Journal of Educational Chemistry (6(1), 2024, 37-48)
Website: https://journal.walisongo.ac.id/index.php/jec/index
ISSN 2715-3029 (p) 2685-4880 (e)
DOI: 10.21580/jec.2024.6.1.12438

Development of a Three-Tier Chemical Equilibrium Multiple-Choice Test to Assess Higher-Order Thinking Skills Based on Brookhart's Categories

Neneng Fauziyah[1*], Tonih Feronika[2], Luki Yunita[3]
[1,2,3)]Department of Chemistry Education, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia
*E-mail Corresponding Author: neneng.fauziyah15@mhs.uinjkt.ac.id

Abstract

This study aimed to develop and validate the Three-Tier Chemical Equilibrium Multiple-Choice Test–HOTS (TTCEMCT–HOTS) instrument. The method employed in this study was Research and Development (R&D) based on Brookhart's (2010) indicators of higher-order thinking skills. The development procedure began with a preliminary study, followed by product draft planning and product development. These stages were succeeded by an empirical analysis of the TTCEMCT–HOTS to determine the test instrument's validity, reliability, difficulty level, and discriminating power. The TTCEMCT–HOTS instrument was implemented with Class XI students of SMAN 4 in South Tangerang City. The results of the study were as follows: (1) the developed TTCEMCT–HOTS instrument met the criteria of 32 valid questions, with a reliability coefficient of 0.82 in the "very high" category, a "moderate" difficulty level, and "good" discriminating power; (2) 46.42% of students could solve HOTS questions, while 53.58% could not; (3) the TTCEMCT–HOTS instrument received a positive response from 80.6% of participants, placing it in the "good" category. This study is expected to provide an alternative method for measuring students' higher-order thinking skills using a three-tier test instrument integrated with HOTS indicators.

Keywords: chemical equilibrium; higher-order thinking skills; three-tier multiple choice

Abstrak

Penelitian ini bertujuan untuk mengembangkan dan memvalidasi Three Tier Chemical Equilibrium Multiple Choice Test–HOTS (TTCEMCT–HOTS). Metode yang digunakan dalam penelitian ini adalah Penelitian dan Pengembangan (R&D) berdasarkan indikator berpikir tingkat tinggi menurut Brookhart (2010). Prosedur pengembangan dimulai dengan studi pendahuluan, perencanaan draft produk, hingga pengembangan produk, dilanjutkan dengan analisis empiris TTCEMCT–HOTS untuk menentukan menentukan instrumen yang dikembangkan berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Implementasi instrumen TTCEMCT–HOTS dilakukan pada siswa kelas XI SMAN Kota Tangerang Selatan. Hasil penelitian: (1) instrumen TTCEMCT–HOTS yang dikembangkan memenuhi kriteria valid sebanyak 32 butir soal, koefisien reliabilitas sebesar 0,82 dengan kategori sangat tinggi, tingkat kesukaran butir soal sedang, dan memiliki daya pembeda baik; (2) teridentifikasi 46,42% siswa mampu menyelesaikan soal HOTS, sedangkan 53,58% siswa lainnya belum mampu; (3) instrumen TTCEMCT–HOTS mendapat respon positif sebesar 80,6% dengan kategori baik. Penelitian ini diharapkan dapat memberikan alternatif dalam mengukur keterampilan berpikir tingkat tinggi dengan menggunakan instrumen three-tier test yang terintegrasi indikator HOTS.

Keywords: kesetimbangan kimia; keterampilan berpikir tingkat tinggi; pilihan gada tingkat tiga

Introduction

In the current era of the Industrial Revolution 4.0, where job competition intensifies, students are expected to develop 21st-century skills. Hence, science education aims to equip students with higher-order thinking skills (HOTS) in line with the demands of the 21st century (Juliarti et al., 2019). However, an evaluation conducted by PISA in 2015 revealed that the higher-order thinking skills of Indonesian students were still low (Ramadhan et al., 2018). This situation was further corroborated by the PISA evaluation in 2018, which ranked Indonesian students' science achievement 71st out of 79 participating countries (Agustina et al., 2021). One of the factors contributing to the low level of higher-order thinking skills is the lack of training for Indonesian students in solving contextual problems that require reasoning, argumentation, and creativity. This issue arises from insufficient test questions specifically designed to measure higher-order thinking skills (Hikmah & Amin, 2019).

