Expert Validation and Revision

The instrument has been revised, and improvements have been made to the items' suitability for indicators, the writing's inappropriateness, the questions' lack of guidance or instructions, the ambiguous and ineffective use of words, the items' suitability for the bloom taxonomy level (C4-C6), and the item's suitability for the concept of the material. There are four questions in expert validation that are invalid: PG-12, PG-15, Es-7, and Es-10. In addition, there are three questions that need to be revised, namely: PG-9, PG-17, and PG-19. example of a revised question

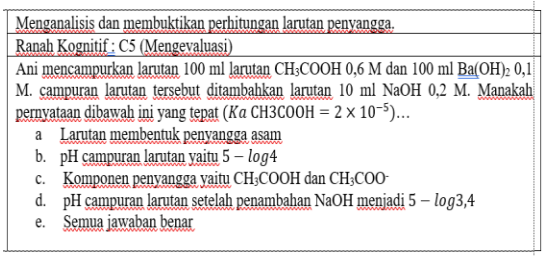
Menganalisis dan membuktikan perhitungan larutan penyangga.
Ranah Kognitif : C5 (Mengevaluasi)
Ani mencampurkan larutan 100 ml larutan CH_3COOH 0,6 M dan 100 ml $Ba(OH)_2$ 0,1 M. campuran larutan tersebut ditambahkan larutan 10 ml NaOH 0,2 M. Manakah pernyataan dibawah ini yang tepat (Ka CH3COOH $= 2 \times 10^{-5}$)…
a Larutan membentuk penyangga asam
b. pH campuran larutan yaitu $5 - log4$
c. Komponen penyangga yaitu CH_3COOH dan CH_3COO^-
d. pH campuran larutan setelah penambahan NaOH menjadi $5 - log3{,}4$
e. Semua jawaban benar

Figure 1 PG-19 Before Revision

In this question, according to the validator students will tend to choose answer E by only knowing the truth of the two detractors in the answer choices. revision results as shown in Figure 1

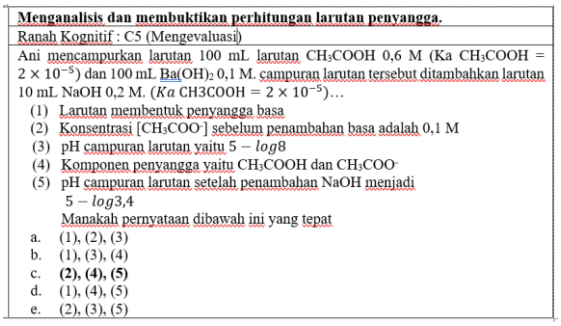
Menganalisis dan membuktikan perhitungan larutan penyangga.
Ranah Kognitif : C5 (Mengevaluasi)
Ani mencampurkan larutan 100 mL larutan CH_3COOH 0,6 M (Ka $CH_3COOH = 2 \times 10^{-5}$) dan 100 mL $Ba(OH)_2$ 0,1 M. campuran larutan tersebut ditambahkan larutan 10 mL NaOH 0,2 M. (Ka CH3COOH $= 2 \times 10^{-5}$)…
(1) Larutan membentuk penyangga basa
(2) Konsentrasi $[CH_3COO^-]$ sebelum penambahan basa adalah 0,1 M
(3) pH campuran larutan yaitu $5 - log8$
(4) Komponen penyangga yaitu CH_3COOH dan CH_3COO^-
(5) pH campuran larutan setelah penambahan NaOH menjadi $5 - log3{,}4$
Manakah pernyataan dibawah ini yang tepat
a. (1), (2), (3)
b. (1), (3), (4)
c. **(2), (4), (5)**
d. (1), (4), (5)
e. (2), (3), (5)

Figure 2 PG-19 After Revision

Empirical Validation Test Results

Rasch modeling is used to examine items that meet three criteria: outfit means-square, outfit Z-standard, and point measure correlation. There are four questions that cannot be validated empirically: PG-9, PG-18, Es-4, and Es-5. The MNSQ logit value of more than 1.5 indicates that *the item* is indicated to be confusing for students to answer, while the ZSTD which is more than 2 and has a positive value indicates that the answers given vary [13]. In the *point measure corr value*, if the *logit* is less than 0.4 it explains that the question items are less able to distinguish students' abilities and if the *point measure corr* negative it explains that the item can mislead students in answering because low-ability students answer correctly and students who are capable high answered incorrectly.

