Thabiea : Journal of Natural Science Teaching
Vol. 8(1), pp. 60-78, 2025
Available online at
http://journal.iainkudus.ac.id/index.php/Thabiea
p-issn: 2580-8474, e-issn: 2655-898X

Assessment tools development for enhancing higher order thinking skills in Islamic contextualized stoichiometry

Didan Sopian[1], Muhammad Zamhari[2]*
[1,2]Chemistry Education Department, State Islamic University Sunan Kalijaga, Yogyakarta, Indonesia
**Correspondence: muhammad.zamhari@uin-suka.ac.id*

Keywords:
HOTS;
Test instrument;
Stoichiometry;
Science-Islamic integration;

Abstract

This research seeks to create a higher-order thinking skills (HOTS) assessment for stoichiometry that integrates Islamic principles, ensuring it meets the criteria of validity, reliability, difficulty level, and distinctiveness. The integration is essential to combine scientific understanding with spiritual values and to enhance ethics in science. The research employed a research and development study with the Tessmer model, ,consisting of preliminary, self-evaluation, prototyping, and field test stages. Product quality was validated by a material expert, a media expert, and an Islamic Integration expert with a score of 98.96%, 99.03%, and 73.13%, respectively. Then, the product was also assessed by six reviewers (chemistry teachers) with a score of 93.26% as a very good category. The product was tested on 75 students in 10th grade with 16 valid items and four invalid items with very high reliability, namely 0.843. The invalid items was then discarded. At the difficulty level, the product contains six easy, ten medium, and four difficult questions, with more than 75% of the questions able to distinguish students with high and low abilities. Therefore, the developed instrument is suitable for measuring HOTS and Islamic integration on the material of stoichiometry. However, the implementation of questions combining the teaching of chemistry and Islam can be utilized to produce better research outcomes.

To cite this article:
Sopian, D., & Zamhari, M. (2025). Assessment tools development for enhancing higher order thinking skills in Islamic contextualized stoichiometry. *Thabiea: Journal of Natural Science Teaching*, 8(1), 60-78.

Introduction

Adopting a learning approach focused on higher-order thinking skills (HOTS) offers a valuable strategy to meet the challenges of the 21st century (Dwijayanti, 2021). Student currently should have 4 C skills, i.e. creativity and innovation, critical thinking and problem-solving, communication and collaboration (Thornhill-Miller et al., 2023). These skills with HOTS synergistically improve the quality of knowledge, conceptualization, recognition, understanding, determination, and reasoning of students (Arthadewi et al., 2024).

HOTS are students' thinking skills at a high cognitive level to solve problems developed from various learning concepts, methods, and taxonomies (Amali et al., 2022). Based on Bloom's Taxonomy which Krathwoll and Anderson have revised Syahri Ramdhani et al., (2024) the ability of students not only at the LOTS (Lower Order Thinking Skills) level, which includes knowing skills (C1), understanding (C2), and applying (C3). However, there must be an increase in the HOTS (Higher Order Thinking Skills) level, namely the skills to analyze

(C4), evaluate (C5), and create (C6). However, the reality on the ground shows that students in Indonesia are still classified as at the LOTS level (Kurniawan et al., 2024). It is evidenced by the results of the HOTS ability evaluation study of students by PISA (Program for International Student Assessment) in 2018, where Indonesia was ranked 72th of 77 participating countries (OECD, 2023).

Implementation of the 2013 Curriculum is a form of government appreciation in response to developments in education at the international level based on the results of PISA from time to time (Elitasari et al., 2023). Also, Puad & Ashton, (2023) state that developing the 2013 Curriculum can improve the quality of national education. The 2013 curriculum strives to provide the best student services, so think creatively, independently, and innovatively (Kharismayuni et al., 2021). The service is divided into two parts, namely content standards and assessment standards. Markhamah (2021) states that content standard services are carried out to classify relevant material for students according to the needs of critical thinking, while assessment legal services are carried out by analyzing and adapting international standard assessment models. In line with this, Setiawan et al., (2021) stated that Indonesia's assessment standards are directed at the higher-order thinking skills (HOTS) assessment model.

The HOTS-oriented learning outcome assessment model has been applied to the National Examination (Harta et al., 2020). In particular, the chemistry questions at the senior high school level contain several questions to measure higher-order thinking skills. Based on the test results, Widarti et al., (2020) 93.3% of the respondents yielded that the chemistry national examination questions have confirmed to the national education standard agency (BSNP), 60% of them believed that these questions have included HOTS aspects, prior to the use of the 2013 curriculum, Most of the respondents then argued that HOTS criteria cover stimulus, contextualization, problem-based questions, analogical thinking, and complexities to be assessed, 62.2% of them finally intended to add more aspects of HOTS in the chemistry national examination questions. It shows that the category of National Examination chemistry questions is still relatively low. Thus, the development of chemistry National Examination questions still needs to be seriously improved to develop students' high-level thinking skills.

Teachers play a significant role in developing students' higher-order thinking skills through assessment instruments (Surjanti et al., 2022). Wena (2020) emphasizes that the success of developing higher-order thinking skills is determined by the assessment instrument developed by the teacher. The assessment instrument is a measuring tool for evaluating student learning outcomes (Ndiung et al., 2024). In line with Arthadewi et al., (2024) suggests that the assessment instrument functions to determine the level of learning effectiveness of students at the cognitive, affective, and psychomotor levels. Teachers can drill students' cognitive learning evaluation with HOTS-oriented test items (Thowiyyah & Hidayatulloh, 2024). Dahlan et al. (2020) showed that only 12% of the teacher's questions were classified as HOTS, while 88% of the other questions were included in Lower Order Thinking Skills (LOTS). In line with the results of this study, it was found that teachers experienced difficulties formulating assessment instruments to measure HOTS. Therefore, teachers need intensive training in compiling assessment instruments, which is one of the competencies that professional teachers must possess.

