

## DEVELOPMENT OF HOTS (*HIGHER ORDER THINKING SKILLS*) BASED ASSESSMENT INSTRUMENTS FOR THE SUBJECT OF MARKET ECONOMICS AND THE FORMATION OF MARKET PRICES IN GRADE X AT MAN 2 CITY OF MADIUN

Binti Munafi'ah<sup>1a</sup>, Siswandari<sup>2b</sup>, and Sudarno<sup>3c</sup>

<sup>123</sup>Master of Economics Education, Postgraduate of the Faculty of Teacher Training and Education, Sebelas Maret University Surakarta

[bintimunafiah3@student.uns.ac.id](mailto:bintimunafiah3@student.uns.ac.id)

[siswandari@staff.uns.ac.id](mailto:siswandari@staff.uns.ac.id)

[sudarno251168@gmail.com](mailto:sudarno251168@gmail.com)

(\*) Corresponding Author

[sudarno251168@gmail.com](mailto:sudarno251168@gmail.com)

### ARTICLE HISTORY

**Received** : 07-08-2025

**Revised** : 15-09-2025

**Accepted** : 05-10-2025

### KEYWORDS

Assessment,  
HOTS,  
Economics,  
Instrument,  
Learning

### ABSTRACT

This study aims to describe the development stages and test the feasibility of HOTS-based assessment instruments in the Economics subject: markets and market price formation for class X at MAN 2 Madiun City. The research approach or model used by the researcher is the Thiagarajan model, however, the researcher only developed up to the 3D model (Define, Design, Develop) because it aims to develop educational products (in this case HOTS instruments) whose feasibility can be tested systematically through expert validation. The results of this study are the stages of developing HOTS-based assessment instruments carried out through a 3D model approach (Define, Design, Develop) modified from the 4D development model by Thiagarajan. In the Define stage, an analysis of the curriculum, student characteristics, teaching materials, and specifications of learning objectives is carried out. The Design stage includes the preparation of question grids, writing HOTS-based question items (C4–C6), and creating assessment guidelines. The Develop stage involves expert validation and instrument revision based on input, resulting in an assessment product that is in accordance with the high-level cognitive demands in the Independent Curriculum. The developed HOTS-based assessment instrument was proven to be suitable for use, as indicated by the results of expert validation of the instrument's content. In the first validation test, the average Content Validity Index (CVI) value from the material experts was 0.73 (fairly valid category), and after improvements were made, it increased to 1.00 (very valid) in the second test. The assessment experts also stated that the instrument was suitable for use with minor revisions. This indicates that the instrument has met the content validity standards to be used as a measuring tool for students' higher-order thinking skills in the material "Markets and Market Price Formation".

*This is an open access article under the CC-BY-SA license.*



### INTRODUCTIONS

21st century education demands a learning model that goes beyond transferring knowledge—but also builds character, creativity, and high-level thinking skills (Mashudi, 2021). One of the approaches that answers this challenge is deep learning, which is a pedagogical approach that emphasizes conceptual understanding, reflection, interdisciplinary connections, and the application of knowledge in real contexts (Fullan et al., 2017). In its implementation, deep learning requires an evaluation instrument that can assess more than memorization, namely an assessment that measures higher-order thinking skills (HOTS) (Z. S. Siregar et al., 2023).

HOTS-based assessments are evaluation instruments developed to assess students' analytical, critical, evaluative, and creative thinking skills—namely levels C4 to C6 in the revised Bloom Taxonomy (Anderson, L.W.; Krathwohl, 2001). This assessment assesses how learners are able to use the knowledge gained to examine problems in depth, evaluate solutions, and create new relevant approaches (Ramdani et al., 2019). In the context of deep learning, HOTS-based assessments are important because they support the learning process that not only answers questions correctly, but also develops thinking rooted in deep understanding (Budiarti & Airlanda, 2019).

HOTS assessments in deep learning approaches are used continuously in formative and summative learning (Amelia et al., 2024). In formative learning, assessments are used to provide feedback during the learning process—for example, in class discussions, project reflections, or case-based assignments (Sholeh et al., 2024). Meanwhile, in summative learning, the HOTS assessment can be used to measure the extent to which students are able to apply knowledge in real projects or develop innovative solutions in the final learning portfolio (Ariyana et al., 2019).