Table 4 Results of Valid Criteria Values

Question code	MNSQ	ZSTD	PT. Measure Corr
PG-9	1,88	3,2	-0,04
PG-18	1,87	2,1	0,12
Es-4	1,52	3,6	0,75
Es-5	0,45	-0,8	0,36

Students select more distracting answers than accurate ones for the PG-9 and PG-18 scales, which include a variety of answers that do not follow Rasch's modeling. Students are prone to guess the answers, which can lead to discrepancies between the answers provided and suggest cheating. The Es-4 item shows that the question is confusing which causes students to still have errors in making a reaction from the added acid-base solution, calculating the pH without involving an analysis of its working principle, and there are still many who lack understanding in relating the concepts that are related to each other. Es-5

questions have a difficulty level of items in the "very difficult" category. If an item has a high level of difficulty, then the item cannot be a measurement parameter because it is difficult to distinguish students' abilities.

According to the reliability test's findings, essays and multiple-choice items were designated as special due to their unique reliability values of 0.94 and 0.98, respectively.

Implementation

Valid and reliable questions based on empirical test results are then implemented to students of 11th grade in high schools, totaling 34 students. Questions that are valid and reliable on empirical test results are then implemented to students in this class. The results of students' HOTS can be seen in Table 7.

Table 5 Results of Students' Higher-order Thinking Ability

Value range	Total	category
$x \geq 19{,}495$	1	Very good
$15{,}165 \leq x < 19{,}495$	5	Good
$10{,}835 \leq x < 15{,}165$	5	Fairly Good
$6{,}505 \leq x < 10{,}835$	22	Not Good
$x < 6{,}505$	1	Very Poor

The implementation results in Table 5 show that most students are in the cognitive domain with a poor category with a total of 22 students, and there is only 1 student in the very good category and 5 students in the good category, 5 students in the fairly good category and 1 student in the good category. very less. There are still many students in the unfavorable category because the ability of students to think at a higher level is still low. According to Kurniat, one of the factors that causes low thinking skills is students' lack of training in solving questions that demand analysis, evaluation, and creativity [14]. In addition, another factor that can also affect students' thinking skills is the learning process experienced by students. This is related to the way teachers transfer their knowledge to students where some teachers only convey material, but do not invite students to actively explore and discover the concept for themselves so that they can understand the relationship between other parts/factors. Most teachers still do not understand about effective and appropriate learning to achieve learning objectives or improve students' thinking skills. Teachers are still rare in training students with examples of HOTS-based questions so that students are less accustomed to solving problems in the cognitive domain of analyzing, evaluating, and creating. As a result, there is a tendency for teachers to carry out learning by only transferring knowledge or material they know from books to their students. In fact, the presence of HOTS questions in the learning process can be the first step in meeting the assessment standards in the 2013 curriculum which demands a national assessment that does not only train students in aspects of memory, understanding and application. There are supporting factors that apply learning accompanied by HOTS practice questions, so the standard assessment carried out on students can apply variations of item questions with the LOTS, MOTS, and HOTS cognitive domains.

Carlgreen's research concluded that there were obstacles faced by students, namely in students' critical thinking, student communication, and problem solving faced by students. This is due to three factors, namely the structure of this education system, the complexity of student skills, and the competence of teachers in teaching [15]. This needs to be anticipated by educators that teaching is not just helping students work on problems, but needs to equip students with the ability to apply and reason about a problem they face.

The results of the analysis using the Winsteps program with Rasch modeling on implementation students, it can be seen that there are 6 students (32P, 01P, 10L, 11L, 5L, 02P) who need to be suspected in answering certain questions, for example students whose answers should be wrong, but the student answered correctly. Conversely, students who should have answered correctly, but answered incorrectly on the item. This relates to students with low abilities being able to answer difficult items correctly, but on the other hand students with high abilities answered wrong items that were easy. This can be caused by several factors such as the possibility of students guessing answers, cheating, or students being careless in answering. In addition, from the results of the analysis of essay answers, most of the students

were still unable to parse the information about the questions to formulate the right solutions

Evaluation

Design: the instrument in collecting data should not only use questionnaires but also need to be interviewed. Planning: you should always discuss with the teacher to find out the description of the problem design with consideration of the student's condition. Development: revision of expert and empirical validation. Implementation: limited time in working on the questions.