The test or non-test procedure was applied to determine which assessment technique would be used (Masruri et al., 2024). The assessment instrument in the form of a test use HOTS-oriented items. The HOTS-oriented assessment instrument is used in the 2013 Curriculum to measure students' cognitive levels (Suswanti et al., 2024). HOTS items have characteristics that are different from questions in general. HOTS questions have the attributes of presenting a stimulus based on contextual problems that are interesting and novel (Shalikhah & Nugroho, 2023). Widana (2017) emphasizes that HOTS questions show real situations in everyday life where students are expected to be able to apply learning concepts in class to solve problems. One of the subject matter that uses a HOTS-oriented test instrument is chemistry.

Chemistry is the science that studies the composition, structure, properties, changes, and interactions of matter with everyday life (Gomollón-Bel, 2022). Chemistry is a compulsory subject for class 10th high school students for science majors. The characteristics of chemistry consist primarily of abstract concepts and forms of simplification of the actual situation, which are sequential and tiered. This characteristic makes chemistry material less attractive and challenging for students to learn (Karimah et al., 2021). Hemayanti et al., (2020) revealed that students were less enthusiastic in working on chemistry questions because the material was too difficult to understand. Jumaera et al., (2024) showed that stoichiometry material was a problematic indicator of chemical material, as indicated by student learning outcomes under the Minimum Completeness Criteria 75. In other words, stoichiometry is one of the basic materials that is considered difficult to understand in chemistry.

Stoichiometry is a chemical material that requires an understanding concepts and calculation skills (Díaz & Aizman, 2023). Simple basic concepts related to stoichiometry are described in the form of quantitative formulas and chemical reactions, which show that the products of the reaction are formed from some of the original substances (reactants) (Rasmawan, 2022). It shows that matter stoichiometry is very important for all aspects of chemistry (Jumaera et al., 2024). It is in line with Habib et al., (2025), who assert that stoichiometry is the core that underlies other chemical materials. The material characteristics of stoichiometry contain basic chemical concepts, laws, and calculation formulas, so it takes a lot of practice so that it is not easily lost in memory.

Based on these characteristics, stoichiometry contains values that can be applied contextually in daily activity, internalized, and added Islamic value to students (Andriyani & Apriantoro, 2023). The application of chemistry in everyday life that is integrated with Islam and cultural values provides many benefits (Zamhari, 2023). However, integrating Islam and chemistry subjects, especially stoichiometry material, has not been fully implemented in school test instruments. It is due to the low understanding of teachers in developing HOTS question instruments and the unavailability of guidelines for preparing questions to measure students' HOTS, especially in the concept of Islamic integration-interconnection (Suhady et al., 2020).

Then, this research developed a HOTS test instrument based on Islamic Integration on stoichiometric material. The HOTS question instrument is not only for testing the ability of chemical concepts but also enriches scientific and Islamic study literacy insights. In addition, the integration of chemistry with religious values aims to improve students' religious character by aligning chemistry with religion (Putri et al., 2023). It strengthens the mastery of values and habituates students' good behaviors in everyday life through learning chemistry (Sulastri et al.,

2018). As a very first research in this area, the development of HOTS questions is expected to become a reference and contribute to teachers as HOTS parameters in training higher-order thinking skills and forming students' religious values.

Method

This study employed a Research and Development (R&D) approach by applying the development of a formative research type, the Tessmer model (Stefaniak et al., 2025). The Tessmer model consists of 4 stages: Preliminary, Self-Evaluation, Prototyping, and Field Test. The product developed in this study is a HOTS test instrument on stoichiometric material based on Islamic integration-interconnection.

a. Preliminary. This stage was started with reviewing various reference sources related to research. Then, the procedure was continued by selecting places and test subjects by contacting the school principal and chemistry subject teacher to arrange research schedules and collaborative procedures with classroom teachers who would be used as research sites. The school was selected because it provides science subjects and religious studies. The sampling method used is saturated sampling.
b. Self-Evaluation. This stage was carried out with several analyzes, namely analysis of the curriculum, students, and material. The goal is to ensure that the product being developed is by the requirements. Furthermore, test instruments were designed and executed, including test grids, questions, answer sheets, assessment guidelines, and tools. This product design was a prototype that focuses on seven aspects: content feasibility, characteristics of HOTS, the role of HOTS, graphics, construction, language, and integrations. The integration aspect was detailed by the compatibility of the integration-interconnection of Islam and chemistry forms religious values.
c. Prototyping. At this stage, the designed product was evaluated. The evaluation phase was tested in 3 groups as revision material: Expert Review, One-to-One, and Small Group Phases.
    1. The expert review stage was a product validity test designed to be scrutinized and evaluated by one material expert lecturer, one media expert lecturer, one Islamic integration expert lecturer, and six reviewers (chemistry teachers).
    2. One-to-one stage involved three students with high, medium, and low abilities as a trial to answer and comment on the questions that have been worked on. The results of the one-to-one evaluation were combined with the results of the expert review evaluation to produce prototype II.
    3. The small group trial stage was carried out to try prototype II which consisted of 6 students with high, medium, and low abilities to test the practicality of the questions so that prototype III was produced (Afni et al., 2025).

d. Field Test. Prototype III, which had been revised and validated, was then re-tested on research trial subjects, namely 75 grade 10 students. The test results were analyzed for validity, reliability, level of difficulty, and discriminating power to measure students' high-level thinking skills.

Product validation was product quality data that contains suggestions and input from material expert lecturers, media expert lecturers, Islamic integration expert lecturers, and reviewer ratings (chemistry teachers). The data obtained was the conversion of letters into scores using a Likert scale ranging from 1 to 4, as seen in Table 1 (Astuti et al., 2023).