The main users of this assessment are educators as learning facilitators, who have the responsibility to design HOTS-based instruments in a contextual manner (Lestari, 2019). In deep learning-based learning, educators not only prepare questions, but also accompany students' thinking processes in exploration, experimentation, and evaluation of learning outcomes (Putri et al., 2021). In addition, schools, principals, and curriculum designers also use the HOTS assessment as an indicator of the effectiveness of in-depth learning, as well as as a material for policy reflection and educator professional development (Wahidin, 2023).

The urgency of developing HOTS assessments in the framework of deep learning is very important to prevent surface learning, which is only oriented towards mastering facts without transformative abilities (Putro, 2024). When learners are only trained to answer multiple-choice questions without being challenged to integrate information and build meaningful understanding, then learning loses its usefulness (Amelia et al., 2024). Therefore, HOTS-based assessments not only serve as a measuring tool, but also as a catalyst for reflective, in-depth, and future-oriented learning (Widiana et al., 2025). In this context, the development of HOTS assessment instruments is an essential need, especially in responding to the complex and dynamic demands of 21st century learning (Kaniawati & Kasmahidayat, 2021). HOTS-based assessment instruments are essential to measure and improve students' high-level thinking skills (Anderson, L.W.; Krathwohl, 2001), especially in the subject of Economics, which requires a deep understanding of the concept and its application (Sholiha & Kurniawan, 2022). The importance of this assessment instrument can improve the quality of learning and provide a clearer picture of students' abilities (Muhibbuddin et al., 2023).

In the study of Siregar et al. (2023), educators at SMAN 2 Percut Sei Tuan had difficulty distinguishing between LOTS and HOTS questions, the questions they had did not touch the real context and were only cognitive tests. Masrina et al. (2023) stated that assessments at MAN 3 North Aceh and SMAN 1 Matangkuli only focused on cognitive aspects, not 21st century skills and the lack of collaboration between educators in the development of HOTS questions. Wahyudi et al. (2023) added that among educators at MAN 1 Semarang, there is still a misconception of the HOTS-based AKM format and students' literacy is still low in HOTS questions. At SMAN 11 Pontianak, class XI IPS students experienced difficulties in solving HOTS Economics problems on the topic of national income, divided into four stages according to Halifah et al. (2024). First, students had difficulty understanding the problem because they were unable to grasp the meaning of the problem and convert it into an equation. Second, they had difficulty creating a solution plan due to choosing the wrong or forgetting the formula. Third, at the stage of implementing the

plan, students often made mistakes in substituting values and applying equations. Fourth, at the stage of re-checking, students were unable to conclude the answer correctly and were less careful in checking the results. Thus, the main obstacles lie in understanding the concept, solution strategies, and accuracy in evaluating answers.

The researcher tried to answer by adjusting the opinions of observation results for class X students on March 1, 2024 and the opinion expressed by one of the Economics educators that madrasas do not yet have a HOTS-based assessment instrument. In fact, in the future, educators want or intend to measure critical thinking skills which are skills needed in the 21st century. In the context of deep learning learning, HOTS assessments are very important to measure students' conceptual understanding, transferability skills, and reflection on real context (Angga & Sari, 2025). This instrument allows learners not only to know what, but also to understand why and how a concept applies in life (Fullan et al., 2017). Without the HOTS assessment, the objectives of deep learning will not be achieved authentically.

The researcher contributes to developing HOTS-based assessment instruments in Economics subjects by compiling questions that measure the ability to analyze, evaluate, and create according to learning outcomes. The instrument was developed based on the revised Bloom taxonomy and validated through expert testing to ensure its feasibility.

## METHOD

This study uses a Research and Development (R&D) design using a 3D model modified from Thiagarajan's 4D framework, namely Define, Design, and Develop, because it is to develop educational products (in this case HOTS instruments) that can be tested for feasibility systematically through expert validation. The initial Define stage involves curriculum analysis, analysis of student characteristics, and analysis of materials related to Market and Price Formation, and specification of learning objectives. The second stage of Design, designing an overview of the learning instrument (Nisfiah, 2019), includes the preparation of instrument grids, creation of test items, and preparation of assessment guidelines. The third stage of Develop, development consisting of initial validation by experts (content validity) and instrument revision and preparation of the final instrument.

## RESULT AND DISCUSSIONS

### Preliminary Study Results

Based on a preliminary study conducted through observation, interviews, and document analysis, an urgent need was identified to develop an assessment instrument based on Higher Order Thinking Skills (HOTS) for 10th-grade Economics at MAN 2 Madiun City. This study identified a significant gap between the demands of the Independent Curriculum and the implementation of assessment and learning strategies in the field.