CONCLUSION AND SUGGESTION

Based on the results of research that has been done, it can be concluded as follows:

1. The HOTS-based cognitive evaluation instrument was constructed, and the results of the expert's feasibility test using the Aiken formula revealed that 26 of the 30 developed questions were valid at the expert validation stage. 18 multiple-choice questions and 8 essay questions make up valid questions. Questions PG-12, PG-15, Es-7, and Es-10 are all invalid. Before moving on to the empirical test, there are three valid questions PG-9, PG-17, and PG-19—that need to be updated.
2. Out of the 26 questions that have been evaluated by experts, 4 questions have not passed the criteria and therefore cannot be used for empirical validation. The ratings of PG-9, PG-18, Es-4, and Es-5 are all bogus. With a reliability score of 0.94 and 0.98 for each, the results of the reliability test on multiple-choice items and essays show very high dependability, demonstrating the instrument's high level of consistency.
3. In implementation, 34 students participated in the analysis of high-level thinking skills. The results showed that very good categories had 1 student, good categories had 5 students, quite good categories had 5 students, poor categories had 22 students, and very less had 1 student.

The suggestion in this research is that it is better for further development, in making a HOTS-based cognitive assessment instrument to also pay attention to the level of difficulty and the differentiating power of the questions. This is to show how the difficulty level of each item is and the differentiating power of the questions related to the students' ability to solve the problem. Thus, it can be an evaluation of the items and the learning process that takes place to improve students' abilities at a high level.

ACKNOWLEDGEMENTS

Thank you to all parties involved and have helped in this research, from preparation, research process, to the final reporting of research results.

REFERENCES

[1] Nurhasanah, N., & Auliyati, Y. (2018). Pengembangan Nilai Karakter Siswa Melalui Pembelajaran Berbasis Higher Order Thinking Skills di Sekolah Dasar. *Jurnal Pemberdayaan Sekolah Dasar (JPSD)*, *1*(1), 7–12.

[2] Khaldun, I., Hanum, L., & Utami, S. D. (2020). Pengembangan Soal Kimia Higher Order Thinking Skills Berbasis Komputer Dengan Wondershare Quiz Creator Materi Hidrolisis Garam Dan Larutan Penyangga. *Jurnal Pendidikan Sains Indonesia*, *7*(2), 132–142.

[3] Nofrion, N., & Wijayanto, B. (2018). Learning Activities in Higher Order Thinking Skill (Hots) Oriented Learning Context. *Geosfera Indonesia*, *3*(2), 122.

[4] Purwanto. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Belajar. 2012.

[5] Gradini, E. (2019). Menilik Konsep Kemampuan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) dalam Pembelajaran Matematika. *Jurnal Numeracy*, *6*(2), 189–203.

[6] Widodo, T & Kadarwati, S. (2013). High Order Thinking *Berbasis* Pemecahan Masalah Untuk Meningkatkan Hasil BelajarBerorientas PembentukanKarakter Siswa. *Cakrawala Pendidikan*, 32(1): 161-171

[7] Agustina, L., Feronika, T., & Yunita, L. (2021). Analysis of Higher Order Thinking Skills Questions in the Brookhart Category in High School Chemistry Textbook: Curriculum 2013. *Journal of Educational Chemistry (JEC)*, *3*(1), 23–34.

[8] Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R. D. (2018).

Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century*, *76*(2), 215–230.

[9] Branch, R. M., *Instructional Design: The ADDIE Approach*. New York: Springer, 2009.

[10] Aiken, L. R. (1985). Three Coefficients For Analyzing The Reliability And Validity Of Ratings. *Educational and Psychological Measurement*, *45*, 131–141.

[11] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH Pada Assessment Pendidikan*. Trim Komunikata.

[12] Himmah, W. I. (2019). Analisis Soal Penilaian Akhir Semester Mata Pelajaran Matematika Berdasarkan Level Berpikir. *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, *3*(1), 55.

[13] Bond, T., Yan, Z., & Moritz, H. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences* (3rd editio). Routledge.

[14] Kurniati, D., Harimukti, R., & Jamil, N. A. (2016). Kemampuan berpikir tingkat tinggi siswa SMP di Kabupaten Jember dalam menyelesaikan soal berstandar PISA. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(2): 142-155.

[15] Carlgreen, T. (2013). Communication, Critical Thinking, Problem Solving: A Suggested Course for A High School Students in the 21st Century. Interchange, 44(1-2), 63-81