Table 1. Likert Scale for Validation of Material Expert Lecturer, Media Expert Lecturer, and Reviewers (High School Chemistry Teachers)

Score	Category Value
4	Very Good
3	Good
2	Poor
1	Very Poor

Interpretation of the validation values obtained using Table 1. Then the data obtained was calculated using the formula for the total average score of each aspect using equation (1) (Sugiyono, 2022).

$$\underline{X} = \frac{\Sigma x}{n} \qquad (1)$$

Information:

$\underline{X}$: average score

Σx : total score

n : number of appraisers

The average score of all aspects of product evaluation is converted into qualitative data based on ideal assessment criteria. The provisions can be seen in Table 2 (Sugiyono, 2020).

Table 2. Conversion of The Ideal Score into a Scale Value 4

Score Range	Category
$x \geq \underline{X} + SBi$	Very Good
$\underline{X} + SBi > x \geq \underline{X}$	Good
$\underline{X} > x > \underline{X} - SB$	Poor
$x < \underline{X} - SBi$	Very Poor

The percentage of ideal product quality for each aspect was calculated using equation (2) to determine the validity.

$$\%\ ideal\ from\ every\ aspect = \frac{the\ average\ score\ of\ each\ aspect}{the\ ideal\ highest\ score\ for\ each\ aspect} \times 100\% \qquad (2)$$

While the ideal product percentage for all aspects of the assessment uses equation (3).

$$\%\ ideal\ from\ all\ aspect = \frac{the\ average\ score\ of\ all\ aspect}{the\ ideal\ highest\ score\ for\ all\ aspect} \times 100\% \qquad (3)$$

The validity test of the item items was carried out using SPSS statistics 20. Whether each item was valid or not was determined by comparing the obtained r and r table with α = 0.05. In comparison, the interpretation of the magnitude of the correlation coefficient can be seen in Table 3 (Magdalena et al., 2021).

Table 3. Interpretation of Validity Value for Assessment Instrument

r_{xy}	Category
$0.80 \leq r_{xy} \leq 1.00$	Very High
$0.60 \leq r_{xy} < 0.80$	High
$0.40 \leq r_{xy} < 0.60$	Adequate
$0.20 \leq r_{xy} < 0.40$	Poor
$0.00 \leq r_{xy} < 0.20$	Very Poor

The reliability test of the items was carried out using SPSS statistics 20. Whether each item was reliable or not was determined by comparing the SPSS output with Croncbach's Alpha value > 0.60. In contrast, the interpretation of the reliability value can be seen in Table 4.

Table 4. Interpretation of Reliability Value for Assessment Instrument

r_{11}	Category
$0.80 \leq r_{11} \leq 1.00$	Very High
$0.60 \leq r_{11} < 0.80$	High
$0.40 \leq r_{11} < 0.60$	Adequate
$0.20 \leq r_{11} < 0.40$	Poor
$0.00 \leq r_{11} < 0.20$	Very Poor

The difficulty level of a question is an opportunity to answer a question correctly at a certain ability level which can be expressed by an index (Sugiyono, 2022). The difficulty level test was carried out using Anates V4. The interpretation of the difficulty level of the table 5 (Setiyawan & Wijayanti, 2020).

Table 5. Interpretation of Item Difficulty Level

Difficulty Index	Interpretation
$P < 0.3$	Difficult
$0.3 < P \leq 0.7$	Medium
$P > 0.7$	Easy

Discriminating power is the ability of questions to distinguish answers between students who are capable and students who are less able to work on questions (Magdalena et al., 2020). Discriminatory test analysis was carried out using Anates V4. The interpretation of discriminating power can be seen in Table 6.

Table 6. Interpretation of the Discriminating Power of the Items

Discriminating Power Index	Interpretation
DP ≤ 0.00	Very Poor
0.00 ≤ DP < 0.20	Poor
0.20 ≤ DP < 0.40	Adequate
0.40 ≤ DP < 0.70	Good
0.70 ≤ DP < 1.00	Very Good

Results and Discussion

Preliminary

This stage aimed to discover the problems in the learning and assessment process in measuring students' cognitive abilities. Novatania & Kamaludin (2021) The initial step was carried out by reviewing several references related to developing the HOTS, the availability of HOTS test instruments was still relatively small. Stoichiometry is one of the chemistry materials, but the availability of HOTS questions is still low. The combination of stoichiometry and HOTS questions is important because it requires a deep understanding of concepts, critical thinking, and problem-solving abilities. Stoichiometric material is material whose context is close to everyday life and is integrated with Islamic values. The discussion of the sea being split due to differences in salt composition in the Qur'an is chemically very relevant to be explained with stoichiometry and discussed with HOTS. However, many learning and assessment processes have not applied stimulus, context, and thinking at the C4-C6 level (Syahidah et al., 2024).

The next stage was a school survey to determine the location and subject of the research trials. The test site for this study was MAN 2 Yogyakarta. At the same time, the test subjects in this study were 10th grade students. Based on the results of interviews and school surveys, Students' perspectives were obtained by one to one and small group stages. teachers still use questions in the form of multiple choice and descriptions with the Low Order Thinking Skills (LOWS) level in measuring students' thinking abilities. The teacher found difficulties obtaining reference questions and preparing HOTS questions.

Self-Evaluation

This stage aimed to design a HOTS test instrument based on Islamic integration-interconnection based on the results of the preliminary stage. The test instrument consisted of test grids, multiple-choice test questions, and test answer keys. This stage consisted of 4 activities: curriculum analysis, student analysis, material analysis, and design.

Curriculum analysis was carried out to check the suitability of the competencies, learning objectives, scope of material, and strategies to be developed in the HOTS test instrument based on Islamic integration-interconnection. This analysis refers to the 2013 curriculum content standards, including Core Competencies and Basic Competencies, as seen in Table 7.