The test questions used in high school chemistry subjects tend to rely on one-tier multiple-choice tests that measure cognitive levels of remembering (C1), understanding (C2), and applying (C3) (Netri et al., 2018). This form of testing is considered less effective in measuring higher-order thinking skills, as there is a 20% chance that students can answer correctly just by guessing (Dewati et al., 2016). This circumstance highlights the lack of instruments that accurately measure students' higher-order thinking skills (Shidiq et al., 2014).

Research and development of a two-tier multiple-choice test instrument by Azmi et al. (2021) succeeded in revealing profiles of students' higher-order thinking skills (HOTS) concerning reaction rates. However, it has not been implemented adequately to measure HOTS accurately, as it is still possible for students to answer correctly by luck or coincidence (Sutiana et al., 2018). Hidayatullah et al. (2022) argue that measuring higher-order thinking skills will be more accurately identified using a more modern form of multiple-choice test. In line with this, Azmi et al. (2021) recommend using an instrument that includes a level of confidence to identify students' HOTS profiles in more detail. The three-tier multiple-choice test instrument incorporates an answer tier, a reason tier, and a student's confidence level in their answers and reasons, allowing for a more accurate assessment of whether students truly possess higher-order thinking skills (Sutiana et al., 2018).

In line with efforts to improve the quality of education in Indonesia, the Ministry of Education and Culture issued a policy in 2018 that learning and assessment in schools should be oriented towards higher-order thinking skills (HOTS) (GTK Pendidikan Dasar, 2021, p. 65). This policy is applied to chemistry subjects, where HOTS is integrated into classroom learning and assessment to foster students' thinking skills and creativity (Mushlihuddin et al., 2021). In this regard, chemical equilibrium is one of the topics in chemistry that includes simple concepts used to build more complex ones, requiring a thought process beyond mere memorization (Netri et al., 2018). The chemical equilibrium material also includes Basic Competence (KD), which requires a high level of reasoning, making it suitable as the main focus for developing the HOTS test instrument.

Due to the COVID-19 pandemic that occurred worldwide, including in Indonesia, students' chemistry thinking skills have decreased. This decline was attributed to 75% of online chemistry learning and assessments, with only 25% being face-to-face (offline) (Mushlihuddin et al., 2021). To enhance the quality of Indonesian education and promote higher-order thinking skills (HOTS) during the pandemic, there has been a need for a practical and easy-to-implement HOTS test instrument (Wahyudi et al., 2021). In this context, a three-tier multiple-choice test instrument is advantageous for several reasons. It offers objective assessment, economic and practical benefits, and can measure a large sample. Additionally, it allows researchers to validate the instrument and predict HOTS scores. The correlation between two-tier scores and confidence levels provides evidence of test validity (Laksono, 2020).

The effectiveness of a three-tier multiple-choice test instrument in measuring higher-order thinking skills has been demonstrated by previous researches. For instance, Laeli et al. (2021) developed a three-tier multiple-choice diagnostic test instrument to identify students' critical thinking skills in learning the nature of light in science. Similarly, Sari et al. (2019) developed a three-tier multiple-choice diagnostic test instrument to identify and enhance students' HOTS. Pradana et al. (2021) also developed a HOTS instrument based on Marzano's Dimensions of Learning, which met validity, reliability, and visibility criteria.

In line with the above previous studies, the present research aimed to develop an appropriate HOTS test instrument for students' chemical equilibrium material using a three-tier test. This instrument, called the Three-Tier Chemical Equilibrium Multiple-Choice Test–HOTS (TTCEMCT–HOTS), integrated higher-order thinking skills indicators outlined by Brookhart (2010). It is expected to serve as an alternative tool for teachers to measure students' HOTS in chemical equilibrium, particularly during the COVID-19 pandemic.

This study strived to measure and identify seven aspects of higher-order thinking skills indicators: analyzing, evaluating, creating, reasoning and logic, decision-making, problem-solving, and creativity (Brookhart, 2010, pp. 144–147).

Method

The research design used was Research and Development (R&D), aiming to develop a TTCEMCT–HOTS instrument based on Brookhart's framework. The research followed the instrument development steps outlined by Nabilah et al. (2019), consisting of three stages: (1) preliminary study, (2) product draft planning, and (3) product development. The development procedure is detailed in Table 1.