#### 1. Observation and Interview Results

Observations indicate that assessments used in class are still dominated by Lower Order Thinking Skills (LOTS) questions that require memorization and basic understanding (C1-C3). Learning tends to be passive, dominated by lecture methods, and does not stimulate students to think critically, analyze, or solve real-world economic problems. This finding was reinforced by interviews with educators, who revealed that the main obstacle in designing HOTS questions was the lack of training, references, and systematic assessment tools from the school.

Table 1. Observation Findings

| No. | Observed Aspects          | Field Findings                                                                                                               |
|-----|---------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 1.  | Types of Assessments Used | The questions presented are predominantly LOTS (C1–C3) cognitive levels, not yet reaching HOTS (C4–C6) levels.               |
| 2.  | Question Characteristics  | They are primarily based on memorization and basic understanding, without contextualization to real-world economic problems. |

| No. | Observed Aspects               | Field Findings                                                                                                                                  |
|-----|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| 3.  | Learning Methods               | They are dominated by lectures and one-way Q&A sessions; with minimal use of participatory or collaborative methods.                            |
| 4.  | Student Activities in Learning | Students are not sufficiently involved in discussions, data exploration, case studies, or project-based assignments.                            |
| 5.  | Indications of Deep Learning   | There is no visible application of in-depth learning components such as reflection, conceptual connections, or contextual meaning.              |
| 6.  | Stimulating Critical Thinking  | There are no activities that stimulate higher-order thinking skills (analysis, evaluation, or solution creation to concrete economic problems). |

## 2. Curriculum Document Analysis

Analysis of the curriculum documents, Learning Objectives Flow (ATP), and Teaching Modules indicates that the school curriculum adheres to the Independent Curriculum, which emphasizes the development of critical thinking competencies. However, the assessment instruments in the Teaching Modules do not consistently reflect HOTS principles. The questions do not fully measure the analytical (C4), evaluation (C5), and creative (C6) skills necessary for a deep understanding of Economics material.

## Product/Instrument Development

The development of HOTS-based assessment instruments for Economics was carried out using the 3D model (Define, Design, Develop), an adaptation of the 4D model. This process aims to create a systematic product that meets the real needs of schools.

### 1. Define Stage

This stage focuses on needs analysis. Activities included: Curriculum Analysis, identified that the Independent Curriculum at MAN 2 Madiun City requires the development of questions that measure higher-order thinking skills (C4-C6). Student Characteristics Analysis, found that 10th-grade students have a good understanding of basic concepts (LOTS), but struggle when faced with contextual and analytical questions (HOTS). This indicates that students' critical thinking skills have not yet developed optimally. Material Analysis, the topic "Markets and the Formation of Market Prices" was chosen because of its significant potential to be developed into HOTS questions relevant to everyday life. Learning Objective Specification, formulate learning objectives aligned with HOTS indicators, including analysis of the concepts of demand, supply, and market mechanisms.

### 2. Design Stage

At this stage, the assessment instrument is designed using the following steps: Development of the Instrument Outline, the outline is created as a guide, ensuring that each question item measures HOTS skills (C4, C5, and C6) using various question formats (multiple choice, essay, etc.). Item Writing, based on the outline, 15 questions are developed that challenge students to analyze, evaluate, and create solutions. Development of Assessment Guidelines, assessment guidelines are created to ensure objectivity, including a scoring rubric for essay questions and an answer key for multiple-choice questions.

### 3. Development Stage

This final stage is the process of testing and refining the instrument: Expert Validation, the instrument was validated by experts (lecturers and Economics teachers) to ensure content validity, alignment with HOTS indicators, and readability. The validation results were processed using the Content Validity Index (CVI). Revision and Finalization, based on input from the validators, revisions were made to the test items. Questions that proved valid and reliable were then finalized into a HOTS assessment instrument ready for use. Overall, this development process resulted in a valid and relevant HOTS-based assessment instrument designed to bridge the gap between curriculum demands and learning practices, and to stimulate students' higher-order thinking skills.

### **Feasibility of HOTS-Based Assessment Instruments**

The feasibility of HOTS-based assessment instruments is a crucial component in the process of developing evaluation tools aimed at measuring students' higher-order thinking skills. In the context of this research, instrument feasibility was tested through two main approaches: expert judgment validation and instrument pilot analysis.

