Eurasian Journal of Educational Research 107 (2023) 140-164

Eurasian Journal of Educational Research
www.ejer.com.tr

Development of High-Order Thinking Skills (HOTs) Instruments with a Socio-Scientific Problem-Based Learning Approach on the Subject of Viruses

Annur Indra Kusumadani[1]*, Sentot Budi Rahardjo[2], Sri Yamtinah[3], Baskoro Adi Prayitno[4]

ARTICLEINFO

Article History:
Received: 27 January 2023
Received in revised form: 25 July 2023
Accepted: 20 September 2023
DOI: 10.14689/ejer.2023.107.009

Keywords
High-Order Thinking Skills, Socio-Scientific, Problem-Based Learning, Assessment Instruments.

ABSTRACT

Purpose: Higher-Order Thinking skills (HOTs) oriented learning is based on socio-scientific problem-based learning in higher education and requires a valid and reliable instrument to measure these skills. This study plans to deliver a high-order thinking expertise instrument model with a sociological issue based-learning approach. **Methodology:** This appraisal model proposes the Plan and Improvement approach, which alludes to the Ellis and Duty model, to be specific comprising: (1) issues, (2) objective setting, (3) model plan and advancement, (4) model testing, (5) model assessment of experimental outcomes, and (6) application models. Participants in this research consisted of eighty (80) prospective science teacher students from three institutions in Indonesia. The expert validationsheetinstrument wasused to validate thecontent. **Findings:** The study assessed the construct validity of the proposed measurement model via the utilization of Confirmatory Factor Analysis (CFA). Evaluation of the unwavering quality of the estimation model incorporates Composite Dependability (CR) with a worth of 0.9 and Normal Fluctuation Removed (AVE) with a worth of 0.7, showing that the instrument is in the solid class. Corroborative variable examination showed that the worth of the Root Mean Square Blunder of Guess (RMSEA) was 0.01 < 0.08, the Chi-Square/or got from the test is 1.0 < 2, and the Decency of fit Record (GFI) is 0.91 > 0.90, or the expressed model compares to the information got in the field and can be utilized in a great many estimations. Confirmation of content legitimacy using the Aiken equation gets a base score of 0.8 on the fantastic classification. **Implications to Research and Practice:** These assessment instruments can be recommended to measure HOTs in socio-scientific problem-based learning but limited to discussing only viruses.

[1] Doctoral Program of Natural Science Education, Universitas Sebelas Maret, Surakarta, Indonesia. Email: aik110390@student.uns.ac.id
& Department of Biology Education, Universitas Muhammadiyah Surakarta, Indonesia. Email: aik120@ums.ac.id
[2] Doctoral Program of Natural Science Education, Universitas Sebelas Maret, Surakarta, Indonesia. Email: sentotbr@staff.uns.ac.id
[3] Doctoral Program of Natural Science Education, Universitas Sebelas Maret, Surakarta, Indonesia. Email: jengtina@staff.uns.ac.id
[4] Doctoral Program of Natural Science Education, Universitas Sebelas Maret, Surakarta, Indonesia Email: baskoro_ap@fkip.uns.ac.id
*Corresponding Author: Annur Indra Kusumadani, Email: aik120@ums.ac.id

1. Introduction

Higher-order thinking skills (HOTs) have the great potential to reform the education system, and prioritize the demands of students' ability to choose intellectually the valid and relevant information (Mulyono & Yamtinah, 2023; Yuliarti et al., 2023). These skills are termed as primary, innovative, open, and cooperative reasoning abilities, which help to grow critical abilities to think, cooperate, use data innovation, refresh, monitor globalization and develop natural mindfulness. These skills also help develop the competence of graduates to master thinking, complex communication and problem-solving skills (Fong et al., 2017; Kembara, Rozak, & Hadian, 2019). HOTs are also linked with the 21st-century abilities known as 4C: corporate decisive reasoning, correspondence, coordinated effort, and creativity (Astuti et al., 2019; Kembara et al., 2019; Ku & Au, 2021). In short, these skills are essential in developing students' potential to become critical, creative, and independent human beings.

Being a primary focal point of 21st-century learning skills (Abdullah, Abidin, & Ali, 2015; Ghobadi & Taki, 2018; Shukla & Dungsungnoen, 2016), HOTs are likewise vital for learners in tackling learning issues (Almerich et al., 2020; Halili, Razak, & Zainuddin, 2015). HOTs-based learning works with undergraduates to think consistently and deliberately as indicated by the attributes of the subjects and learning materials and develop higher-request thinking abilities in them. In addition, HOTs have a positive effect on learning outcomes as well. This is in line with the results of Ganiron Jr (2014) research which shows that learning that emphasizes HOTs can improve learning outcomes. HOTs also assist teachers in shifting from simple to complex actions highlighting the need to transform in teaching the lower-order thinking skills into higher-order thinking skills (Afandi et al., 2019). Anderson et al. (2001) consider HOTs supporting legitimate thinking to get answers or perceive and chase the investigation of realities logically. This suggests that when individuals learn and think, they might experience questions or unforeseen issues and circumstances (Makmuri & Aziz, 2021). Consequently, HOTs incorporate decisive and inventive reasoning abilities, so they should be sustained, created, and improved through viable learning techniques.

The problem-based learning (PBL) is one of the practical learning strategies to develop HOTs. By providing real problems and problem-solving steps, students can build abilities in conducting research, sharing insights, performing actual experiments, and applying them with a high-order thinking ability (Bayat & Tarmizi, 2012; Mtetwa & Ndemo, 2022; Prayitno & Jaedun, 2022). PBL always starts with a problem so that this learning model can improve HOTs. In line with the research results of Şendağ and Odabaşı (2009), blended-PBL is significant in increasing HOTs. PBL helps students become independent learners (Arends, 2012). Therefore, prospective teacher students are expected to be able to study independently both in the network and outside the network.

Contextually, PBL can be integrated with socio-logical issue (SSI) based learning of the contemporary issues dubious in the public eye. It develops discourse, conversation, or discussion of moral thinking or assesses moral issues in dynamic cycles regarding potential answers for social issues. This approach also develops basic and imaginative reasoning in a multidisciplinary and integrative way to tackle logical issues and virtues. One of the advantages of SSI coordination is that it powers learners to foster a scope of decisive abilities to reason and manners through conversations and discussions on dubious social subjects. Furthermore, examination, induction, clarification, judgment, understanding, and self-guideline rely upon the capacity to think inventively (Salleh, Tasir, & Shukor, 2012; Wan Mohamed Salleh et al., 2022). Along these lines, consolidating SSI parts can help undergraduates find open, logical, calculated, creative, critical thinking. SSI provides a real-life context, allowing students to grasp these issues' scientific and social impact, involving scientific understanding, discussion, collaboration, and decision-making.

Developing appropriate assessment instruments for HOTs in SSI and PBL is also challenging, requiring measurement of conceptual understanding, higher-order thinking, and effective communication, ideally through authentic assessments like projects or simulations. Teacher training is crucial for designing and implementing HOTs assessments in SSI and PBL, fostering a deep understanding of integrating higher-order thinking skills. The current study aimed to address these challenges to positively contribute to developing students' skills for real-life challenges. This study is in accordance with the existing studies including Romine, Sadler, and Kinslow (2017), which developed the Assessment of Socio-Scientific Reasoning; Setambah (2018), which developed an instrument model based on socio-scientific issues to measure students' reasoning skills; Rahayu and Rosawati (2023), which developed an assessment instrument higher order thinking abilities (HOTs) in the field of chemistry to examine the context of sociospatial issues; and Mulyono and Yamtinah (2023), which developed a socio-scientific issue-based instrument to measure critical and creative thinking skills.

The current study also complements subsequent research by developing socio-scientific issue-based instruments starting with problem-based learning. This study also promotes how students' HOTs on viruses can be measured with valid and reliable instruments. It should be noted that HOTs competency is listed as one of the significant achievements in the Ministry of Science Education curriculum. To measure HOTs, however, there is a need to develop a valid and trustworthy instrument model, and an evaluation tool that must be combined with a supporting strategy to measure HOTs better successfully. This study aims to develop valid HOTs instruments that can measure both content and construction variables. The development of the HOTs instrument is important as an evaluation tool and to measure the achievement of the curriculum for prospective science teacher students.

The current study embarks upon the investigation in the light of applicable hypothetical examinations and other requirements in the field, namely: (1) What are the aftereffects of the build legitimacy of the HOTs instrument with the socio-logical issue-based learning approach created? (2) Does the HOTs evaluation instrument with a socio-logical issue-based learning approach comprise of suitable inquiries considering separating power and trouble level? The fundamental standard of creating HOTs instruments in learning is that the learning accomplishment to be accomplished alludes to the most noteworthy mental space of Sprout's Scientific classification (examination, assessment, and creation). HOTs evaluation instruments are expected to survey decisive reasoning abilities, critical thinking abilities, and inventiveness, so capability-based apparatuses connected with learning are required (Yanto, Subali, & Suyanto, 2019).