Table 7. Core Competencies and Basic Competencies in the 2013 Curriculum

Competency standards	Fill
Core Competency (CC)	3. Understanding, applying, and analyzing knowledge based on curiosity about science, technology, art, culture, and humanities with human, national, state, and civilization insights related to the causes of phenomena and events and applying procedural knowledge in specific fields of study according to their talents and interests to solve the problem
Basic Competency (BC)	3.10 Apply the fundamental laws of chemistry, the concept of relative molecular mass, chemical equations, and the concept of moles and levels of matter to complete chemical calculations

Based on the two competencies mentioned in Table 7, they are translated into indicators adjusted to the criteria for HOTS questions in the cognitive domain of Bloom's Taxonomy, namely C4, C5, and C6. These indicators were used as a reference for the initial design of the HOTS test instrument based on Islamic integration-interconnection.

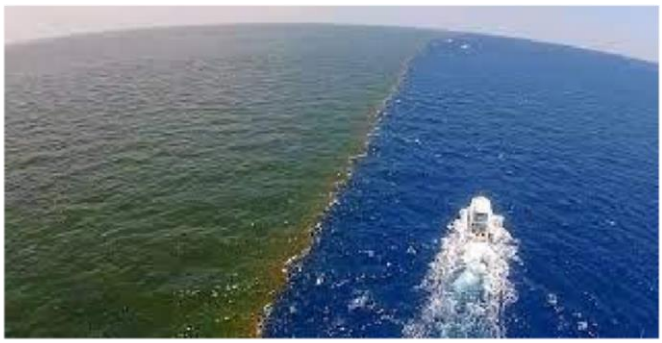
Material analysis was conducted to identify the chemical material students would study. Based on the results of the curriculum analysis, the chemistry material complies with the 2013 curriculum content standards, namely stoichiometry material. Stoichiometry is one of the learning materials that requires students to understand and analyze phenomena and events in everyday life that are integrated with Islam and solve problems. It is used as the basis for research to develop a product as a HOTS test instrument on stoichiometric material based on Islamic integration-interconnection.

Students at 10^{th} grade was test subjects because they had learned stoichiometry. The test subjects were selected in three science classes, namely classes 10^{th} C, 10^{th} D, and 10^{th} E. The number of test subjects from the three classes was 75 students. Based on the results of observations and interviews, the thinking abilities of 10^{th} grade students varied, namely high, medium, and low skills. This ability was reviewed from the evaluation of the Mid-Semester Assessment, Final Semester Assessment, or a separate assessment from the teacher. It allowed for factors of interest that each student has in chemistry lessons.

The design stage involved designing the HOTS test instrument as prototype 1, including test grids, multiple choice test questions, and answer keys. The test grid guides in compiling relevant and meaningful HOTS items. Test grids helped select subject matter related to Basic Competency, formulate question indicators, and determine the cognitive level in questions (Widana, 2017). Design test questions were developed in the form of multiple choice. According to Liu et al., (2024) the multiple choice test instrument was often used because it was more objective and more straightforward to process data. The HOTS test instrument

developed consists of 20 questions. Examples of HOTS questions and answer keys, can be seen in Figure 1.

Allah says: "*He allowed two seas to flow which then met. Between the two, there is a limit that is not exceeded each*" (QS. Ar-Rahman: 19-20). One of the natural phenomena that become the power of Allah is the phenomenon of the meeting of the Mediterranean and Atlantic seas which do not dissolve each other in the Strait of Gibraltar, as shown below:



Source. https://s.id/1K9kl

Based on research, the Mediterranean Sea has a higher water salinity than the Atlantic Ocean. To find out the salt content of NaCl in the Mediterranean Sea, a researcher reacted seawater samples with an excess of $AgNO_3$ solution, resulting in the formation of 28.7 grams of AgCl precipitate as shown in the following reaction equation:

$$NaCl(aq) + AgNO_3(aq) \rightarrow AgCl(s) + NaNO_3(aq)$$

If it is assumed that seawater only contains NaCl salt, then the mass of NaCl dissolved in the sample is... (Mr AgCl = 143.5 grams/moles; NaCl = 58.5 grams/moles)

A. 0.2 grams
B. 28.7 grams
C. 5.0 grams
D. 11.7 grams
E. 28.7 grams

The answer to the question is : D

$$\text{Mole AgCl} = \frac{\text{Mass}}{\text{Molecular mass (Mr)}} = \frac{28{,}7 \text{ grams}}{143{,}5 \text{ grams/mole}} = 0.2 \text{ mole}$$

$$\text{Mole NaCl} = \frac{\text{coefficient NaCl}}{\text{coefficient AgCl}} \times \text{mole AgCl}$$

$$= \frac{1}{1} \times \text{mole AgCl}$$

$$= \frac{1}{1} \times 0.2 \text{ mole}$$

$$= 0.2 \text{ mole}$$

$$\text{Mass of NaCl} = \text{mole} \times \text{Mr}$$

$$= \text{mole NaCl} \times \text{Mr NaCl}$$

$$= 0.2 \text{ moles} \times 58.5 \text{ grams/mole}$$

$$= 11.7 \text{ grams}$$

So, the mass of NaCl dissolved in the sample is 11.7 grams

Figure 1. Examples of HOTS Question Items on Islamic Integration-Integration-Based Stoichiometry Material

The characteristics of HOTS questions based on Islamic integration-interconnection in Figure 1 provided a stimulus in the form of images and contextual discourse integrated with Islam. The questions are at cognitive level C4 (analysis) by demanding students' abilities to analyze or analyze information critically. Furthermore, multi-concepts are applied to work on the questions because students must be able to master chemical calculations without a limiting reaction to determine the mass of NaCl.