Table 1
Development Procedures

Stages	Activities
Preliminary study	1. Determining the problem 2. Reviewing literature related to relevant topics 3. Analyzing the chemical equilibrium basic competence 4. Analyzing Brookhart's HOTS indicators 5. Creating an essay test grid based on Brookhart's HOTS indicators
Product draft planning	1. Arranging the essay test based on Brookhart's HOTS indicators 2. Conducting validation with expert validators 3. Conducting a preliminary test of the essay test based on Brookhart's HOTS indicators 4. Converting Brookhart's HOTS indicators-based essay test into a two-tier test 5. Conducting final validation with expert validators
Product development	1. Adding a confidence rating to the two-tier test to form the TTCEMCT–HOTS instrument 2. Implementing the TTCEMCT–HOTS instrument 3. Distributing student response questionnaires 4. Conducting instrument quality tests using the ANATES software 5. Analyzing students' questionnaire responses and HOTS achievement

The population in this study included all Class XI MIPA students at SMAN 4 in South Tangerang City during the 2021/2022 academic year. A purposive sampling technique was employed, selecting classes that were studying or had studied chemical equilibrium material. This selection resulted in a research sample of 76 students from Class XI MIPA 1 and Class XI MIPA 2 at SMAN 4 in South Tangerang City. The research was conducted during the even semester from January 17 to January, 2022.

The research instruments employed included (1) a TTCEMCT–HOTS instrument validation sheet evaluated by expert validators, comprising two lecturers from the Chemistry Education Study Program at UIN Syarif Hidayatullah Jakarta, (2) a student response questionnaire regarding the developed TTCEMCT–HOTS instrument, and (3) a TTCEMCT–HOTS instrument.

The results of the TTCEMCT–HOTS instrument were analyzed quantitatively using the ANATES software, which is known for its ease of use, speed, and practicality (Nabilah et al., 2019). The empirical validity analysis included validity, reliability, difficulty level, and discriminating power tests. Conducting a quantitative analysis of the empirical data on the developed instrument is essential to obtain high-quality items that accurately reflect students' higher-order thinking skills (Ulum, 2017). Difficulty level scale and discriminating power scale is shown in Table 2 and Table 3.

Table 2

Difficulty Level Scale

Difficulty Level	Criteria
0.71 – 1.00	Easy
0.31 – 0.70	Moderate
0.00 – 0.30	Difficult

(Sudijono, 2013, p. 372)

Table 3

Discriminating Power Scale

Discriminating Power	Classification	Interpretation
0.70–1.00	Excellent	Highly Discriminative
0.40–0.69	Good	Moderately Discriminative
0.20–0.39	Satisfactory	Fairly Discriminative
0.00–0.19	Poor	Poorly Discriminative
Negative	–	Non-Discriminative

(Magdalena et al., 2021)

The data from the student response questionnaire were qualitative and quantified using a Likert scale (5 levels of criteria). The conversion of student questionnaire response scores is shown in Table 4.

Table 4

Linkert Scale

Alternative Answer	Item Score	
	(+)	(–)
Strongly agree (SA)	5	1
Agree (A)	4	2
Neutral (N)	3	3
Disagree (D)	2	4
Strongly disagree (SD)	1	5

(Mamondol, 2019, p. 162)

The instructions for scoring students' answers on the TTCEMCT–HOTS instrument are shown in Table 5. The processing of student score data on the TTCEMCT–HOTS instrument was conducted using a percentage formula. The results were then determined based on Table 6.

Table 5
TTCEMCT–HOTS Scoring

Tier				
I	II	III	Score	Student Category
Correct	Correct	Confident	7	Proficient
Correct	Incorrect	Confident	6	Not proficient
Incorrect	Correct	Confident	5	
Incorrect	Incorrect	Confident	4	
Correct	Correct	Uncertain	3	
Correct	Incorrect	Uncertain	2	
Incorrect	Correct	Uncertain	1	
Incorrect	Incorrect	Uncertain	0	

Table 6
HOTS Achievement

Percentage	HOTS Category
81 – 100%	Excellent
61 – 80%	Good
41 – 60%	Satisfactory
21 – 40%	Needs Improvement
0 – 20%	Poor