Additionally, this study means to foster an instrument that can quantify the HOTs of Science instructor competitors on infections and to get the attributes of HOTs appraisal, which incorporates parts of investigation, assessment, and creation. The consequences of this study would foster another instrument model for estimating HOTs in the light of socio-logical issues. The definition of pointers in this instrument alludes to the detailing of specialists and is orchestrated by the qualities of learning considering socio-logical issues. So, the consequences of this study give another reference connected with learning gadgets.

2. Literature Review

HOTS is essential for 21st-century abilities with numerous definitions (Saido et al., 2018). For instance, HOTs can be implanted in advanced education educational plan improvement as an exchange of information in the mental space of examination, assessment, and making of Blossom's scientific classification modified by Anderson (Magsino, 2014). Through the HOTs-based educational plan, tertiary foundations can set themselves up for change. HOTs can be shown through different learning models and techniques, for example, research center practicum; (Liu et al., 2017); flipped homeroom and e-learning (Lee, Lim, & Kim, 2017); cooperative conversation (Afandi et al., 2019); mixed learning (Fryer & Bovee, 2018); and versatile learning (Sung, Chang, & Liu, 2016).

There are three categories in defining HOTs (Brookhart, 2010), namely (1) high-level thinking as a transfer of knowledge; (2) high-level thinking as critical thinking; (3) high-level thinking as problem-solving. As a transfer of knowledge, HOTs consider meaningful learning related to the top end of Bloom's taxonomy: analyze, evaluate, and create (Anderson et al., 2001). Rismi (2021) stated that HOTs are thinking models that include various cognitive skills and divide HOTs into reasoning skills, argumentation skills, problem-solving & critical thinking, creativity, and metacognition. Reasoning skills include deductive and inductive processes. The deductive process uses facts, claims, and evidence to support a conclusion, while the inductive process is the process by which verifiable conclusions are generalized to new cases. Argumentation skills are the skills to make statements, collect and evaluate evidence, combine multiple sources of evidence to support a claim and refute or reply to claims. Metacognition is knowledge about cognition and theregulation of cognition (Langdon, Kurz, & Coppola, 2023).

Shukla and Dungsungnoen (2016) indicate that HOTs are a piece of imaginative and decisive reasoning. In the interim, Magsino (2014) said that HOTs come from the mental space of Blossom's scientific classification. Considering the aftereffects of concentrating on mental regions in Blossom's scientific classification, there are six mental levels. The three fundamental abilities incorporate the memorable capacity, comprehend and apply, which are classified as low-level reasoning (Parcel), and the accompanying three levels, which have been overhauled by Anderson et al. (2001), including the capacity to dissect, assess and make which are ordered as significant level reasoning (HOT). Thus, HOTs encompass a range of cognitive abilities, including analysis, synthesis, comparison, conclusion, interpretation, evaluation, and inductive and deductive reasoning. These skills effectively address and resolve unexpected situations (Budsankom et al., 2015). Alkhatib (2019) posits that higher-order thinking encompasses critical thinking, problem-solving, decision- making, and creative thinking.

The synthesis of higher-order thinking theory based on experts can be seen in Table 1.

Table 1

Theory Synthesis of HOTs Definition.

Anderson & Krathwohl (2001)	Brookhart (2010)	(Singh et al., 2018)	Synthesis
Analyze	Transfer of knowledge	Decision-making	Transfer of knowledge
Evaluate	Solution to problem	Solution to problem	Solution to problem
Create	Critical thinking	critical thinking	Critical thinking
		Creative thinking	Meta cognition

Based on the theories of the experts mentioned in Table 1, it can be concluded that higher-order thinking is the ability to initiate specific knowledge by combining acquired knowledge that is related to one another to form a unified whole of knowledge so that it can criticize problems, solve problems then reflect on solving solutions problem. In other words, higher-order thinking is a continuous thinking process so that one can analyze a problem faced and then criticize the problem so that finally one can make a solution to solve the problem and reflect on the solution to the problem. Criticizing a problem is the ability to think deductively to formulate a problem from the information obtained and then formulate temporary conjectures and arguments to solve a problem. Problem-solving is an ability that begins with identifying problems and then analyzing information and planning solutions to solve problems to create alternative temporary solutions in the form of ideas, procedures/steps, or products in the form of work, written or oral. Reflecting on problem solutions is the ability to find their weaknesses or solutions to problems that have been put forward.

The educational program for Training Staff Schooling Establishments (LPTK) is organized so that its alums can foster understudy potential through public schooling objectives, which are set apart by the development of perspectives, information, and abilities in a coordinated way as per the skills of alums of each degree of schooling in light of the new educational plan and by current requests, alluding to the System Indonesian Public Capabilities (KKNI) and the future to ensure the nature of imminent expert teachers (Rustad, 2013). The IQF-based educational plan at level 6 underlines HOTs capacities in learners because, other than supporting their mental capacities, HOTs are additionally required to take care of issues in learners' lives (Amalia & Wuryandani, 2020; Budi & Farcis, 2021). That is upheld by the learning outcomes (LO) of science learning in the tertiary educational program alluding to the KKNI level 6, specifically, in addition to other things, when learners can think consistently, basically, efficiently, and creatively with regards to the turn of events or execution of science and innovation that focuses on and applies the qualities humanities.

Higher education systems with faculties that produce prospective teachers have a more significant responsibility than others (Philip, Schuler-Brown, & Way, 2013). Since educators are the principal entryway in passing on material data and information with the most recent events (Abremski & Roben, 2021), subsequently, they should be available to a wide range of changes and have the option to adjust to progressively progressed innovative turns of events and data frameworks (Garcia-Martin & Garcia-Sanchez, 2017; Uerz, Volman, & Kral, 2018; Utama & Nurkamto, 2020). That inexorably open and associated world implies that information can be acquired effectively anyplace (Bilyalova, Salimova, & Zelenina, 2020).

Wellsprings of information that can be gotten from any place expect undergraduates to access, use, and utilize it. Similarly, tertiary institutions should have the option to work with and enable their HR, the learners and instructing staff can rapidly adjust to the most recent advancements in science and innovation (Jeffrey et al., 2014). The institutions can also plan for graduates to confront these progressions by fostering an advanced education educational program.

In this context, institutions should design well and include both teachers and learners in learning exercises that can energize and foster HOTs. The instrument utilizes questions requiring information and abilities in new circumstances by elucidating articulations (Prayitno & Titikusumawati, 2018). Numerous specialists have analyzed the significance of HOTs, like Halili et al. (2015), affirming that HOTS is essential for 21st-century abilities that are vital for people in the future. They additionally analyzed HOTs and Parts, which demand a similar investment, cost, and remembered to create. HOTs are fundamental to critical thinking Rubbo and Moore (2012). Heong et al. (2016) study named "The Job of Higher Request Thinking Abilities in Green Expertise Advancement" demonstrates that HOTs can make new capabilities that align with modern improvements, such as the modern transformation 4.0. One element of HOT-based instruments is that every learner is urged to take part in coming to conclusions about different logical issues. Consequently, they should foster thinking and rational abilities given logical information (Kim et al., 2018). Expanding HOTs also influence learning rehearses (Yanto et al., 2019).

The HOTs assessment instrument based on socio-scientific problem-based learning measures students' higher-order thinking skills in complex socio-scientific issues. The following are several elements and characteristics of assessment instruments that can be used in this framework (1) Emphasis on Higher Order Thinking Skills, (2) Relevance to SSI, (3) Student Involvement in PBL, (4) Process and Outcome Assessment, (5) Collaborative Assessment, (6) Emphasis on Ethics and Social Values, (7) Multiple Assessment Formats, (8) Formative and Summative Assessments, and (9) Opportunities for Reflection and Self-Monitoring. Each of these instruments should assess HOTs, such as critical analysis, information synthesis, evidence evaluation, and decision-making. It includes students' ability to apply their knowledge to solve complex problems (Damaianti, Abidin, & Rahma, 2020).

Moreover, assessment instruments need to be relevant to the socio-scientific issues being studied. Assessment questions or tasks should reflect the real-life context of the issues to measure understanding and application of scientific concepts in real-world situations (Rahayu & Rosawati, 2023). Instruments should include PBL elements, such as problem-solving, collaboration, and presentation (Noviana et al., 2019). For example, an assessment task may ask students to design a solution to a specific socio-scientific issue or participate in a simulated scientific discussion. Instruments should assess the result of problem-solving and the processes involved in achieving that solution (Shanta & Wells, 2022). It includes the steps of analysis, research, critical thinking, and other steps students take during PBL. If PBL involves group work, assessment instruments should reflect students' collaboration ability. That may include assessing individual contributions, communication skills, and the ability to reach consensus within the group.

Instruments should also ensure ethical considerations and social values related to socio-scientific issues (Subiantoro & Treagust, 2021). Students can be assessed based on their understanding of the ethical implications of proposed solutions and the extent to which they consider social values in problem-solving. Assessment instruments can take various forms, including written exams, oral presentations, media-based projects, or portfolios (Koç & Ölmez-Çağlar, 2023). Using a variety of formats can allow students to demonstrate their skills in a way that suits their learning styles and strengths. Assessment instruments can include formative and summative aspects (Gallardo et al., 2023). Formative assessments can provide feedback during the learning process, while summative assessments provide an overall picture of a student's understanding and skills at the end of a unit or program. Assessment instruments may include elements that enable students to reflect on their learning and self-monitor the development of their HOTs. By designing a socio-scientific PBL-based HOTs assessment instrument that considers these aspects, teachers can measure and encourage the development of students' critical thinking skills in relevant and in-depth contexts.