#### **1. Expert Validation (Theoretical and Methodological Foundations)**

Expert validation aims to test the content validity of the developed instruments. The assessment was conducted by three subject matter experts: Dr. Muhammad Sabandi, S.E., M.Si., (Lecturer in the Master's Program in Economics Education), two Economics teachers (Mr. Nurhuda, S.Pd., Mr. Ahmad Baidhowi, S.Pd., M.Pd.), and one assessment expert, Ms. Leny Noviani, S.Pd., M.Si. (Lecturer in Master's Degree in Economics Education), using an assessment instrument covering 10 key aspects, such as conformity to core competencies, cognitive level (C4–C6), stimulus clarity, and contextual relevance. The technique used to analyze this validity was the Content Validity Index (CVI), as referenced in the study by Chasanah et al. (2022), which allows for quantification of the level of agreement among experts regarding the validity of each aspect.

The results of the first validation are attached in Appendix 3 and show that the average CVI score from the subject matter experts was 0.73—still in the "fairly valid" category—indicating a need for improvement, particularly in the aspects of question stimulus and originality of context. Meanwhile, the assessment experts' assessment yielded a CVI of 0.23, noting the need for major revisions to the substance and structure of the questions.

After improvements were made based on qualitative and quantitative input from the experts, the second validation showed significant improvement. All three subject matter experts gave full marks on all aspects, with a CVI score of 1.00, indicating optimal content validity. The assessment expert also stated that all aspects of the instrument were valid (score  $\geq 3$ ) and suggested minor revisions. These changes indicate that the revised instrument has successfully improved the quality of the questions, both in terms of substance and technique.

#### **2. Instrument Trial: Limited Empirical Validity**

Although the instrument demonstrated very high content validity, empirical reliability and validity testing on students was not comprehensively conducted in this study. This was due to time and scope limitations, which focused solely on the content development and validation stages. Therefore, testing aspects such as difficulty level, discriminatory power, and reliability coefficients (e.g., using Cronbach's Alpha) remains an important agenda for further research.

Considering the above findings, it can be concluded that the HOTS-based assessment instrument developed in this study meets the theoretical feasibility criteria, particularly in terms of content validity. This instrument is suitable for use in the context of Economics learning based on the Independent Curriculum, specifically on the topic "Markets and the Formation of Market Prices." However, further implementation still requires thorough empirical testing to ensure the validity and reliability of the instrument in a real field context.

### **Expert Validation Results and Instrument Trial**

#### **1. Expert Validation Results**

Instrument validation was conducted to assess the feasibility of the HOTS-based assessment questions developed for the 10th-grade Economics subject on the topic of Markets and Market Price Formation. A total of 15 questions were developed, consisting of 10 multiple-choice questions, 2 matching questions, 2 short answer questions, and 1 essay question. Based on the expert validation results for the HOTS-based assessment instrument for the 10th-grade Economics subject on the topic of Markets and Market Price Formation, two validation tests were conducted:

before and after revision. These assessments were conducted by three experts: Mr. Nurhuda, S.Pd., Mr. Ahmad Baidhowi, S.Pd., M.Pd., and Dr. Muhammad Sabandi, S.E., M.Si., with the instrument covering 10 assessment aspects. The assessment results were processed using the Content Validity Index (CVI), which measures the proportion of expert agreement in assessing each aspect as valid. Polit et al. (2005) stated that an instrument item can be considered to have good content validity if it obtains an Item-Level Content Validity Index (I-CVI) value of  $\geq 0.78$ , especially when the number of validators is at least three people. This value indicates expert agreement that the instrument item is relevant to the construct being measured. Thus, items that obtain a value below this criterion should be revised or eliminated because they are considered less representative of the intended construct. In line with this, the results of research by Zahra & Subekti (2025) also show that an I-CVI value  $\geq 0.78$  is considered adequate when the number of validators ranges from three to five people, while the average scale value (S-CVI/Ave) of 0.91 is categorized as very high and indicates that the instrument has very good content validity. Thus, both authors agree that instruments with an I-CVI value above 0.78 can be categorized as sufficient or good, and an S-CVI/Ave value approaching or exceeding 0.90 indicates very good content validity for use in research and educational practice.

In the first test, the results from the three experts showed that some aspects still needed improvement. The score details are as follows. The 1st expert, Mr. Nurhuda, S.Pd., assessed 8 out of 10 aspects with a score  $\geq 3$ , so the CVI result = 0.70. The 2nd expert, Mr. Ahmad Baidhowi, S.Pd., M.Pd., assessed 6 out of 10 aspects with a score  $\geq 3$ , so the CVI result = 0.60. The 3rd expert, Mr. Dr. Muhammad Sabandi, S.E., M.Si., assessed 8 out of 10 aspects with a score  $\geq 3$ , so the CVI result = 0.80. The average CVI in the first stage is as follows.