(Arikunto, 2013, p. 44)

Results and discussion

Development of TTCEMCT–HOTS

The preliminary study stage included a literature review on relevant HOTS test instruments, such as the use of a two-tier test form, descriptions, and the development of a three-tier multiple choice test. The concept of chemical equilibrium was analyzed based on the Basic Competence (KD) of the 2013 Curriculum regarding senior high school chemistry subjects. Indicators of higher-order thinking skills, as outlined by Brookhart (2010), were analyzed using stimuli and question formats following Brookhart's guidelines for item preparation (2010). The development of the TTCEMCT–HOTS instrument began with integrating Brookhart's HOTS indicators into a description test, serving as a preliminary test to identify the range of student responses. This approach clarified and focused on preparing alternatives in the answer and reason tiers, making them more effective and deceptive (Nabilah et al., 2019). The term "more deceptive" refers to the ability of alternative distractors in the answers and reasons to meet one of the essential rules for multiple-choice question construction: homogeneity. This allows alternative distractors to be selected by students who cannot answer the TTCEMCT–HOTS questions accurately (Widiyaningrum et al., 2020). As a result, a HOTS-based essay test instrument consisting of 40 questions was validated by expert validators.

The product draft planning stage included the preparation of a HOTS-based essay test that had been validated. A total of 40 items were tested on 38 students in Class XI MIPA 1 at SMAN 4 in South Tangerang. In this sub-stage, the researchers obtained data on the variation of students' actual answers. The next step involved converting Brookhart's HOTS indicators-based essay test into a two-tier HOTS test, referencing Brookhart's guidelines for preparing HOTS items. Data analysis of students' actual answers and analysis of the chemical equilibrium concept were then used as a reference in constructing answer keys and correct reasons, while students' ambiguous answers were used as alternatives for distractors (Nabilah et al., 2019). As a result, a total of 43 items of the two-tier HOTS test were validated by expert validators.

Volume 6, Issue 1, 2024

The product development stage included adding a confidence rating to the validated two-tier HOTS test items as the third tier, forming the TTCEMCT–HOTS instrument. The reference scale and confidence rating criteria are shown in Table 7.

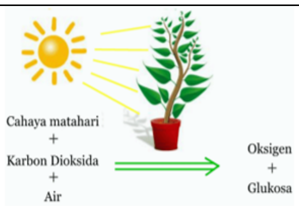
Table 7
Confidence Rating Matrix

CR Scale	Criteria
1	Guessing
2	Very unsure
3	Unsure
4	Moderately Sure
5	Sure
6	Very sure

(Caleon & Subramaniam, 2010)

The TTCEMCT–HOTS instrument, consisting of 43 questions, was then implemented with 76 students in Class XI MIPA 1 and Class XI MIPA 2 at SMAN 4 in South Tangerang City. Examples of questions on the TTCEMCT–HOTS instrument are presented in Figure 1.

Figure 1
TTCEMCT-HOTS Instrument Item



Seperti yang kita tahu, fotosintesis merupakan suatu proses biokimia oleh tumbuhan hijau dalam mendapatkan makanannya. Reaksi yang terjadi di dalam tumbuhan tersebut adalah sebagai berikut:

$$6CO_{2(g)} + 6H_2O_{(g)} \rightleftharpoons C_6H_{12}O_{6(s)} + 6O_{2(g)}$$

Identifikasi pernyataan yang paling tepat untuk reaksi reversibel pada proses fotosintesis?

a. Reaksi fotosintesis terjadi secara statis
b. Hasil fotosintesis kembali menjadi pereaksi akibat dari proses respirasi tumbuhan
c. Fotosintesis berlangsung tuntas karena karbondioksida dan air habis bereaksi
d. Reaksi maju ketika terjadi penguraian glukosa menjadi karbondioksida dan air
e. Reaksi balik ketika pembentukan molekul glukosa dan oksigen

Alasan yang tepat untuk jawaban Anda adalah

a. Jika suhu diturunkan, produksi glukosa akan semakin banyak
b. Jika volume air ditambah, produksi glukosa akan semakin melimpah
c. Jika suhu dinaikkan, produksi glukosa akan semakin banyak
d. Jika konsentrasi oksigen ditambah, produksi glukosa juga semakin banyak
e. Jika tekanan karbondioksida diturunkan, produksi glukosa akan semakin melimpah

Seberapa yakin Anda atas jawaban dan alasan yang Anda berikan?