3. Method

- *Research Design*

This study employed the Design and Development (D&D) technique, which consists of the following stages: (1) issue identification, (2) goal establishment, (3) model design and development, (4) model testing, (5) assessment of model findings, and (6) application of the developed models. This methodology rigorously examines the processes of design, development, and assessment to establish an evidence-based framework for creating novel instructional products, tools, and models. The study resulted in developing a HOTs instrument paradigm grounded on socio-scientific issues. Field studies and needs analysis were conducted to carry out problem analysis. The outcomes of this research encompass the establishment of goal-oriented frameworks, namely the creation of HOTs tools centered around socio-scientific matters.

- *Research Instrument*

The instruments for this study were compatible with the planning and development phase and the established theoretical frameworks. The instruments' compatibility with the field data was assessed based on specific criteria. Each instrument included items to identify HOTs indicators, and integrating socio-scientific concerns to create a comprehensive set of indicators. The instrument grid was devised using these indicator formulations. The prototype instrument was designed with the grid as its foundation. The expert validation sheet instrument was used to validate the content. To assess the instrument's validity, it is essential to employ content validity, explicitly utilizing the Aiken formula, which involves the participation of seven expert validators.

- *Population and Sample*

Participants in this research consisted of prospective science teacher students from three institutions in Indonesia who were sampled randomly, with a total sample size of 80 respondents. The sufficiency of the sample size was assessed using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy.

- *Data Analysis*

The study analyzed the construct validity of the proposed measurement model via the utilization of Confirmatory Factor Analysis (CFA) within the framework of a Linear Structural Model (Lisrel). It also conducted an analysis using the RASCH model to assess items' difficulty level and discriminatory power utilizing the Winsteps software. The SPSS software was used for descriptive statistics. The study measured the p-value with a minimum threshold of 0.5; which was more significant than or equal to the predetermined significance level (α = 0.05), while the goodness of fit indices (GFI) needed to be at least 0.90. Additionally, the adjusted goodness of fit index (AGFI) and the root mean square error of approximation (RMSEA) should both be greater than or equal to 0.90 and less than or equal to 0.08, respectively (Budiyono & Rohrlich, 2017). A chi-square statistical test was also employed to evaluate the following hypotheses:

H0: *Model fit (fits) with empirical data in the field*
H1: *The model does not fit (does not match) the empirical data in the field.*

If it met these criteria, it was concluded that H0 was accepted and vice versa. Item fit indicated the expected level between the actual item characteristics and the Rasch model characteristics. Items were categorized to fit to the model when the MNSQ outfit values were in the range 0.5 to 1.5, the ZSTD outfit values were in the range -2 to +2, and the correlation measurement point values were in the range 0.4 to 0.85 (Wei, Chen, & Kinshuk, 2012). Assessment of the reliability of measurement models included Composite Reliability (CR) and Average Variance Extracted (AVE). The output of this study was a valid and reliable instrument that met the criteria for measuring HOTs.

4. Results and Discussion

The study involved three major steps: Needs Analysis, Planning and Development. Each step required a protocol to be followed.

- *Needs Analysis*

In the first stage of Needs Analysis, the problem-based learning approach emphasized developing higher-order thinking skills, such as critical analysis, problem-solving, and evaluation. While it was known that HOTs assessment measured how students applied their knowledge in real-life contexts, SSI provided a real-life context for learning. Assessment of HOTs in the context of SSI allowed students to understand these issues' scientific and social impact. SSI also involved scientific understanding and required involvement in discussion, collaboration, and decision-making. Assessments reflected students' ability to participate in scientific and social dialogue. PBL involved presenting students with real-world problems, which require applying knowledge across subjects.

It was evident that HOTs assessments should reflect the integration of material and the student's ability to solve complex problems. PBL encouraged the development of metacognitive skills, such as reflection and self-monitoring. Assessment included students' ability to manage and regulate their learning processes. Developing appropriate assessment instruments to measure HOTs in the context of SSI and PBL was challenging. Assessment instruments were able to measure conceptual understanding, higher-order thinking skills, and the ability to communicate effectively. Assessments should be authentic and reflect real-world situations as much as possible. This included projects, presentations, or simulations illustrating scientific and social demands.

During this stage, it is also important that teachers must be trained to design and implement HOTs assessments in SSI and PBL contexts. There is a need to develop a deep understanding of HOTs and integrate them into assessment. If done in a proper way, involvement in real-world problem-solving can increase student motivation and strengthen the connection between classroom learning and everyday life. By understanding and addressing these challenges, SSI and PBL-based HOTs assessment approaches can positively contribute to developing students' skills and preparing them to face real-life challenges.

- *Planning Stage*

The second phase of planning is the stage when a model or prototype is designed and developed. During this phase, a comprehensive literature review was conducted, synthesizing the data from the literature research. Such activities were undertaken to provide substantiation for the development model. Moreover, the preliminary phase of designing the prototype 1 model was concluded and afterward progressed to the model testing stage. The current part of the project focused on doing a literature review to explore the notion of HOTs and various assessment methodologies. This review was the foundation for constructing an initial design model or prototype. Upon conducting a comprehensive examination of the HOTs ideas, valuable research findings were acquired about the underlying theories, methodologies, procedural stages, and techniques employed in developmental research. This activity aimed to conduct an empirical study using HOTs assessment abilities to evaluate virus-related content among science teacher candidates. An examination of pertinent research findings indicated that the HOTs assessment significantly supports advancing scientific education within university settings. The assessment of the student's abilities was determined by implementing the HOTs assessment approach.

An examination of the identified issues and requirements through preparatory activities subsequently followed the findings of the literature review. The examination of the problem indicated that the test instrument that was previously established had yet to be able to assess HOTs effectively. The assessment model for the HOTs test instrument that was previously established remains at a rudimentary stage. The existing assessment techniques still needed to demonstrate the capacity to evaluate and enhance higher-order thinking skills effectively. Finally, a problem analysis was undertaken to find research models to enhance students' HOTs skills in learning about environmental challenges. The present study's requirements analysis revealed two key findings. Firstly, it was imperative to construct an instrument model that could effectively gauge HOTs skills. Secondly, the constructed assessment model demonstrated its ability to identify the areas of weakness in prospective science teacher students' learning.

- *Development Stage*

The third stage of development led to creation of an instrument testing product designed to assess HOTs, specifically on viral content, intended for science teacher candidate students. The initial phase of the developmental process involved creating and verifying a prototype. This activity aimed to provide items that may serve as effective and accurate instruments for application. The prototype comprised a set of assessment tools, recommendations, and methodologies for interpreting the obtained assessment findings. The process of prototype planning involved engaging in design activities, which encompassed establishing objectives, selecting instruments, and developing execution guidelines. The assessment prototype's initial design underwent a review process that included the participation of specialists in the fields of environmental science and environmental education evaluation.

A comprehensive synthesis of expert definitions and indicators developed the HOTs instrument. Researchers synthesize signs of critical thinking by relying on the formulation of experts. This approach is chosen because it aligns with various aspects of model syntax, its indicators, and functions to study the virus content, as presented in Table 2.

Table 2

Hots Indicators.

Aspect	Indicators	Functions
Transfer of knowledge	The ability to initiate specific knowledge by combining acquired knowledge that is related to one another to form a unified whole of knowledge	- Evaluate - Create - Identify problems
Solve the problem	Ability to analyze information, identify problems, and plan to problem-solve to create problem-solving solutions	- Analyze information - Planning problem solutions - Create problem solutions in the form of ideas/conclusions/ products, written/oral works
Critical thinking	Ability to think deductively inductively to be able to formulate problems and formulate hypotheses and provide arguments from the results of inductive-deductive thinking	- Analyze - Formulate the problem - Formulate hypotheses - Explain the argument
Metacognition	The ability to measure oneself to be able to estimate the possibility of other ways of solving problems	- Assess self-abilities - Estimating the possibility of other ways

The creation of indicators guided the development of the prototype instrument. A content validity test was conducted using the Aiken method to assess the instrument's validity, with input from five validators. The design of the evaluation model prototype received several recommendations and contributions from experts, which should be considered when designing an assessment framework. These recommendations are: Firstly, it is crucial to establish clear and well-defined objectives that align with the principles of assessment. Secondly, the characteristics of the assessment instrument should be explicitly stated and organized in a manner that provides a distinct advantage. Thirdly, the assessment instrument should accurately reflect the principles of valuation. Lastly, it is essential to develop assessment guidelines that are practical, comprehensive, and transparent in order to ensure effective implementation.

Following the necessary repairs and revisions, the prototype of the assessment model underwent validation by specialists, resulting in the acquisition of a viable and practical model. According to the Aiken table, which was utilized to assess claims evaluated by a panel of seven experts, the criteria range of scores employed ranged from 1 to 5. The significance level set for this analysis was 0.05. The reference value for Aiken, as determined by this study, is 0.80. The findings from the expert evaluation indicated that all items achieved a minimum score of 0.8. Based on the obtained calculations, the instrument under consideration possesses legitimate content.