Prototyping

This stage aimed to evaluate the results of prototype I in 3 groups, namely Expert Review, One-to-One, and Small Group. Expert Review is usually called the prototype I validity test, which experts or experts will examine, assess, and evaluate. The validators comprised one material expert lecturer, one media expert lecturer, one Islamic integration expert lecturer, and six reviewers (chemistry teachers).

Based on the quality assessment of the material expert lecturers in Table 8, the product gets a percentage of 98.96% and is categorized as Very Good (VG). This performance confirms that the product represents competency-appropriate material in the 2013 curriculum syllabus and that a stimulus is contextual at cognitive levels C4-C6.

Table 8. Quality Assessment Data for HOTS Assessment Instruments on Stoichiometry Material Based on Islamic Interconnection by Material Experts

Assessment Aspects	Σ Score	Σ Ideal Maximum Score	Ideal Percentage (%)	Category
Content Eligibility	235	240	97.92%	very good
HOTS characteristics	160	160	100.00%	very good
The Role of HOTS Questions	80	80	100.00%	very good
Total	**475**	**480**	**98.96%**	**very good**

According to the media expert's assessment in Table 9, 99.03% was obtained and considered very good category. This performance shows that the presentation of the questions is relatively straightforward and does not give directions for the correct answer. In addition, the language used is good and right, so students easily understand it.

Tabel 9. Quality Assessment Data for HOTS Assessment Instruments on Stoichiometric Materials Based on Islamic Interconnection by Media Experts

Assessment Aspects	Σ Score	Σ Ideal Maximum Score	Ideal Percentage (%)	Category
Graphics	153	160	95.63%	very good
Construction	240	240	100.00%	very good
Language	320	320	100.00%	very good
Total	**713**	**720**	**99.03%**	**very good**

Furthermore, according to the assessment of the Islamic integration expert in Table 10, 73.13% was obtained as good category. The superior performance formulated shows that the product includes the truth of the concept of verses of the Qur'an or Hadits or relevant Islamic studies supporting the topic of stoichiometry in chemistry. The dietary guidelines in Islam can be detailed by the hydrochloric acid (HCl) reaction in the stomach, characterized by its stoichiometric properties and functions which cleared in stoichiometry topic. The material in the problem reflects integration with students' understanding. In addition, presenting the

material in the items can instill Islamic values in students, such as making awareness of Allah SWT greatness and power, increasing students' righteousness, and adding insight into chemical concepts from an Islamic point of view.

Tabel 10. Quality Assessment Data for HOTS Assessment Instruments on Stoichiometric Materials Based on Islamic Interconnection by Islamic Integration Experts

Assessment Aspects	Σ Score	Σ Ideal Maximum Score	Ideal Percentage (%)	Category
Integration-Interconnection	117	160	73.13%	good
Total	**117**	**160**	**73.13%**	**good**

The assessment was carried out by six chemistry teachers in Table 11, which obtained a percentage of 93.26% and was considered very good. This performance confirms that the product is accepted and can be tested on more prominent students.

Tabel 11. Quality Assessment Data for HOTS Assessment Instruments on Stoichiometric Materials Based on Islamic Interconnection by Reviewers (High School Chemistry Teachers)

Assessment Aspects	Σ Score	Σ Ideal Maximum Score	Ideal Percentage (%)	Category
Content Eligibility	1308	1440	90.83%	very good
HOTS characteristics	905	960	94.27%	very good
The Role of HOTS Questions	430	480	89.58%	very good
Graphics	859	960	89.48%	very good
Construction	1353	1440	93.96%	very good
Language	1857	1920	96.72%	very good
Integration-Interconnection	898	960	93.54%	very good
Total	**7610**	**8160**	**93.26%**	**very good**

At this stage, one-to-one interviews were carried out simultaneously with the expert review stage. The prototype 1 was tested on three students with different abilities: high, medium, and low. The three students worked on the questions and provided comments or suggestions on the HOTS test instrument being developed. The goal is to see the readability of the question. Students' criticisms and suggestions at the one-to-one stage can be seen in Table 12.

Tabel 12. Student Criticism and Suggestions at The One-To-One Stage

Group	Respondents	Comments (Criticism and Suggestions)
High	Learners 1	The difficulty level of the questions is high, especially in the equations for the reactions of substances that lack detailed explanations. Therefore, it is necessary to include information on the name of the essence and the formula of the compound in the problem.
Medium	Learners 2	The narrative of Islamic studies presented is too long, so it takes time to understand the essence of what is being asked in the problem.
Low	Learners 3	The questions look foreign because they are related to Islamic values, so it isn't easy to understand the essence of the questions.

Based on the results of the one-to-one trial, the students' responses felt it was difficult to answer the questions because they were rare and foreign. It took a lot of concepts to solve problems, so sound reasoning is needed in solving problems. The results of the one-to-one evaluation were combined with the results of the expert review evaluation to produce prototype II.

Small group, at this stage, the results of prototype II were tested on six students with high, medium, and low abilities. Based on the test results and student comments, the product was revised so that Prototype III was produced. The results of this stage were expected to make a test instrument that can measure students' higher-order thinking skills. Students' criticisms and suggestions at the one-to-one stage can be seen in Table 13.

Tabel 13. Student Criticism and Suggestions at The Small Group Stage

Group	Respondents	Comments (Criticism and Suggestions)
High	Learners 1	Statements and questions are so long but still understandable
	Learners 2	Some of the questions contain narratives of Islamic studies that are too long and unrelated to chemistry, so they need to be corrected so as not to cause double meanings.
Medium	Learners 3	The narrative of Islamic studies presented is too long, so it needs to be read over and over to make it easy to understand
	Learners 4	Questions are too long and complicated, so it takes extra time to find the core of the problem.
Low	Learners 5	Islamic narration in questions is very helpful in adding insight. However, the narration is too long, so finding the problem's core problem is complex.
	Learners 6	The difficulty level of the questions is not equivalent to what is learned.