Table 2. Aspects of Assessment from Experts

| No.                                                 | Aspects of Assessment                | Expert 1    | Expert 2    | Expert 3    |
|-----------------------------------------------------|--------------------------------------|-------------|-------------|-------------|
| 1                                                   | Compliance with Core Competencies    | 1           | 1           | 1           |
| 2                                                   | Question Construction                | 1           | 0           | 1           |
| 3                                                   | Stimulus Clarity                     | 0           | 0           | 1           |
| 4                                                   | Cognitive Level (C4–C6)              | 1           | 1           | 1           |
| 5                                                   | Contextual Integration               | 1           | 1           | 1           |
| 6                                                   | Clarity of Answer Alternatives       | 1           | 0           | 1           |
| 7                                                   | Clarity of Scoring Guidelines        | 1           | 0           | 1           |
| 8                                                   | Content Validity                     | 1           | 1           | 1           |
| 9                                                   | Ability to Measure Critical Thinking | 1           | 1           | 0           |
| 10                                                  | Ability to Measure Creativity        | 0           | 1           | 0           |
| <b>Number of Valid Scores (<math>\geq 3</math>)</b> |                                      | <b>8</b>    | <b>6</b>    | <b>8</b>    |
| <b>CVI for Each Expert</b>                          |                                      | <b>0,80</b> | <b>0,60</b> | <b>0,80</b> |

$$CV_i = \frac{0,80 + 0,60 + 0,80}{3} = 0,73$$

The assessment validation was conducted by Dr. Leny Noviani, S.Pd., M.Sc., a lecturer specializing in Economic Education. The assessment was conducted on 10 aspects using a Likert scale of 1–4, with scores of 3 and 4 considered to indicate that the item has met the validity criteria. In the first test, she scored the 10 aspects of the instrument. From the assessment, all aspects received a score of 3 (Good), with none reaching 4 (Very Good). Since all aspects were scored  $\geq 3$ , all were counted as validity indicators.

$$CV_i = \frac{0,30 + 0,30 + 0,10 + 0,20 + 0,20 + 0,20 + 0,20 + 0,30 + 0,20 + 0,30}{1} = 0,23$$

The CVI value was <1, and the expert's conclusion notes indicated that the instrument needed major revision. This was due to qualitative considerations regarding the substance of the questions that required further improvement, such as aspects of stimulus clarity, variation in question format, and the current context that was not optimally depicted. Therefore, these results indicate that quantitative assessment using the CVI needs to be balanced with qualitative assessment to obtain a comprehensive picture.

After the revisions were made, the instrument was revalidated by the same three experts. The assessment results showed significant improvements as follows. The 1st expert, Mr. Nurhuda, S.Pd., gave a score of 3 or 4 on all aspects with a CVI = 1.00. The 2nd expert, Mr. Ahmad Baidhowi, S.Pd., M.Pd., all aspects were scored  $\geq 3$ , so the CVI result = 1.00. The 3rd expert, Mr. Dr. Muhammad Sabandi, S.E., M.Si., all aspects were scored  $\geq 3$ , so the CVI result = 1.00. The average CVI in the second stage is as follows.

Table 3. Aspects of Assesment from Experts

| No. | Aspects of Assessment                               | Expert 1    | Expert 2    | Expert 3    |
|-----|-----------------------------------------------------|-------------|-------------|-------------|
| 1   | Compliance with Core Competencies                   | 1           | 1           | 1           |
| 2   | Question Construction                               | 1           | 1           | 1           |
| 3   | Stimulus Clarity                                    | 1           | 1           | 1           |
| 4   | Cognitive Level (C4–C6)                             | 1           | 1           | 1           |
| 5   | Contextual Integration                              | 1           | 1           | 1           |
| 6   | Clarity of Answer Alternatives                      | 1           | 1           | 1           |
| 7   | Clarity of Scoring Guidelines                       | 1           | 1           | 1           |
| 8   | Content Validity                                    | 1           | 1           | 1           |
| 9   | Ability to Measure Critical Thinking                | 1           | 1           | 1           |
| 10  | Ability to Measure Creativity                       | 1           | 1           | 1           |
|     | <b>Number of Valid Scores (<math>\geq 3</math>)</b> | <b>10</b>   | <b>10</b>   | <b>10</b>   |
|     | <b>CVI for Each Expert</b>                          | <b>1,00</b> | <b>1,00</b> | <b>1,00</b> |

$$CV_i = \frac{1,00 + 1,00 + 1,00}{3} = 1,00$$

This perfect CVI score indicates that the assessment instrument was deemed highly valid by all three subject matter experts. The second validation by the assessment experts also demonstrated significant improvements in quality. Reassessment was conducted on the same 10 aspects. The second validation sheet revealed that all aspects again received a score of  $\geq 3$ , with most improving to a score of 4 (Very Good), indicating improvements in substance and stronger test writing techniques.