- *Product Trial Results*

Instrument trials were conducted to measure construct validity. The results of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy test using SPSS obtained a value of 0.797 or more than 0.5 so that the number of samples was declared eligible for the construct validity test. The results of the construct validity test using CFA using Lisrel are presented in Figure 1.

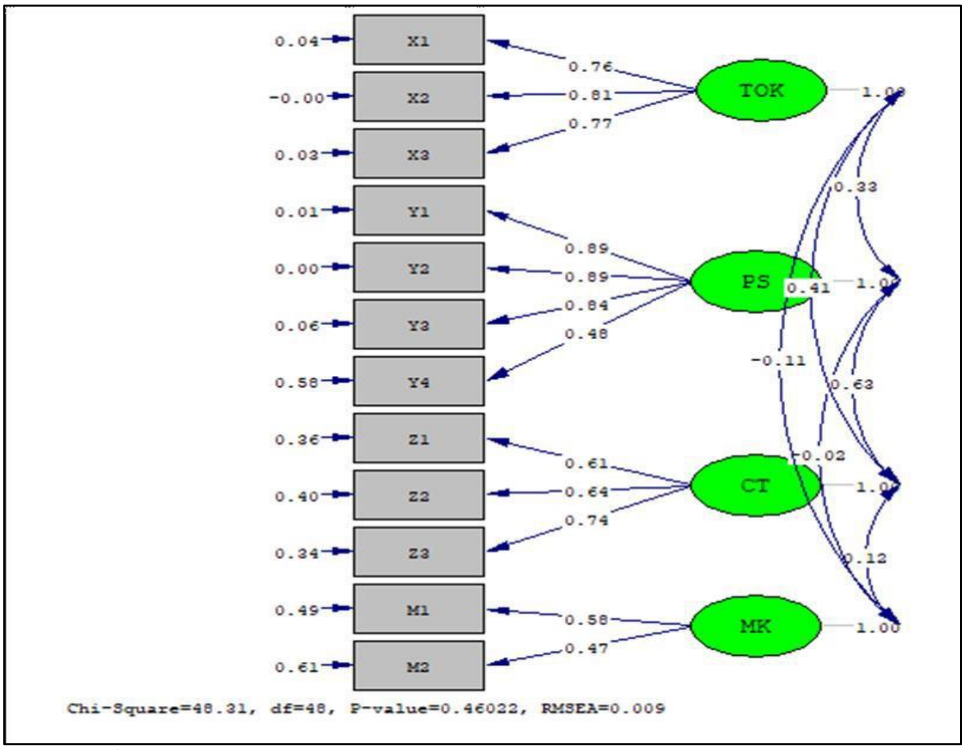

Figure 1: *Output Diagram of Estimating Construct Validity Test Using CFA.*
Note: TOK: Transfer of knowledge; PS: Problem solving; CT: Critical thinking; MK: Metacognition

Figure 1 shows that the model is fit. The detailed interpretation of the analysis results is presented in Table 3.

Table 3

Model Fit Test Results.

Indicator	Reference Score	Earned Score	Fit Models
Chi-Square/dr	≤ 2.00	1.00	Fit
Probability (p-value)	≥ 0.08	0.46	Fit
Root Mean Square Error of Approximation (RMSEA)	≤ 0.08	0.01	Fit
Goodness of Fit Index (GFI)	≥ 0.90	0.91	Fit
Comparative Fit Index (CFI)	≥ 0.90	0.99	Fit
Relative Fit Index (RFI)	≥ 0.90	0.92	Fit
Incremental Fit Index (IFI)	≥ 0.90	0.99	Fit
Parsimony Goodness of Fit Index (PGFI)	≥ 0.05	0.56	Fit

Theoretical conceptions that had been developed were substantiated by empirical facts in the area. According to (Sörbom, 1989), a model was considered to fulfill the necessary model requirements if it satisfied a minimum of three model fit indices. The model fit index comprised the Chi-Square ($\chi 2$), Root Mean Square Error of Approximation (RMSEA), p-Value, and Goodness of Fit Index (GFI) test. Overall, the Goodness of Fit criteria had been satisfied since the generated values fall within the specified intervals, indicating that the model was suitable. Specifically, the instrument's construct had favorable qualities and was suitable for measuring variables.

The importance of the relationship between variables was observed by examining the t-values shown in Figure 2. Once a satisfactory model had been obtained, the subsequent stage involved assessing the parameters' statistical significance by utilizing the t-statistical test. If the t-values are red, they are not significant, whereas if there are no red t-values, they are significant.

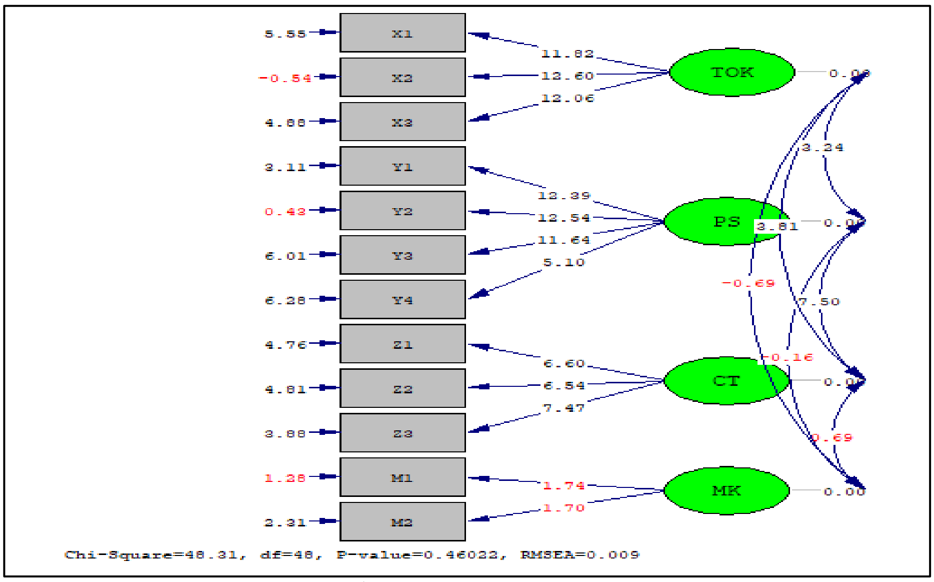

Figure 2: *Output t-Values.*

Note: TOK: Transfer of knowledge; PS: Problem solving; CT: Critical thinking; MK: Metacognition

Figure 3 shows an item correlation that is not significant, namely metacognition items. The process of path testing involves the comparison of the calculated t value with the critical t value. At a significance level of 0.05 (5% significance level), a crucial t-value of 1.97 is found. The study results indicated that the t-value exceeded 1.97 for all items except those related to metacognition. Subsequently, a comprehensive examination was conducted to assess the impact of various factors on indicators related to information transfer, problem-solving, critical thinking, and metacognition for each item. These results as generated from the analysis using LISREL are depicted in Figure 3.

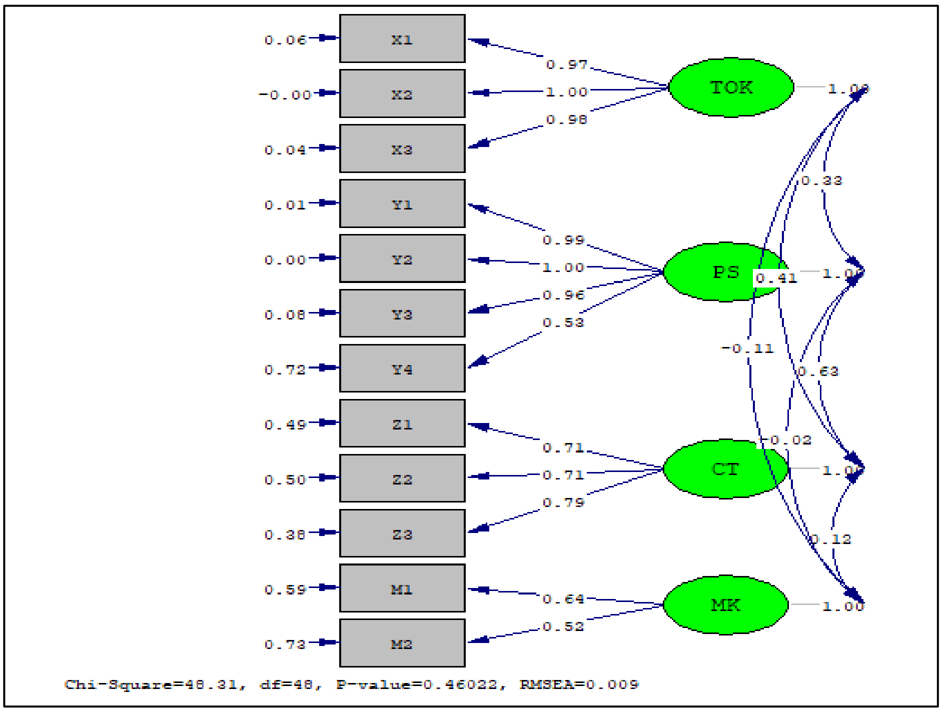

Figure 3: *Standardized Solution Output Path Diagram.*
Note: TOK: Transfer of knowledge; PS: Problem solving; CT: Critical thinking; MK: Metacognition

After establishing the adequacy of the model, a subsequent study was conducted to assess the construct's validity. The construct validity test was carried out by considering each indicator's standard loading factor (LF) value. After the LF standards were set, items with a minimum of 0.5 were declared valid (Retnawati, 2016). The analytical findings indicate that the standard LF value satisfies the necessary conditions to be deemed legitimate. Therefore, the four indicators that have been created were deemed valid for describing the model. The convergent validity test in this study yielded satisfactory findings, as the AVE exceeded the minimum threshold of 0.5. In addition to meeting construct validity, the fit test/goodness of fit model satisfies all the requirements outlined in Table 3. It may be inferred that all items possessed validity in gauging indicators.