(1) (2) (3) (4) (5) (6)

Student questionnaire responses were distributed to determine the readability of the questions and the practicality of the TTCEMCT–HOTS instrument. A total of 13 validated statements in the student response questionnaire were presented and given to students who had completed the developed TTCEMCT–HOTS instrument. The percentage of student responses obtained is shown in Table 8.

Table 8
Results of Student Response Questionnaire

No.	Indicators	Percentage	Category
1.	Alignment of chemical equilibrium material with higher cognitive levels	88.6%	Good
2.	Alignment of TTCEMCT–HOTS instrument with chemical equilibrium material	73.4%	Fair
3.	Readability of TTCEMCT–HOTS instrument items	93.4%	Good
4.	Clarity of TTCEMCT–HOTS instrument items	80.8%	Good
5.	Adequacy of processing time and number of questions provided	70.5%	Fair
6.	Benefits of TTCEMCT–HOTS instrument	80%	Good
7.	Student interest in TTCEMCT–HOTS instrument	77.6%	Good
	Average	80.6%	Good

Quality of Instrument

Validity

The first quality test conducted on the TTCEMCT–HOTS instrument was a validity test. It was carried out to determine the instrument's accuracy in identifying students' higher-order thinking skills (HOTS) in the context of chemical equilibrium material (Mubarak et al., 2016). Of 43 items on the TTCEMCT–HOTS test, 32 items, or about 74%, were significant and valid for identifying students' higher-order thinking skills in chemical equilibrium material.

Reliability

The reliability test was conducted to determine the consistency of the TTCEMCT–HOTS instrument (Laeli et al., 2021). The average reliability coefficient for both the answer and reason tiers was 0.82, indicating a very high level of consistency. This result suggests that the TTCEMCT–HOTS instrument was highly reliable in measuring higher-order thinking skills. Accordingly, a higher reliability coefficient (closer to 1) signifies greater measurement accuracy (Kusaeri & Suprananto, 2012, p. 177).

Difficulty Level

A difficulty level test was conducted on both tiers of answers and reasons, resulting in 32 valid items of moderate difficulty. Questions of moderate difficulty are neither easy nor difficult, ensuring students are appropriately challenged (Putri & Ofianto, 2019). A well-balanced question encourages students to put in the effort to solve HOTS questions without feeling overwhelmed. If questions are too easy, students are not motivated to exert more effort; if too difficult, they may feel discouraged and lose motivation (Pradana et al., 2021). The purpose of the prepared test instrument was to measure higher-order thinking skills, which do not necessarily correlate with a high level of difficulty. Hence, the dimensions of LOTS vs. HOTS and easy vs. difficult do not imply the same thing (Subakti, 2021).

Discriminating Power

Discriminating power refers to the ability of a question to distinguish between high-ability and low-ability students (Putri & Ofianto, 2019). A good HOTS item can effectively differentiate students who possess higher-order thinking skills from those who do not. Thus, the test objectives cannot be achieved if the items cannot make this distinction (Mubarak et al., 2016). Based on the results of the discriminating power test using the ANATES software, it was found that 32 items in the three-tier multiple-choice test instrument could be categorized as follows: 1 excellent item, 26 good items, and 5 fair items to identify higher-order thinking skills in the chemical equilibrium material.

The quality test on the TTCEMCT–HOTS instrument confirmed the validity of 32 items. The empirical test results of the items are summarized in Table 9.

Volume 6, Issue 1, 2024

Table 9
Instrument Quality Test Results

Reliability	Difficulty Level		Valid Item		DP		DP Interpretation	
	Tier							
	I	II	I	II	I	II	I	II
0.82	Moderate		19		0.70 – 1.00		Excellent	
			3, 4, 5, 6, 7, 9, 10, 12, 13, 17, 22, 23, 24, 25, 26, 27, 28, 29, 30, 33, 37, 38, 39, 40, 41, 43		0.40 – 0.69		Good	
			11, 14, 15, 34, 35		0.20 – 0.39		Fair	

Student Response and Higher-Order Thinking Skills

The analysis results of the student response questionnaire revealed that 88.6% of respondents agreed that the chemical equilibrium material requires higher-order thinking skills, as it necessitates higher-order thinking processes to build more complex concepts (Netri et al., 2018). Additionally, 73.4% of respondents agreed that the test instrument in the form of three-tier multiple-choice questions was adequate for measuring higher-order thinking skills in chemical equilibrium material. Besides being used for diagnostic tests, the three-tier multiple-choice format integrated with HOTS indicators could also identify and improve higher-order thinking skills (Sari et al., 2019).