Field Test

The data obtained from this trial stage analyzed the validity, reliability, difficulty level, and differentiating power of the HOTS test instrument on stoichiometric material based on Islamic integration-interconnection.

The validity test of the items was analyzed using SPSS statistics 20. The analysis was done by comparing the r count and r table values at a significance level of 0.05. Decision-making is based if the r count exceeds the r table, and the item is declared valid. R table for 75 respondents (students) was 0.227. The output results of the SPSS test for the validity of the items can be seen in Table 14.

Tabel 14. Validity Test of HOTS Questions

Category	Question Number	Number of Questions	Percentage (%)
Valid	1,2,3,4,5,6,8,9,10,11,12,13,15,16,18,20	16	80%
Invalid	7,14,17,19	4	20%

Based on the results of the analysis, 16 items consisting of numbers 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 18, and 20 are declared valid because they have a value of r count > 0.227. The other four items, consisting of numbers 7, 14, 17, and 19, have r count < 0.227, so they are declared invalid. So, from these results, the best 16 items were taken as the final product of the HOTS test instrument being developed. If you look at the interpretation of the validity value in Table 3, 80% of the HOTS test instruments are in a sufficiently valid category because the r count value is > 0.40. The sufficient validity of the items indicates that the HOTS test instrument is suitable for the teacher's assessment instrument measuring students' abilities (Ningsih & Kamaludin, 2023).

The reliability test of the items was analyzed using SPSS statistics 20. The reliability test was conducted only on valid questions, namely 16 items. Cronbach's Alpha value obtained is 0.843. The test instrument is reliable because the Cronbach's Alpha value is more than 0.6. The SPSS output results of the reliability test can be seen in Table 15. Suppose you look at the interpretation of the reliability value in Table 4. The HOTS test instrument is highly reliable because the Cronbach's Alpha value is more than 0.80. The very high reliability of the items indicates that the HOTS test items are sufficient and can be used in further research.

Tabel 15. The Result of Reliability Test

Reliability Statistics			
Cronbach's Alpha	**N of Items**	**Information**	**Category**
0.843	16	reliable	Very High

Test the difficulty level of the items analyzed using Anates V4. Based on the analysis conducted on 20 multiple-choice items, it was found that 6 items (30%) were in the easy category, 10 items (50%) were in the medium category, and 4 items (20%) were in the difficult category. The difficulty level of the questions as a whole is included in the medium category because the number of items (percentage of questions) is higher than the number of items (percentage of questions) easy and difficult. The distribution of the difficulty level of the questions can be seen in Table 16.

Tabel 16. Distribution of Item Difficulty Levels

Difficulty Index	Category	Question Number	Number of Questions	Percentage (%)
P < 0.3	Easy	2,8,12,15,16,18	6	30%
0.3 < P < 0.7	Medium	1,4,5,6,7,9,10,11,14,17	10	50%
P > 0.7	Difficult	3,13,19,20	4	20%

The item discriminating power test was analyzed using Anates V4. Based on the analysis conducted on 20 multiple choice items, it is known that 3 items (15%) are in the very poor category, 2 items (10%) are in the poor category, 6 items (30%) are in the adequate category, 3 items (15%) are in the good category, and 6 items (30%) are in the very good category. The discriminating power of the items is included in the good category because 9 questions have a discriminating power of questions > 0.40. Thus, these 9 questions can distinguish students with high and low abilities. The distribution of the power of discriminating questions can be seen in Table 17.

Tabel 17. Power Distribution of Differentiating Items

Discriminating Power Index	Category	Question Number	Number of Questions	Percentage (%)
DP ≤ 0.00	Very Poor	7,14,17	3	15%
0.00 ≤ DP < 0.20	Poor	8,19	2	10%
0.20 ≤ DP < 0.40	Adequate	2,6,13,15,16,20	6	30%
0.40 ≤ DP < 0.70	Good	3,12,18	3	15%
0.70 ≤ DP < 1.00	Very good	1,4,5,9,10,11	6	30%

Conclusion

Research on developing an Islamic-integrated HOTS stoichiometry test using the Tessmer R&D model was successfully done. Product quality was validated by a material expert, a media expert, and an Islamic Integration expert with a score of 98.96%, 99.03%, and 73.13%, respectively. Then, the product was also assessed by six reviewers (chemistry teachers) with a score of 93.26% as a very good category. The product was tested on 75 students in 10th grade with 16 valid items and four invalid items with very high reliability, namely 0.843. At the difficulty level, the product contains six easy, ten medium, and four difficult questions, with more than 75% of the questions able to distinguish students with high and low abilities. Therefore, the developed instrument is suitable for measuring HOTS and Islamic integration on the material of stoichiometry.

Credit Authorship Contribution Statement

Didan Sopian: Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing – original draft, Writing – review & editing. **Muhammad Zamhari**: Conceptualization, Methodology, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration.

References

Afni, N., Ulya, H., & Rahmawan, S. (2025). Development of Pictorial-Based Five-Tier Multiple Choice Diagnostic Test to Identify Misconceptions of Electrochemical Cell Material. Paedogoria: Jurnal Kajian, Penelitian Dan Pengembangan Kependidikan, 6356(2011), 145–154. https://doi.org/10.31764

Amali, L. N., Anggani Linggar Bharati, D., & Rozi, F. (2022). The Implementation of High Order Thinking Skills (HOTS) Assessment to Evaluate the Students' Reading Comprehension Achievement. *English Education Journal*, *12*(1), 10–18. https://doi.org/10.15294/eej.v12i1.52571

Andriyani, N., & Apriantoro, M. S. (2023). Enhancing Students Cognitive Abilities in Integrated Islamic High Schools: A Need-Based Analysis for Developing Chemistry Modules Infused with Islamic Values. International Journal of Humanities and Social Science Studies, 1(1), 1–10. https://doi.org/https://ejournal.grninstitute.com/index.php/Ethica/article/view/1