The results of the analysis of item difficulty and a person's abilities using the Rasch model are shown in Figure 4. The separation of items shows the grouping of questions, while the separation of persons shows the grouping of student abilities. To the left of the vertical line is the distribution of student abilities, and to the right is the distribution of the difficulty levels of the questions. For example, the tested instrument produced a person's separation value 2.30. This value can be rounded to 3, which means that the instrument can classify students' transfer of knowledge, problem-solving, critical thinking and metacognitive into three levels: high ability, medium ability, and low ability.

Figure 4 also shows that the transfer of knowledge, problem-solving, critical thinking and student metacognition varies but is dominated by moderate abilities, which are grouped in the middle, from 0 logs to 2 logs. Thus, students with high abilities are at the top, while students with low abilities are at the bottom.

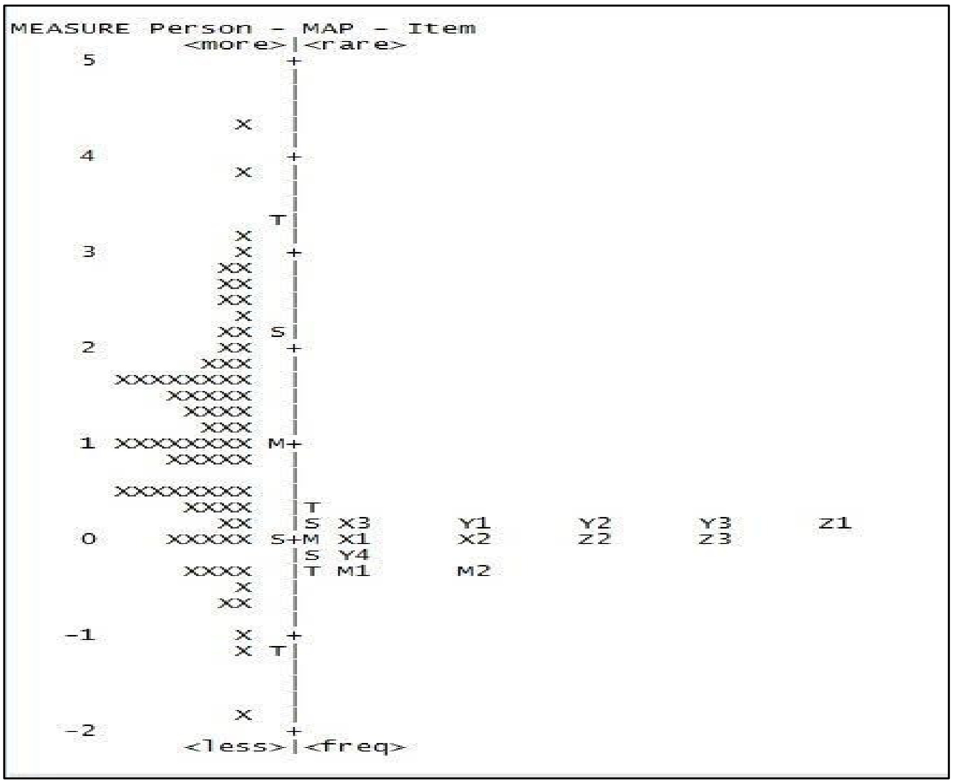

Figure 4: *Logit Distribution for Person and Item Difficulty Levels.*

The grain separation value generated through Rasch modelling was measured 2.00. This value suggested that the instrument consisted of 12 items with a high and low level of difficulty. The most accessible item questions were at the bottom of the logit, and the higher they were, the greater was the difficulty level. Furthermore, the analysis of instrument fit items is shown in Table 4.

Table 4

Item Fit Statistics.

Item	Difficulty Value	S.E.	Infit		Outfit		Pt. Mean Corr.
			MNSQ	ZSTD	MNSQ	ZSTD	
Y3	0.21	0.16	0.67	0.52	0.66	0.33	0.78
Z1	0.21	0.16	0.94	-0.65	0.95	-0.38	0.63
Y2	0.13	0.16	0.68	-0.99	0.67	-1.34	0.80
X3	0.10	0.16	0.71	-1.26	0.67	-1.55	0.66
Y1	0.10	0.16	0.69	-0.79	0.68	-1.13	0.79
X1	0.08	0.16	0.76	-1.33	0.73	-1.67	0.64
X2	0.05	0.16	0.77	-1.33	0.72	-1.67	0.65
Z2	-0.06	0.16	1.03	-0.14	1.06	0.14	0.62
Z3	-0.06	0.16	0.92	-0.38	0.93	0.02	0.73
Y4	-0.24	0.16	0.97	-0.50	0.99	-0.38	0.64
M1	-0.26	0.16	1.83	2.83	1.73	2.64	0.20
M2	-0.26	0.16	1.95	3.85	1.99	3.74	0.16

The analysis results show that question M2 had the highest MNSQ score of 1.95, and item Y3 had the lowest MNSQ score of 0.67. Both are closest to the specified maximum limit (-2 < MNSQ < 2). However, all the items were not problematic for most students. The fit analysis on the MNSQ infit and outfit found that the Y3 question was relatively the easiest and could be done by all students with low and high abilities. This shows that this item can distinguish students with high or low abilities. All questions in the easy and medium categories can be used. Item M2 has an MNSQ score for both infit and outfit closest to 2.0, indicating that this question is only considered easy for those with high abilities and too tricky for those with low abilities. Item Y4, as the most challenging question, had an infit value, and the MNSQ outfit was close to 1, still within the allowable range.

The subsequent phase was conducting a dependability assessment, which was done to assess the degree to which the measurement model accurately measured the specified latent component. The reliability evaluation employed the Cronbach Alpha measuring model and Rasch analysis. The reliability test results obtained a reliability coefficient of 0.84 ≥ 0.7, declared reliable. Table 5 displays the inclusion of CR and AVE in the reliability evaluation.

Table 5

Reliability Analysis Results.

	Score	Description
∑LF	9.79	
(∑LF)2	95.84	
∑LF2	8.39	
∑(1-LF2)	3.61	
CR	0.96	Reliable
AVE	0.70	Reliable

Table 5 presents the reliability test outcomes, indicating that the CR values are more than or equal to 0.7, hence confirming their dependability. Similarly, the AVE findings are equal to or greater than 0.5, satisfying the established reliability standards. The AVE is computed by dividing the root mean normalized loading factor by the number of indicators. The AVE can serve as evidence for the capacity of latent variable values to represent the original data's scores accurately. The higher the AVE value, the better the capacity to elucidate the value of the hidden variable measurement indicator. The commonly employed threshold for the AVE is 0.50, with the least acceptable AVE value also being 0.50. According to Sürücü and Maslakci (2020), a variable's indicator is considered trustworthy if its AVE value is equal to or more than 0.5 and its CR is equal to or greater than 0.7.

The study's findings indicate that the convergent validity is strong, implying that the likelihood of an indicator from one construct influencing another variable is 0.50. Consequently, the chance of an indicator converging and influencing a construct, where the value inside the block is 50%, is established. This finding demonstrates that the instrument exhibits a notable consistency over several usage instances, even when administered to diverse participants.

5. Discussion

The outcome of this research make evident that a HOTs assessment model can be developed for aspiring science teacher candidates, explicitly tailored for educational materials about virus concerns, focusing on its applicability. Cantos et al. (2015) argue that evaluation should accurately capture students' comprehensive abilities, including knowledge, attitudes, and skills, while also catalyzing for students to maximize their potential (Hidayati, Kusmanto, & Kiswantoro, 2023). In order to effectively assess students' talents and competencies, the test instrument must be carefully designed to satisfy specific requirements (Amelia & Kriswantoro, 2017; Hidayati et al., 2023). The development of this instrument model was driven by a requirements analysis, specifically addressing two key objectives: (1) the creation of an instrument model capable of measuring HOTs and (2) the development of an assessment model that effectively identifies the learning deficiencies of prospective science teacher students.