The readability of the HOTS-based three-tier multiple choice test instrument also received a positive response, with 93.4% of respondents indicating that the sentences, question stimuli, and instructions presented in the instrument were easy to read. Questions designed to demand high reasoning are required to be based on fundamental questions (stimuli) in the form of reading texts, paragraphs, drama texts, novel fragments, stories, fairy tales, poems, cases, pictures, graphics, photos, tables, formulas, lists of words, symbols, examples, maps, films, or recorded sounds (Kusaeri & Suprananto, 2012, p. 152).

The presentation of the questions in the test instrument, developed using the Indonesian standard language, was deemed communicative and free from misinterpretation, with 80.8% of students agreeing. Furthermore, 77.6% of students showed interest in the three-tier chemical equilibrium multiple-choice test instrument based on Brookhart's HOTS indicators that had been developed. The students reported feeling new experiences when working on questions integrated with indicators of higher-order thinking skills in the three-tier multiple-choice format, and they supported the implementation of similar tests for other materials.

The pattern of student responses that can be categorized as capable of answering HOTS questions is as follows: if in the first tier, students provide correct answers, and in the second tier, they give the correct reasons for their chosen answers with a high level of confidence (Peşman & Eryilmaz, 2010). The high level of confidence was measured on a scale of 4 (moderately sure), 5 (sure), and 6 (very sure). In this regard, students with low confidence levels were considered to lack confidence in answering HOTS questions, implying they could not complete them (Caleon & Subramaniam, 2010). The identification of the pattern of student responses in the answer tier–reason tier–confidence level was as follows: Correct–Correct–Confident was given a score of 7; Correct–Incorrect–Confident was scored 6; Incorrect–Correct–Confident was scored 5; and Incorrect–Incorrect–Confident was given a score of 4. This scoring system refers to the guidelines provided by Sari et al. (2019).

The analysis of the score data indicated the achievement of higher-order thinking skills among students of Class XI MIPA 1 and Class XI MIPA 2 at SMAN 4 in

South Tangerang City on the chemical equilibrium material. The results were categorized as fair, with 46.42% of students able to correctly complete the HOTS questions and confidently provide reasons. 51.57% of students could not complete both the answer tier and the reason tier correctly on the HOTS questions provided. Meanwhile, 2.01% of students could only complete one of these tiers (either the answer or the reason) but had a high level of confidence in their responses, indicating possible misconceptions (Nabilah et al., 2019).

In the present study context, misconceptions are associated with inaccuracies in students' understanding of a concept. Specifically, students are considered to be at the stage of higher-order thinking if, in addition to understanding a concept, they can apply theory, analyze, evaluate, and make decisions (Fitria, 2014). The percentage of students' achievement in higher-order thinking skills for each indicator was as follows: 76.9% could analyze; 75.4% could evaluate; 76.7% could create; 75.3% could reason and use logic; 80.1% could make decisions; 79.1% could solve problems; and 71.8% could think creatively. Ultimately, the average percentage of achievement in higher-order thinking for all HOTS indicators was 76.4%, which was categorized as good.

Based on Brookhart's HOTS indicators, the student HOTS profile exhibited the highest achievement in decision-making. Specifically, 80.1% of students could answer correctly on both the answer and reason tiers as well as demonstrated confidence in their responses. Brookhart's HOTS indicators for decision-making assessed students' higher-order thinking skills in (1) determining which parts of the presented information were reliable and explaining the reasons, (2) explaining the necessary assumptions to make their arguments or explanations coherent, and (3) choosing an implicit assumption from a set of options.

Conversely, according to Brookhart's HOTS indicators, the lowest achievement in the student HOTS profile was in creating and creativity. This indicator assessed students' higher-order thinking skills in (1) creating something original, (2) organizing existing materials in new ways, and (3) reframing the question or problem in a different way.