$$CV_i = \frac{1,00}{1} = 1,00$$

Based on the assessment experts' CVI scores, the instrument is considered suitable for use with minor revisions. This means that its content validity is very good, but there are some minor improvements that can be made without changing the main substance of the questions. This result reflects the success of the revision process in improving the quality of the HOTS assessment instrument, including clarity, cognitive depth, contextual relevance, and originality of the questions. Based on validation results by three subject matter experts and one assessment expert, after improvements based on expert input, the HOTS-based assessment instrument was declared highly valid and suitable for use in measuring students' higher-order thinking skills in Economics.

Although validation has demonstrated the instrument's theoretical feasibility, it is important to acknowledge that this feasibility remains conceptual and has not been empirically tested on students through field trials. Therefore,

empirical validity, such as reliability, difficulty index, and item discrimination, could not be analyzed in this study. This means that the instrument's feasibility is still in its early stages, as a product of development that requires further testing in real-world implementation. This limitation does not diminish the instrument's quality; rather, it is part of the step-by-step scientific procedure required in research and development (R&D).

By considering all these aspects, a systematic design process, robust theoretical validation, revisions based on expert input, and alignment with learning outcomes and HOTS principles, this instrument can be declared content-feasible and ready to proceed to the empirical testing stage in real classrooms. This process will allow for a comprehensive evaluation of the instrument's functionality in real-world learning situations and measure its effectiveness in fostering students' critical thinking skills regarding contextual economic issues. Only through this next stage can the instrument truly become a valid and reliable measuring tool in HOTS-based learning in madrasas and public schools.

## CONCLUSION

This research was conducted to develop a Higher-Order Thinking Skills (HOTS)-based assessment instrument for Economics learning on the topic of *Market and Price Formation* at MAN 2 Kota Madiun. The development process employed the 3D model (Define, Design, Develop) to ensure that the resulting instrument was not only theoretically sound but also practical and aligned with the *Merdeka Curriculum*.

The findings can be summarized as follows. First, the preliminary study revealed that existing assessment practices at MAN 2 Kota Madiun were still dominated by Lower-Order Thinking Skills (LOTS) items, with a heavy focus on memorization and simple comprehension. Teachers faced difficulties in creating HOTS-oriented questions due to a lack of references and insufficient training, resulting in a gap between curriculum expectations and classroom implementation.

Second, the developed product consisted of 15 assessment items in various formats, including multiple-choice, matching, short-answer, and essay questions. Each item was designed to measure cognitive levels C4 (analysis), C5 (evaluation), and C6 (creation). The integration of different item types enriched the instrument and allowed for a more comprehensive measurement of students' higher-order thinking skills.

Third, the validation process demonstrated that the instrument achieved significant improvement after revisions. The initial expert review resulted in a Content Validity Index (CVI) score of 0.73, indicating "adequate validity." Following expert feedback and revisions—particularly in improving contextual clarity, strengthening stimulus materials, and refining instructions—the CVI score increased to 1.00, categorized as "excellent validity." Experts confirmed that the instrument was valid, feasible, and aligned with the learning objectives of the *Merdeka Curriculum*.

Based on these results, it can be concluded that the HOTS-based assessment instrument developed in this study is valid and suitable for use in Economics learning, particularly in the topic of market and price formation. The instrument has the potential to help teachers measure students' critical, evaluative, and creative thinking skills, moving assessment beyond rote learning towards more meaningful and reflective learning experiences. Nevertheless, the study acknowledges its limitations. The validation was restricted to expert judgment, and empirical trials with students were not conducted. As a result, item reliability, difficulty, and discrimination indices remain unexplored. Future research is strongly recommended to test the instrument in real classroom contexts to evaluate its statistical reliability and effectiveness.

In conclusion, this research contributes practically and theoretically to the field of educational assessment. Practically, it provides teachers with a structured and validated instrument to support HOTS-based assessment in Economics. Theoretically, it demonstrates that systematic development using the 3D model can produce valid and feasible tools that align with 21st-century education demands. By implementing this instrument, schools can enhance

the quality of learning assessment and foster students' ability to think critically, analytically, and creatively in accordance with the goals of modern education.