Arthadewi, N. M., Bayu, G. W., & Rati, N. W. (2024). HOTS-Based Science Learning Outcomes Assessment Instrument for Measuring Elementary School Learning Outcomes. Sinergi International Journal of Islamic Studies, 2(1), 24–37. https://doi.org/https://doi.org/10.23887/mi.v29i1.36521 HOTS-Based

Astuti, R. T., Dewi, M., & Pratiwi, R. Y. (2023). Online Test Instrument to Measure Students' Critical Thinking Ability. DWIJA CENDEKIA: Jurnal Riset Pedagogik, 7(1), 276. https://doi.org/10.20961/jdc.v7i1.71328

Dahlan, D., Permana, L., & Oktariani, M. (2020). Teacher Competence and Difficultes in Contructing HOTS Instruments in Economics Subject. *Jurnal Cakrawala Pendidikan*, *39*(1), 111–119. https://doi.org/https://doi.org/10.21831/cp.v39i1.28869

Díaz, B., & Aizman, A. (2023). Design and impact of a stoichiometry voluntary online course for entering first-year STEM college students. Chemistry Education Research and Practice. https://doi.org/https://doi.org/10.1039/D3RP00179B

Dwijayanti, N. (2021). Pembelajaran Berbasis HOTS sebagai Bekal Generasi Abad 21 di Masa Pandemi. *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, *9*(1), 333–336. https://doi.org/https://doi.org/10.20961/jkc.v9i1.53837

Elitasari, H. T., Yeni Rakhmawati, Irma Yulianti Budi Safitri, & Muhammad Asip. (2023). Alteration: 2013 Curriculum Becomes Merdeka Curriculum in Elementary Schools. Progres Pendidikan, 4(2), 122–130. https://doi.org/10.29303/prospek.v4i2.350

Gomollón-Bel, F. (2022). IUPAC Top Ten Emerging Technologies in Chemistry 2022. Chemistry International, 44(4), 4–13. https://doi.org/10.1515/ci-2022-0402

Habib, M., Shiddiqi, A., Ikhsan, J., Purwaningsih, M. D., & Ayubi, M. (2025). Stoichiometry in Chemistry as a Reflection of The Values of The Qur'an. SERUMPUN : Journal of Education, Politic, and Social Humaniora, 3(1), 7–9. https://doi.org/10.61590/srp.v3i1.160

Harta, J., Rasuh, N. T., & Seriang, A. (2020). Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students. Journal of Science Learning, 3(3), 143–148. https://doi.org/10.17509/jsl.v3i3.22387

Hemayanti, K. L., Muderawan, I. W., & Selamat, I. N. (2020). Analisis Minat Belajar Siswa Kelas XI MIA Pada Mata Pelajaran Kimia. *Jurnal Pendidikan Kimia Indonesia*, *4*(1), 20. https://doi.org/https://doi.org/10.23887/jpk.v4i1.24060

Jumaera, S., Temitope, O., & Rukondo, N. (2024). Optimizing Student Activities and Learning Outcomes through Problem Solving Models in Stoichiometry Material. Journal of Chemical Learning Innovation, 1(2), 39–44. https://doi.org/10.37251/jocli.v1i2.1147

Karimah, U., Retnawati, H., Hadiana, D., Pujiastuti, P., & Yusron, E. (2021). The characteristics of chemistry test items on nationally-standardized school examination in Yogyakarta City. REID (Research and Evaluation in Education), 7(1), 1–12. https://doi.org/10.21831/reid.v7i1.31297

Kharismayuni, E., Feronika, T., & Yunita, L. (2021). Implication of Thinking Maps Assisted Inquiry Model for Higher Order Thinking Skills (HOTS) on Chemistry. *Journal of Physics: Conference Series*, *1836*(1), 1–12.

Kurniawan, E. S., Mundilarto, & Istiyono, E. (2024). Improving student higher order thinking skills using Synectic-HOTS-oriented learning model. International Journal of Evaluation and Research in Education , 13(2), 1132–1140. https://doi.org/10.11591/ijere.v13i2.25002

Liu, Q., Wald, N., Daskon, C., & Harland, T. (2024). Multiple-choice questions (MCQs) for higher-order cognition: Perspectives of university teachers. Innovations in Education and Teaching International, 61(4), 802–814. https://doi.org/10.1080/14703297.2023.2222715

Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya Evaluasi dalam Pembelajaran dan Akibat Memanipulasinya. *Bintang: Jurnal Pendidikan Sains*, *2*(2), 244–257. https://doi.org/https://doi.org/10.36088/bintang.v2i2.986

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Bintang: Jurnal Pendidikan Sains*, *3*(2), 198–21. https://doi.org/https://doi.org/10.36088/bintang.v3i2.1291

Markhamah, N. (2021). Pengembangan Soal Berbasis HOTS (Higher Order Thinking Skills) Pada Kurikulum 2013. *Nusantara: Jurnal Pendidikan Indonesia*, *1*(2), 385–418. https://doi.org/https://doi.org/10.14421/njpi.2021.v1i2-8

Masruri, M. S. I., Sari, M. K., & Azwary, K. (2024). Cognitive Assessment Techniques on Student Learning Outcomes in SKI Subjects at MA Darul Ukhuwwah Putri 2 Singosari Malang. 9(2), 201–210.

Ndiung, S., Atika, M. T., Jediut, M., & Helmon, A. (2024). Higher Order Thinking Skills in Mathematics at Elementary School. Jurnal Cakrawala Pendas, 10(3), 523–534. https://doi.org/10.31949/jcp.v10i3.9658

Novatania, D. W., & Kamaludin, A. (2021). Development of High Order Thinking Skills (HOTS) Test Instruments on Thermochemistry Topics. *JTK (Jurnal Tadris Kimiya)*, *6*(2), 174–184. https://doi.org/https://doi.org/10.15575/jtk.v6i2.12746

OECD. (2023). Programme for international student assessment (PISA) results from PISA 2018. In *OECD*.