Conclusion

The TTCEMCT-HOTS instrument to measure higher-order thinking skills was developed through three main stages: (1) preliminary study, (2) product draft planning, and (3) product development. The developed instrument achieved empirical validity with 32 valid items and a reliability value of 0.82 (very high). The items' difficulty level was moderate, complemented with good discriminating power. Among the students of Class XI MIPA 1 and Class XI MIPA 2 at SMAN 4 in South Tangerang City, 46.42% could solve higher-order thinking questions in the form of three-tier multiple-choice test quite well. In contrast, 53.58% could not solve the HOTS questions effectively. The test instrument received a positive response from 80.6% of respondents, indicating it was categorized as good for measuring higher-order thinking skills in chemical equilibrium material.

Acknowledgement

The authors extend their sincere gratitude to all individuals and institutions whose assistance and contributions were instrumental in the successful completion of this study.

References

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.

Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik Edisi Revisi*. Jakarta: PT Rineka Cipta.

Azmi, L. N., Nurhayati, S., Priatmoko, S., & Wardani, S. (2021). Pengembangan Instrumen Tes untuk Mengukur Hots Peserta Didik pada Materi Laju Reaksi. *School Science and Mathematics*, *2*(3), 133–139.

Brookhart, S. M. (2010). *How to Assess Thinking Higher-Order Skills in Your Classroom*. ASCD.

Caleon, I., & Subramaniam, R. (2010). Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves. *International Journal of Science Education*, *32*(7), 939–961.

Dewati, D., Hadiarti, D., & Fadhilah, R. (2016). Pengembangan Instrumen Penilaian Tes Diagnostik Pilihan Ganda Dua Tingkat untuk Mengukur Hasil Belajar Siswa Materi Hidrokarbon di SMA 10 Negeri Pontianak. *AR-RAZI Jurnal Ilmiah*, *4*(2), 23–32.

Fitria, A. (2014). Miskonsepsi Mahasiswa dalam Menentukan Grup pada Struktur Aljabar Menggunakan Certainty of Response Index (CRI) di Jurusan Pendidikan Matematika IAIN Antasari. *JPM IAIN Antasari*, *01*(2), 45–60.

GTK Pendidikan Dasar. (2021). *Modul Belajar Mandiri Calon Guru Aparatur Sipil Negara (ASN) Pegawai Pemerintah dengan Perjanjian Kerja (PPPK)* (pp. 1–184). Direktorat Jenderal Guru dan Tenaga Kependidikan. Diakses tanggal 13 Juli 2022 dari http://repositori.kemdikbud.go.id/id/eprint/22649

Hidayatullah, A. R., Yamtinah, S., & Masykuri, M. (2022). Development of A Two-Tier Multiple-Choice Instrument Based on Higher Order Thinking Skills (HOTS) on Acids, Bases, and Salts. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 932–938.

Hikmah, H., & Amin, N. (2019). Pengembangan Instrumen untuk Mengukur Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Matematika di SMA Kabupaten Majene. *Saintifik*, *5*(1), 1–7.

Juliarti, H., Widiarti, H. R., & Nazriati. (2019). Prosiding Kajian Literatur : Kemampuan Berpikir Tingkat Tinggi dalam Pembelajaran Kimia. *Prosiding*, *November*, 313–317.

Kusaeri, & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan* (1st ed.). Yogyakarta: Graha Ilmu.

Laeli, C. M. H., Gunarhadi, & Muzzazinah. (2021). Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning. *İlköğretim Online*, *20*(2), 277–287.

Laksono, P. J. (2020). Pengembangan Three Tier Multiple Choice Test Pada Materi Kesetimbangan Kimia Mata Kuliah Kimia Dasar Lanjut. *Orbital: Jurnal Pendidikan Kimia*, *4*(1), 44–63.

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 Dan Sains*, *3*(2), 198–214.

Mamondol, M. R. (2019). *Dasar-Dasar Statistika*. Surabaya: Scopindo Media Pustaka.

Mubarak, S., Susilaningsih, E., & Cahyono, E. (2016). Pengembangan Tes Diagnostik Three Tier Multiple Choice Untuk Mengidentifikasi Miskonsepsi Peserta Didik Kelas Xi. *Journal of Innovative Science Education*, *5*(2), 101–110.