## REFERENCES

- Amelia, N., Rahmawati, R., & Adnan, A. (2024). STRATEGI ASESMEN UNTUK MENGUKUR KOMPETENSI ABAD 21 DALAM PEMBERALAJARAN BIOLOGI. *Biogenerasi Jurnal Pendidikan Biologi*, 10(1), 437–442.
- Anderson, L.W.; Krathwohl, D. R. (2001). *A Taxonomy for Learning , Teaching and Assesing; A Revision of Bloom's Taxonomy of Education Objectives*. Addison Wesley Lonman Inc.
- Angga, P. D., & Sari, A. J. (2025). Deep Learning : Bagaimana Implementasinya Pada Pendidikan Jasmani , Olahraga dan Kesehatan ( PJOK )? *Jurnal Ilmiah Profesi Pendidikan*, 10(2), 1373–1391.
- Ariyana, Y., Setiawati, W., Asmira, O., Bestary, R., & Pudjiastuti, A. (2019). *Buku Penilaian Berorientasi Higher Order Thinking Skills (HOTS)*. Direktorat Jenderal GTK, Kemendikbud.  
<https://repositori.kemdikbud.go.id/15158/>
- Budiarti, I., & Airlanda, G. S. (2019). Penerapan Model Problem Based Learning Berbasis Kearifan Lokal untuk Meningkatkan Keterampilan Berpikir Kritis. *Jurnal Riser Teknologi Dan Inovasi Pendidikan*, 2(1), 167–183.
- Fullan, M., Quinn, J., & McEachen, J. (2017). Deep learning: Engage the world change the world. *Corwin Press*.
- Halifah, S., Asrianti, N., & Okianna, O. (2024). Analisis Kesulitan Peserta Didik dalam Menyelesaikan Soal HOTS Mata Pelajaran Ekonomi Kelas XI IPS. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 13, 700–707.  
<https://doi.org/10.26418/jppk.v13i4.76613>
- Kaniawati, I., & Kasmahidayat, Y. (2021). Pembelajaran STEM: Upaya mencerdaskan kehidupan bangsa di abad 21. *Jurnal Majelis*, 2(1). [https://www.researchgate.net/profile/Yuliawan-Kasmahidayat/publication/389191744\\_Jurnal\\_Majelis\\_Ed\\_3\\_-\\_Arah\\_Kebijakan\\_Pembangunan\\_Nasional\\_Bidang\\_Pendidikan\\_dan\\_Kebudayaan/links/67b842e8207c0c20fa907212/Jurnal-Majelis-Ed-3-Arah-Kebijakan-Pembangunan-Nasio](https://www.researchgate.net/profile/Yuliawan-Kasmahidayat/publication/389191744_Jurnal_Majelis_Ed_3_-_Arah_Kebijakan_Pembangunan_Nasional_Bidang_Pendidikan_dan_Kebudayaan/links/67b842e8207c0c20fa907212/Jurnal-Majelis-Ed-3-Arah-Kebijakan-Pembangunan-Nasio)
- Lestari, S. A. P. (2019). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) pada Materi Himpunan Kelas VII SMP. *JKPM (Jurnal Kajian Pendidikan Matematika)*, 4(2), 111.  
<https://doi.org/10.30998/jkpm.v4i2.3862>
- Mashudi, M. (2021). Pembelajaran Modern: Membekali Peserta Didik Keterampilan Abad Ke-21. *Al-Mudarris (Jurnal Ilmiah Pendidikan Islam)*, 4(1), 93–114. <https://doi.org/10.23971/mdr.v4i1.3187>
- Masrina, M., Alvina, S., Fakhrah, F., & Mellyzar, M. (2023). Pengembangan Instrumen Penilaian Keterampilan 4C (Critical Thinking, Collaboration, Communication, Creativity) Siswa Pada Materi Sifat Koligatif Larutan. *Jurnal Genta Mulia*, 245–261.  
<https://ejournal.stkipbbm.ac.id/index.php/gm/article/view/555%0Ahttps://ejournal.stkipbbm.ac.id/index.php/gm/article/download/555/423>
- Muhibbuddin, M., Artika, W., & Nurmaliah, C. (2023). Improving Critical Thinking Skills Through Higher Order Thinking Skills (HOTS)-Based Science. *International Journal of Instruction*, 16(4), 283–296.  
<https://doi.org/10.29333/iji.2023.16417a>
- Nisfiah, L. S. (2019). PENGEMBANGAN LEMBAR KEGIATAN PESERTA DIDIK (LKPD) BERBASIS HIGHER ORDER THINKING SKILLS (HOTS) PADA MATA PELAJARAN PRAKTIKUM AKUNTANSI LEMBAGA UNTUK KELAS XI SMK. *Jurnal Pendidikan Akuntansi*, 07 Nomor 0, 409–414.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2005). Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing & Health*, 459–467. <https://doi.org/10.1002/nur>
- Putri, C. A., Rofiqoh, E., Wulandari, F. A., Amalia, F., Prastiningrum, P., & Eva, N. (2021). *Asesmen Autentik: Pengembangan Asesmen HOTS Mata Pelajaran Matematika pada Siswa SMP*.  
<http://conference.um.ac.id/index.php/psi/article/viewFile/1229/630%0A%0A>
- Putro, N. Y. (2024). Pembelajaran Ekonomi Berbasis Project : Guna Menumbuhkan High Order Thinking Skill (