Puad, L. M. A. Z., & Ashton, K. (2023). A Critical Analysis of Indonesia's 2013 National Curriculum: Tensions Between Global and Local Concerns. *Curriculum Journal*, *34*(3), 521–535. https://doi.org/10.1002/curj.194

Putri, R., Winarni, S., & Erlidawati, E. (2023). Development of chemistry questions integrated with religious values on acid-base topic. *Thabiea : Journal of Natural Science Teaching*, *6*(2), 213. https://doi.org/10.21043/thabiea.v6i2.20408

Rasmawan, R. (2022). Stoichiometry E-Book Based on Creative Problem Solving (CPS) to Solve Conceptual Problems. Jurnal Pendidikan Kimia Indonesia, 6(1), 70–78. https://doi.org/10.23887/jpk.v6i1.43441

Setiawan, J., Sudrajat, A., Aman, & Kumalasari, D. (2021). Development of higher order thinking skill assessment instruments in learning Indonesian history. International Journal of Evaluation and Research in Education, 10(2), 545–552. https://doi.org/10.11591/ijere.v10i2.20796

Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis Kualitas Instrumen untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring di Masa Pandemi. *Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistik*, *1*(2), 130–139. https://doi.org/https://doi.org/10.46306/lb.v1i2.26

Shalikhah, N. D., & Nugroho, I. (2023). Implementation of Higher-Order Thinking Skills in Elementary School Using Learning Model, Media, and Assessment. AL-ISHLAH: Jurnal Pendidikan, 15(3), 3978–3990. https://doi.org/10.35445/alishlah.v15i3.3091

Stefaniak, J. E., Yang, F., Gilstrap, S., & Yang, L. (2025). An Exploration of Instructional Design Tasks Performed by Designers in Higher Education. TechTrends. https://doi.org/https://doi.org/10.1007/s11528-025-01041-9

Sugiyono. (2020). *Metode Penelitian Kuantitatif*. Alfabeta.

Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Suhady, W., Roza, Y., & Maimunah, M. (2020). Pengembangan Soal untuk Mengukur Higher Order Thinking Skill (HOTS) Siswa. *Jurnal Gantang*, *5*(2), 143–150. https://doi.org/https://doi.org/10.31629/jg.v5i2.2518

Sulastri, S., Rusman, R., & Arifa, A. (2018). Pengembangan Soal-Soal Kimia Bermuatan Nilai-Nilai untuk Memperkokoh Karakter Siswa SMA. *JTK (Jurnal Tadris Kimiya)*, *3*(2), 171–181. https://doi.org/10.15575/jtk.v3i2.3512

Surjanti, J., Prakoso, A. F., Kurniawan, R. Y., Sakti, N. C., & Nurlaili, E. I. (2022). Development of High Order Thinking Skills in Indonesian Teachers. Obrazovanie i Nauka, 24(3), 104–125. https://doi.org/10.17853/1994-5639-2022-3-104-125

Suswanti, N., Solikhin, F., & Ginting, S. M. (2024). Development of Cognitive Assesment Instrument on Buffer Solution Material Based on HOTS (Higher Order Thinking Skills). Journal of Chemistry Education Research, 8(2), 97–103. https://doi.org/https://doi.org/10.26740/jcer.v8n2.p97-103

Syahidah, N., Kusasi, M., Analita, R. N., & Sholahuddin, A. (2024). Pengembangan E-Instrumen Tes Materi Stoikiometri Menggunakan Blooket Game untuk Mengukur Kemampuan Berpikir Kritis Pada Siswa SMA. *Journal of Chemistry And Education*, *7*(3), 161–174. https://doi.org/https://doi.org/10.20527/jcae.v7i3.2878

Syahri Ramdhani, S., Susanti, R., & Meilinda. (2024). Cognitive Level of Program for International Student Assessment (PISA) Questions Based on the Revised Bloom's Taxonomy. European Journal of Education and Pedagogy, 5(2), 104–112. https://doi.org/10.24018/ejedu.2024.5.2.785

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, *11*(3). https://doi.org/10.3390/jintelligence11030054

Thowiyyah, H., & Hidayatulloh, M. K. Y. (2024). Development of a HOTS-based Multiple Choice Question Test Instrument on the Basic Competency of Understanding Marriage Provisions According to Legislation. APPLICATION: Applied Science in Learning Research, 3(3), 13–21. https://doi.org/10.32764/application.v3i3.4549

Wena, I. M. (2020). Pembelajaran Berorientasi HOTS (Higher Order Thinking Skills) di Era Revolusi Industri 4.0 untuk Mewujudkan Generasi Indonesia Emas 2045. Mahasaraswati Seminar Nasional Pendidikan Matematika (MAHASENDIKA), 15–25.

Widana, I. W. (2017a). Higher Order Thinking Skills Assessment (HOTS). *JISAE: Journal of Indonesian Student Assesment and Evaluation*, *3*(1), 32–44.

Widana, I. W. (2017b). *Modul Penyusunan Soal Higher Order Thinking Skills (HOTS)*. Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan.

Widarti, H. R., Amalia, R., Sigit, D., Habbidin, & Sulistina, O. (2020). Higher order thinking skills (HOTS) aspects portrayed in the chemistry national examination questions: Voices of teachers. Innovative Research in Science and Mathematics Education in The Disruptive Era. https://doi.org/https://doi.org/10.1063/5.0043362

Zamhari, M. (2023). Culinary in Islamic-Javanese Cultural Heritage from a Chemistry Perspective. *Journal of Tropical Chemistry Research and Education*, *5*(2). https://doi.org/https://doi.org/10.14421/jtcre.2023.52-05