The developmental phase involves the comprehensive examination and integration of HOTs markers. HOTs refer to (Anderson & Krathwohl, 2001; Hassan et al., 2018; Singh et al., 2018) so that 12 indicators are obtained, namely analyzing, (2) evaluating, (3) creating, (4) identifying problems, (5) analyzing information, (6) planning problem solutions, (7) making problem solutions in the form of ideas/ conclusions/products, written/oral works, (8) formulating problems, (9) formulating hypotheses, (10) explaining arguments, (11) assessing one's abilities, (12) predicting other possible ways. Each indicator is derived by considering the content and instructional procedures implemented within the educational material.

During its deployment phase, the HOTs assessment model underwent a comprehensive evaluation process, including qualitative analysis, validation, testing, and measurement. As previously elucidated, the model testing process was conducted by domain experts and involved empirical validation through a series of experiments. The preliminary evaluation conducted by the expert in modeling indicated that the produced model exhibited the requisite HOTs components. The validity of an instrument is determined by the expert's perception of its ability to accurately measure the intended variables (Connell et al., 2018; Högberg, Hamari, & Wästlund, 2019; Moyano-Fuentes, Bruque-Cámara, & Maqueira-Marín, 2019). Furthermore, the validity and acceptance of the 12 components in the development model will be contingent upon expert feedback and ideas. The feedback provided by the experts revised the comments and ideas. Once the model has undergone repairs, it becomes viable for validation, as shown by a minimum Aiken index of 0.8.

Consequently, it may be employed in the subsequent phase, which involves empirical testing via trials. The present study designated the findings as Revision I of the HOTs Assessment Model, obtained through field experiments conducted at three institutions. The study findings indicate that the 12 produced items exhibited loading factor values. The metrics employed in the study were the Root Mean Square Error of Approximation (RMSEA) and the Goodness of Fit Index (GFI) (Suksanchananun, Juicharoen, & Wangthong, 2023). The Goodness of Fit criteria has been met since the acquired values fall within the specified intervals. Consequently, the model may be considered fitting, indicating that the instrument build is satisfactory and can be utilized in terms of the distribution of this measurement model.

The focus of validation is on both internal validation, which involves confirming the components and processes of the model, and external validation, which involves validating the impact of utilizing the model. The validation process entails critically evaluating and assessing a subject matter by individuals possessing specialized knowledge and expertise. Expert review is a systematic procedure in which individuals with specialized knowledge and expertise evaluate many aspects, including components, structure, and potential future applications. The review process and evaluation input rely on pre-established criteria. The subsequent iterations of the model are developed by incorporating input data and recommendations provided by domain experts.

The validation technique in question might be referred to as formative evaluation, as discussed by Elwy et al. (2020), van Groen and Eggen (2019). The Delphi approach is employed as a means of conducting a validation process wherein experts engage in the critical assessment and evaluation of various components and structures within the overarching model that has been constructed (Amelia & Kriswantoro, 2017). The study found two features of the Delphi approach to be of great value. The efficacy of this approach can be attributed to the qualities of the evaluation. The panel of judges possesses a diverse range of experience across several domains. Professionals play a crucial role in the internal validation process. The expert reviewers were allocated one week to assess each cycle thoroughly. During the initial cycle, a focus group discussion (FGD) was conducted, which resulted in substantial improvements to the model.

Additionally, the reviewers were required to address open-ended questions as part of their evaluation process. This feature enables each specialist to evaluate and provide feedback within a customizable timeframe. The evaluation and recommendations others provide are crucial for consideration throughout the subsequent modification process. Further studies refer to empirical investigations that provide a more comprehensive description of the processes involved in building or refining. This study elucidates the methodology and framework employed in creating the created model. This strategy facilitates the development of new models by emphasizing various characteristics, ensuring sufficient clarity. Furthermore, the model's outcomes may be readily comprehended and utilized by users, educators, and learners. This study presents a model that has demonstrated content validity and may be utilized by educators who cannot be dissociated from the contextual factors, instructional material, and student characteristics.

Enhancing cognitive abilities include more than just comprehending concepts; it also entails the development of critical thinking skills and the ability to effectively address real-world issues by evaluating information and arguments within social contexts, as well as making informed life choices (Bailin, 1987). The deficiency in critical and creative thinking abilities among students may be attributed to challenges encountered in the process of selecting, creating, and establishing connections between experimental data that can serve as substantiating evidence for arguments (Akmam et al., 2019; Mahanal et al., 2019; Supena, Darmuki, & Hariyadi, 2021). Several methods may be employed to enhance HOTs through instruction and evaluation. Ideally, the assessment activities conducted should align with the instructional methodology employed. The HOTs evaluation would be suitable when the learning strategy incorporates these talents. Nevertheless, if the educational process fails to foster these cognitive skills, it should not be assumed that critical and creative thinking evaluations are unattainable.

Student's cognitive abilities are significantly impacted by the possibilities afforded to them (Runco & Basadur, 1993). Lecturers accustomed to assigning unusual tasks to their students, hence fostering the development of their critical thinking abilities, are more likely to enhance their student's cognitive capabilities than lecturers who provide regular activities. The lecturer's facilitation of students' thinking skills should align with the lecturer's thinking talents. Educators are exemplars in cultivating pupils who can think critically and exhibit creativity. The development of critical thinking skills in pupils can be facilitated by instructors who possess the ability to think critically (Shutaleva, 2023; Taddeo & Tirocchi, 2021).

The framework of socio-scientific problem-based learning offers students a contextualized approach to comprehending scientific knowledge's socially and culturally constructed nature, achieved via meticulously designed learning experiences (Karisan & Zeidler, 2017). In addition to enhancing learning patterns across various domains, fostering a paradigm shift among educators, enhancing learning patterns within institutions responsible for producing educational staff, and cultivating a supportive family and surrounding environment is imperative.

6. Conclusion

The instrument model devised in this study has commendable attributes and demonstrates a high-quality level, rendering it a valuable tool for assessing and meeting the criteria for evaluating students' skills. The assertion is supported by the findings of data analysis, which demonstrate that the HOTs instrument model had successfully satisfied the criteria of expert judgment, construct validity, and content validity. Additionally, the model has been deemed trustworthy by utilizing confirmatory factor analysis on data obtained from field trials. The instrument model had strong validity and reliability as it regularly draws upon established theories and indications from professionals in the field. This study had inherent constraints primarily constrained by the requirement for a more focused selection of materials.

Based on the available evidence and analysis, it is recommended that further research be conducted in order. First, and foremost, educators must possess the necessary competence and knowledge to employ an appropriate assessment model tailored to their students effectively. It is essential to propagate the additional benefit of this instrument, that of its incorporation of socio-scientific problem-based learning, which encourages students to consider interdisciplinary subjects, namely those about science, technology, and social studies. It is also required that this integration should facilitate the development of HOTs in students, fostering a multi-disciplinary and integrative approach to scientific problem-solving and the cultivation of moral values. The instrument model that has been constructed demonstrates the ability to activate higher-order thinking skills (HOTs), therefore making it a suggested tool for use in learning evaluation.

One limitation of this instrument is that the acquisition of knowledge necessitates the utilization of problem-based learning approaches. If this action is executed, then the utilization of this tool becomes more convenient. The current instrument model is also constrained to a singular viral material, necessitating further modification to accommodate diverse materials conducive to study and learning for other scientists. This aspect is essential and necessitates consistent implementation to amass a diverse range of assets to create a financial instrument. The creation of instruments aimed at measuring skills that align with the demands of the 21st century has significant value for educators and students while also making a valuable contribution to the advancement of educational science and learning.

References

Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving Higher Order Thinking Skills (HOTS) problems for the topic of fraction. *Asian Social Science, 11*(21), 133. http://.dx.doi.org/10.5539/ass.v11n21p133

Abremski, D., & Roben, P. (2021). UC San Diego, The Military and Building a Unique, Diversified Economic Growth Ecosystem. Journal of Commercial Biotechnology, 26(1). doi: https://doi.org/10.5912/jcb974

Afandi, A., Sajidan, S., Akhyar, M., & Suryani, N. (2019). Development frameworks of the Indonesian partnership 21st-century skills standards for prospective science teachers: A Delphi Study. *Jurnal Pendidikan IPA Indonesia, 8*(1), 89-100. https://doi.org/10.15294/jpii.v8i1.11647

Akmam, A., Anshari, R., Jalinus, N., & Amran, A. (2019). Factors influencing the critical and creative thinking skills of college students in computational physics courses. *Journal of Physics: Conference Series. 1317*(1) (pp. 012172). IOP Publishing. https://doi.org/10.1088/1742-6596/1317/1/012172

Alkhatib, O. J. (2019). A framework for implementing higher-order thinking skills (problem-solving, critical thinking, creative thinking, and decision-making) in engineering & humanities. *2019 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1-8). IEEE. https://doi.org/10.1109/ICASET.2019.8714232

Almerich, G., Suárez-Rodríguez, J., Díaz-García, I., & Cebrián-Cifuentes, S. (2020). 21st-century competences: The relation of ICT competences with higher-order thinking capacities and teamwork competences in university students. *Journal of Computer Assisted Learning, 36*(4), 468-479. https://doi.org/10.1111/jcal.12413