- HOTS ) dalam Konteks Pendidikan Ekonomi. *Waqaf Ilmu Nusantara*.
- Ramdani, A., Jufri, A. W., Gunawan, G., Hadisaputra, S., & Zulkifli, L. (2019). Pengembangan Alat Evaluasi Pembelajaran Ipa Yang Mendukung Keterampilan Abad 21. *Jurnal Penelitian Pendidikan IPA*, 5(1), 98–108. <https://doi.org/10.29303/jppipa.v5i1.221>
- Sholeh, M. I., 'Azah, N., Tasya', D. A., Soki, S., Syafi'i, A., Sahri, S., Rosyidi, H., Arifin, Z., & Fatimmah, S. (2024). PENERAPAN PEMBELAJARAN BERBASIS PROYEK (PJBL) DALAM MENINGKATKAN KEMAMPUAN BERPIKIR KRITIS SISWA Muh. *Jurnal Tinta*, 6(2), 158–176. <https://doi.org/10.31004/cendekia.v8i2.3361>
- Sholiha, I. N., & Kurniawan, R. Y. (2022). Pengembangan Instrumen Penilaian Berbasis Higher Order Thinking Skills pada Mata Pelajaran Ekonomi Sekolah Menengah Atas. *Edukatif: Jurnal Ilmu Pendidikan*, 4(1), 123–132. <https://doi.org/10.31004/edukatif.v4i1.1736>
- Siregar, N. 'Athiyyah M., Muchtar, Z., Dibyanti, R. E., Sutiani, A., & Sinaga, M. (2023). Pengembangan Instrumen Evaluasi untuk Mengukur Keterampilan Berpikir Tingkat Tinggi pada Materi Keseimbangan Kimia. *JiIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(7), 4834–4842. <https://doi.org/10.54371/jiip.v7i8.4901>
- Siregar, Z. S., Ristiono, R., & Alberida, H. (2023). PENGEMBANGAN INSTRUMEN PENILAIAN KEMAMPUAN BERPIKIR TINGKAT TINGGI TENTANG MATERI SISTEM GERAK PADA MANUSIA UNTUK PESERTA DIDIK KELAS XI SMA/MA. *BIOCHEPHY: Journal of Science Education*, 03(2), 98–110. <https://doi.org/10.52562/biochephy.v3i2.526>
- Wahidin, W. (2023). Pelaksanaan evaluasi pembelajaran berbasis higher order thinking skills (HOTS). *Jurnal Ilmu Pendidikan & Sosial*, 01(03), 107–116.
- Wahyudi, E., Destiniar, D., & Fuadiah, N. F. (2023). Pengembangan Instrument Tes Berbasis Asesmen Kompetensi Minimum Materi Trigonometri Pembelajaran Matematika Kelas X SMA. *JIPM (Jurnal Ilmiah Pendidikan Matematika)*, 12(1), 59–69. <https://doi.org/10.25273/jipm.v12i1.15075>
- Widiana, I. W., Jampel, I. N., & Suranata, K. (2025). *Bunga Rampai Peran Teknologi dalam Deep Learning*. Penerbit Widina Media Utama Dilarang.
- Zahra, D. Z. 'Aqila, & Subekti, H. (2025). Validasi Modul Ajar Berbasis Keterampilan Proses Sains Menggunakan Content Validity Index (Cvi). *BIOCHEPHY: Journal of Science Education*, 5(1), 613–617. <https://doi.org/10.52562/biochephy.v5i1.1602>