Mushlihuddin, R., Wahyuni, S., Harahap, T. H., & Hafni, R. C. (2021). Penerapan Pembelajaran dan Penilaian Berorientasi HOTS pada Masa Pandemi COVID 19 bagi Guru-guru Muhammadiyah. *JURNAL PRODIKMAS Hasil Pengabdian Kepada Masyarakat*, *6*(2), 144–148.

Nabilah, L. Y., Ruslan, R., & Rusli, R. (2019). Pengembangan Instrumen Diagnostik Three Tier Test pada Materi Pecahan Kelas VII. *Issues in Mathematics Education (IMED)*, *3*(2), 184–193.

Netri, N., Holiwarni, B., & Abdullah. (2018). Development of Test Instruments Based Higher Order Thinking Skill (HOTS) on Chemical Equilibrium at Second Grade in Senior High School. *Jurnal Online Mahasiswa Fakultas Keguruan Dan Ilmu Pendidikan*, *5*(2), 1–11.

Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *Journal of Educational Research*, *103*(3), 208–222.

Pradana, I. G. Y., Suma, K., & Sujanem, R. (2021). Pengembangan Tes Keterampilan Berpikir Tingkat Tinggi Tiga Level (Three-Tier) Berbasis Marzano Dimensions Of Learning. *Wahana Matematika Dan Sains: Jurnal Matematika, Sains, Dan Pembelajarannya*, *15*(3), 119–128.

Putri, R. H., & Ofianto. (2019). Efektivitas Analisis Butir Menggunakan Anajohn, Anates dan Iteman Studi Soal USBN Pelajaran Sejarah Kota Padang. *Jurnal Mahasiswa Ilmu Sejarah Dan Pendidikan*, *1*(2), 1–11.

Ramadhan, G., Dwijananti, P., & Wahyuni, S. (2018). Analisis Kemampuan Berpikir Tingkat Tinggi (High Order Thinking Skills) Menggunakan Instrumen Two Tier Multiple Choice Materi Konsep Dan Fenomena Kuantum Siswa Sma Di Kabupaten Cilacap. *UPEJ Unnes Physics Education Journal*, *7*(3), 85–90.

Sari, N. W. N., Sunyono, S., & Abdurrahman, A. (2019). Development Of The Three Tier Diagnostic Test Based "Higher Order Thinking Skills" Instrument. *Dinamika Jurnal Ilmiah Pendidikan Dasar*, *11*(2), 86–93.

Shidiq, A., Masykuri, M., & Van Hayus, E. S. (2014). Pengembangan Instrumen Penilaian Two-Tier Multiple Choice Untuk Mengukur Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa Sma/Ma Kelas Xi. *Jurnal Pendidikan Kimia*, *3*(4), 83–92.

Subakti, Y. R. (2021). Tantangan Guru Sejarah dalam Mengajar dan Belajar Berbasis HOTS. *HISTORIA VITAE*, *01*(01), 52–71.

Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan*. Yogyakarta: Raja Grafindo Persada.

Sutiana, Silitonga, H. T. M., & Oktavianty, E. (2018). Pengembangan Tes Diagnostik Three Tier Test pada Materi Suhu dan Kalor Untuk SMP. *Gastrointestinal Endoscopy*, *10*(1), 279–288.

Ulum, M. (2017). Analisis Butir soal Ulangan Harian pada Mata Pelajaran Ekonomi KD 3.1 Pendapatan Nasional Kelas XI IPS 1 di SMA Negeri 1 Gresik. *Jurnal Pendidikan Ekonomi*, *5*(3), 1–5.

Wahyudi, F., Didik, L. A., Bahtiar. (2021). Pengembangan Instrumen Three Tier Test Diagnostik untuk Menganalisis Tingkat Pemahaman dan Miskonsepsi Siswa Materi Elastisitas. *RELATIVITAS: Jurnal Riset Inovasi Pendidikan Fisika*, *4*(2), 48–58.

Widiyaningrum, D. K., Syamsiah, N., & Septiani, R. (2020). Analisis Kualitas Butir Soal Multiple Choice pada Tes Akademik Matematika Seleksi Penerimaan Mahasiswa Baru Sekolah Tinggi Sandi Negara (SPMB STSN) Tahun Akademik 2019/2020. *Jurnal IPSIKOM*, *8*(2), 1–12.