Amalia, S., & Wuryandani, W. (2020). Socio-cultural based learning module for critical thinking ability in elementary school: Systematic search. *Acta Educationis Generalis, 10*(2), 180-187. https://doi.org/10.2478/atd-2020-0017

Amelia, R. N., & Kriswantoro, K. (2017). Implementation of Item Response Theory for Analysis of Test Items Quality and Students' Ability in Chemistry. *JKPK (Jurnal Kimia dan Pendidikan Kimia), 2*(1), 1-12. https://doi.org/10.20961/jkpk.v2i1.8512

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*, 1-32. https://doi.org/10.1023/A:1026471702353

Arends, R. I. (2012). *Learning to teach*. McGraw-Hill. https://hasanahummi.files.wordpress.com/2017/04/connect-learn-succeed-richard-arends-learning-to-teach-mcgraw-hill-2012.pdf

Astuti, A. P., Aziz, A., Sumarti, S. S., & Bharati, D. A. L. (2019). Preparing 21st century teachers: Implementation of 4C character's pre-service teacher through teaching practice. *Journal of Physics: Conference Series. 1233*(1) (pp. 012109). IOP Publishing. https://doi.org/10.1088/1742-6596/1233/1/012109

Bailin, S. (1987). Critical and creative thinking. *Informal logic, 9* (1), 23-30. https://doi.org/10.22329/il.v9i1.2656

Bayat, S., & Tarmizi, R. A. (2012). Effects of problem-based learning approach on cognitive variables of university students. *Procedia-Social and Behavioral Sciences, 46*, 3146-3151. https://doi.org/10.1016/j.sbspro.2012.06.027

Bilyalova, A., Salimova, D., & Zelenina, T. (2020). Digital transformation in education. *Integrated Science in Digital Age: ICIS 2019. 78* (pp. 265-276). Springer. https://doi.org/10.1007/978-3-030-22493-6_24

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd. https://www.ascd.org/books/how-to-assess-higher-order-thinking-skills-in-your-classroom?variant=109111

Budi, G. S., & Farcis, F. (2021). Students' Critical Thinking Skills in Innovating Problem Solving in the Physics Entrepreneurship Course. *Berkala Ilmiah Pendidikan Fisika, 9*(1), 39-46. https://doi.org/10.20527/bipf.v9i1.9517

Budiyono, A., & Rohrlich, D. (2017). Quantum mechanics as classical statistical mechanics with an ontic extension and an epistemic restriction. *Nature communications, 8*(1), 1306. https://doi.org/10.1038/s41467-017-01375-w

Budsankom, P., Sawangboon, T., Damrongpanit, S., & Chuensirimongkol, J. (2015). Factors Affecting Higher Order Thinking Skills of Students: A Meta-Analytic Structural Equation Modeling Study. *Educational Research and Reviews, 10*(19), 2639-2652. https://doi.org/10.5897/ERR2015. 2371

Cantos, A. E., Alday, M., Alog, K. J. A., Asi, K. J. G., Calacal, R. H. U., & Britiller, M. C. (2015). Changing learning needs of student nurses: input to the nursing curriculum. *Asia Pacific Journal of Multidisciplinary Research, 3*(3), 108-119. https://api.semanticscholar.org/CorpusID:44042760

Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of life research, 27*, 1893-1902. https://doi.org/10.1007/s11136-018-1847-y

Damaianti, V. S., Abidin, Y., & Rahma, R. (2020). Higher order thinking skills-based reading literacy assessment instrument: An Indonesian context. *Indonesian Journal of Applied Linguistics, 10*(2), 513-525. https://doi.org/10.17509/ijal.v10i2.28600

Elwy, A. R., Wasan, A. D., Gillman, A. G., Johnston, K. L., Dodds, N., McFarland, C., & Greco, C. M. (2020). Using formative evaluation methods to improve clinical implementation efforts: Description and an example. *Psychiatry research, 283*, 112532. https://doi.org/10.1016/j.psychres.2019.112532

Fong, C. J., Kim, Y., Davis, C. W., Hoang, T., & Kim, Y. W. (2017). A meta-analysis on critical thinking and community college student achievement. *Thinking Skills and Creativity, 26*, 71-83. https://doi.org/10.1016/j.tsc.2017.06.002

Fryer, L. K., & Bovee, H. N. (2018). Staying motivated to e-learn: Person-and variable-centred perspectives on the longitudinal risks and support. *Computers & education, 120*, 227-240. https://doi.org/10.1016/j.compedu.2018.01.006

Gallardo, J., Cruz, N. D., Opinga, G., & Albino, F. S. (2023). Visual Learning-Style-Based Activities in Improving the Reading Comprehension Skills of Second-Year English Major Students. *International Journal of Multidisciplinary: Applied Business and Education Research, 4*(10), 3530-3536. https://doi.org/10.11594/ijmaber.04.10.09

Ganiron Jr, T. U. (2014). The Effect of Study Group Activity Guide in Expository Reading and Writing Course at the College of Architecture in Qassim University. *International Journal of Education and Learning, 3* (1), 23-34. http://dx.doi.org/10.14257/ijel.2014.3.1.03

Garcia-Martin, J., & Garcia-Sanchez, J.-N. (2017). Pre-service teachers' perceptions of the competence dimensions of digital literacy and of psychological and educational measures. *Computers & education, 107*, 54-67. https://doi.org/10.1016/j.compedu.2016.12.010

Ghobadi, S., & Taki, S. (2018). Effects of Telegram stickers on English vocabulary learning: Focus on Iranian EFL learners. *Research in English language pedagogy, 6*(1), 139-158. https://doi.org/10.30486/relp.2018.538824

Halili, S. H., Razak, R. A., & Zainuddin, Z. (2015). Exploring the use of 'Wiggio'to support online collaborative learning for adult learners. *Economics, Social Sciences and Information Management: Proceedings of the 2015 International Congress on Economics, Social Sciences and Information Management (ICESSIM 2015)* (pp. 15-21). https://doi.org/10.1201/b19921-5

Heong, Y. M., Sern, L. C., Kiong, T. T., & Mohamad, M. M. B. (2016). The role of higher order thinking skills in green skill development. *MATEC Web of Conferences. 70* (pp. 05001). EDP Sciences. https://doi.org/10.1051/matecconf/20167005001

Hidayati, R., Kusmanto, A. S., & Kiswantoro, A. (2023). Development and Construct Validation of Indonesian Students Self-Confidance Scale Using Pearson Product Moment. *Pegem Journal of Education and Instruction, 13*(3), 94-103. https://doi.org/10.47750/pegegog.13.03.11

Högberg, J., Hamari, J., & Wästlund, E. (2019). Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction, 29*(3), 619-660. https://doi.org/10.1007/s11257-019-09223-w

Jeffrey, L. M., Milne, J., Suddaby, G., & Higgins, A. (2014). Blended learning: How teachers balance the blend of online and classroom components. *Journal of Information Technology Education, 13*, 121-140. https://doi.org/10.28945/1968

Karisan, D., & Zeidler, D. L. (2017). Contextualization of Nature of Science within the Socioscientific Issues Framework: A Review of Research. *International Journal of Education in Mathematics, Science and Technology, 5*(2), 139-152. https://doi.org/10.18404/ijemst.270186

Kembara, M. D., Rozak, R. W. A., & Hadian, V. A. (2019). Based Lectures to Improve Students' 4C (Communication, Collaboration, Critical Thinking, and Creativity) Skills. *International Symposium on Social Sciences, Education, and Humanities (ISSEH 2018). 306* (pp. 22-26). Atlantis Press. https://doi.org/10.2991/isseh-18.2019.6

Kim, S. S., Kim, E. J., Lim, J. Y., Kim, G. M., & Baek, H. C. (2018). Korean nursing students' acquisition of evidence-based practice and critical thinking skills. *Journal of Nursing Education, 57*(1), 21-27. https://doi.org/10.3928/01484834-20180102-05

Koç, F. Ş., & Ölmez-Çağlar, F. (2023). Alternative Assessment in ELT: Conceptualization, Implementation, and Further Perspectives. In *Global Perspectives on Effective Assessment in English Language Teaching* (pp. 235-254). IGI Global. https://doi.org/10.4018/978-1-6684-8213-1.ch008

Ku, K. Y., & Au, R. H. Y. (2021). Exploring University Students' Perceived Values of Knowledge and Critical Thinking Using Triple-Response Mode Assessment. *Anatolian Journal of Education, 6*(2), 1-12. https://doi.org/10.29333/aje.2021.621a

Langdon, C., Kurz, C., & Coppola, M. (2023). The importance of early number concepts for learning mathematics in deaf and hard of hearing children. *Perspectives on Early Childhood Psychology and Education, 5*(2), 6. https://doi.org/10.58948/2834-8257.1061

Lee, J., Lim, C., & Kim, H. (2017). Development of an instructional design model for flipped learning in higher education. *Educational Technology Research and Development, 65*, 427-453. https://doi.org/10.1007/s11423-016-9502-1

Liu, C.-Y., Wu, C.-J., Wong, W.-K., Lien, Y.-W., & Chao, T.-K. (2017). Scientific modeling with mobile devices in high school physics labs. *Computers & Education, 105*, 44-56. https://doi.org/10.1016/j.compedu.2016.11.004

Magsino, R. M. (2014). Enhancing higher order thinking skills in a marine biology class through Problem-Based Learning. *Asia Pacific Journal of Multidisciplinary Research, 2*(5), 1-6. https://www.apjmr.com/wp-content/uploads/2014/09/APJMR-2014- 2-116.pdf

Mahanal, S., Zubaidah, S., Sumiati, I. D., Sari, T. M., & Ismirawati, N. (2019). RICOSRE: A Learning Model to Develop Critical Thinking Skills for Students with Different Academic Abilities. *International Journal of Instruction, 12* (2), 417-434. https://doi.org/10.29333/iji.2019.12227a

Makmuri, M., & Aziz, T. A. (2021). Exploring mathematics teachers perspectives on the effectiveness of blended learning for professional development program. *AIPConference Proceedings. 2331*(1) (pp. 020008). AIP Publishing. https://doi.org/10.1063/5.0041693

Moyano-Fuentes, J., Bruque-Cámara, S., & Maqueira-Marín, J. M. (2019). Development and validation of a lean supply chain management measurement instrument. *Production Planning & Control, 30*(1), 20-32. https://doi.org/10.1080/09537287.2018.1519731

Mtetwa, D. K., & Ndemo, Z. (2022). An integrated model for teacher continuing professional learning in Zimbabwean primary schools. *African Journal of Teacher Education and Development, 1*(1), 13. https://doi.org/10.4102/ajoted.v1i1.9

Mulyono, Y., & Yamtinah, S. (2023). Development of Critical and Creative Thinking Skills Instruments Based on Environmental Socio-Scientific Issues. *International Journal of Instruction, 16*(3), 691. https://doi.org/10.29333/iji.2023.16337a

Noviana, A., AbdurRahman, A., Rosidin, U., & Herlina, K. (2019). Development and validation of collaboration and communication skills assessment instruments based on project-based learning. *Journal of Gifted Education and Creativity, 6*(2), 133-146. https://doi.org/10.2991/assehr.k.210423.082

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning, 18*, 103-120. https://doi.org/10.1007/s10758-013-9202-4

Prayitno, B. A., & Titikusumawati, E. (2018). Enhancing Students' Higher Order Thinking Skills in Science through INSTAD Strategy. *Journal of Baltic Science Education, 17*(6), 1046-1055. https://doi.org/10.33225/jbse/18.17.1046

Prayitno, S. H., & Jaedun, A. (2022). Technical education teachers' perception on higher order thinking skills and their ability to implementation it in Indonesia. *Jurnal Pendidikan Vokasi, 12*(3), 245-256. https://doi.org/10.21831/jpv.v12i3.54335

Rahayu, S., & Rosawati, E. E. (2023). The development of higher-order thinking skills (HOTS) assessment instrument in chemistry using socioscientific issues context: A preliminary trial. *AIP Conference Proceedings. 2572*(1) (pp. 030009). AIP Publishing. https://doi.org/10.1063/5.0118624

Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *REiD (Research and Evaluation in Education), 2*(2), 155-164. https://doi.org/10.21831/reid.v2i2.11029

Rismi, O. D. (2021). A Learning Design to Improve Higher Order Thinking Skills (HOTS). *Jurnal Riset Pendidikan Matematika Jakarta, 3*(2), 34-41. https://doi.org/10.21009/jrpmj.v3i2.22265

Romine, W. L., Sadler, T. D.,& Kinslow, A. T. (2017). Assessmentof scientificliteracy: Developmentand validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR). *Journal of Research in Science Teaching, 54*(2), 274-295. https://doi.org/10.1002/tea.21368

Rubbo, L., & Moore, J. (2012). Are We Teaching Students to Think Like Scientists? *Connecting People to Science: A National Conference on Science Education and Public Outreach. 457* (pp. 343). https://www.researchgate.net/publication/48180080

Runco, M. A., & Basadur, M. (1993). Assessing ideational and evaluative skills and creative styles and attitudes. *Creativity and Innovation Management, 2* (3), 166-173. https://doi.org/10.1111/j.1467-8691.1993.tb00088.x

Rustad, M. L. (2013). Twenty-First-Century Tort Theories: The Internalist/Externalist Debate. *Indiana Law Journal, 88*(2), 419. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2246494

Saido, G. M., Siraj, S., Nordin, A. B. B., & Al_Amedy, O. S. (2018). Higherorderthinking skills among secondary school students in science learning. *MOJES: Malaysian Online Journal of Educational Sciences, 3*(3), 13-20. https://www.researchgate.net/publication/321423293

Salleh, S. M., Tasir, Z., & Shukor, N. A. (2012). Web-based simulation learning framework to enhance students' critical thinking skills. *Procedia-Social and Behavioral Sciences, 64*, 372-381. https://doi.org/10.1016/j.sbspro.2012.11.044

Şendağ, S., & Odabaşı, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & education, 53*(1), 132-141. https://doi.org/10.1016/j.compedu.2009.01.008

Shanta, S., & Wells, J. G. (2022). T/E design based learning: assessing student critical thinking and problem solving abilities. *International Journal of Technology and Design Education, 32*(1), 267-285. https://doi.org/10.1007/s10798-020-09608-8

Shukla, D., & Dungsungnoen, A. P. (2016). Student's Perceived Level and Teachers' Teaching Strategies of Higher Order Thinking Skills: A Study on Higher Educational Institutions in Thailand. *Journal of education and Practice, 7*(12), 211-219. https://api.semanticscholar.org/CorpusID:73564387

Shutaleva, A. (2023). Ecological culture and critical thinking: building of a sustainable future. *Sustainability, 15*(18), 13492. https://doi.org/10.3390/su151813492

Sörbom, D. (1989). Model modification. *Psychometrika, 54* (3), 371-384. https://doi.org/10.1007/BF02294623

Subiantoro, A. W., & Treagust, D. F. (2021). Development and validation of an instrument for assessing high-school students' perceptions of socio-scientific issues-based learning in biology. *Learning Environments Research, 24*, 223-237. https://doi.org/10.1007/s10984-020-09332-z

Suksanchananun, W., Juicharoen, N., & Wangthong, S. (2023). A Confirmation Factor Analysis of the Potential Level Scale for Halal Tourist Sites in Lower Songkhla Lake. *International Journal of Social Service and Research, 3*(6), 1352-1362. https://doi.org/10.46799/ijssr.v3i6.420

Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers& education, 94*, 252-275. https://doi.org/10.1016/j.compedu.2015.11.008

Supena, I., Darmuki, A., & Hariyadi, A. (2021). The Influence of 4C (Constructive, Critical, Creativity, Collaborative) Learning Model on Students' Learning Outcomes. *International Journal of Instruction, 14*(3), 873-892. https://doi.org/10.29333/iji.2021.14351a

Sürücü, L., & Maslakci, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal, 8* (3), 2694-2726. https://doi.org/10.15295/bmij.v8i3.1540

Taddeo, G., & Tirocchi, S. (2021). Transmedia teens: the creative transmedia skills of Italian students. *Information, communication & society, 24* (2), 241-257. https://doi.org/10.1080/1369118X.2019.1645193

Uerz, D., Volman, M., & Kral, M. (2018). Teacher educators' competences in fostering student teachers' proficiency in teaching and learning with technology: An overview of relevant research literature. *Teaching and Teacher Education, 70*, 12-23. https://doi.org/10.1016/j.tate.2017.11.005

Utama, C., & Nurkamto, J. (2020). The Instrument Development to Measure Higher-Order Thinking Skills for Pre-Service Biology Teacher. *International Journal of Instruction, 13*(4), 833-848. https://doi.org/10.29333/iji.2020.13451a

van Groen, M. M., & Eggen, T. J. (2019). Educational test approaches: The suitability of computer-based test types for assessment and evaluation in formative and summative contexts. *Journal of Applied Testing Technology, 21*(1), 12-24. https://www.researchgate.net/publication/343501948

Wan Mohamed Salleh, W. N., Che Ahmad, C. N., Hussain, N., Setyaningsih, E., Shaharim, S. A., & Hashim, R. (2022). Evaluation of a Higher Order Thinking Skills Test for the Topic of Cell Division and Subtopic of Gametogenesis (UKBATG). *Central Asia & the Caucasus (14046091), 23*(1). https://doi.org/10.37178/ca-c.23.1.192

Wei, C.-W., Chen, N.-S., & Kinshuk. (2012). A model for social presence in online classrooms. *Educational Technology Research and Development, 60*, 529-545. https://doi.org/10.1007/s11423-012-9234-9

Yanto, B. E., Subali, B., & Suyanto, S. (2019). Improving Students' Scientific Reasoning Skills through the Three Levels of Inquiry. *International Journal of Instruction, 12*(4), 689-704. https://doi.org/10.29333/iji.2019.12444a

Yuliarti, Y., Marlina, L., Siahaan, S. M., Fathurohman, A., & Sudirman, S. (2023). Profile of High School Students' Critical Thinking Skills about Renewable Energy Materials. *Jurnal Penelitian Pendidikan IPA, 9*(11), 10151-10160. https://doi.org/10.29303/jppipa.v9i11.5418