Journal of Educational Sciences Vol. 9 No. 5 (Sept, 2025) 3336-3351

Journal of Educational Sciences
Journal homepage: https://jes.ejournal.unri.ac.id/index.php/JES

P-ISSN 2581-1657
E-ISSN 2581-2203

Development of Higher Order Thinking Skills (HOTS) Test Items in IPAS to Evaluate Fifth-Grade Students' Problem-Solving Abilities

Marselina Juita*, Sri Lastuti, Hairunisa, Muslim, Arif Hidayad
Primary School Teacher Education Program, STKIP Taman Siswa, Bima, 84119, Indonesia

ARTICLE INFO

Article history:

Received: 15 Aug 2025
Revised: 21 Aug 2025
Accepted: 03 Sept 2025
Published online: 10 Sept 2025

Keywords:

Development,
HOTS,
IPAS,
Problem Solving

** Corresponding author:*

E-mail: marselinajuita18@gmail.com

Article Doi:

https://doi.org/10.31258/jes.9.5.p.3336-3351

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

A B S T R A C T

Higher Order Thinking Skills refer to advanced cognitive abilities involving analysis, evaluation, and creation in solving problems critically and logically. Problem-solving skills are defined as an individual's ability to identify problems, formulate alternative solutions, and evaluate outcomes systematically. This study aims to develop a test instrument based on higher order thinking skills for the Natural and Social Sciences subject for fifth-grade elementary students that is valid and reliable in measuring problem-solving skills. The research employed a research and development method using a modified Borg and Gall model consisting of seven stages. The developed instrument consists of 25 multiple-choice items. Content validity, assessed using Aiken's V formula, obtained a score of 0.81, indicating a very valid category. Construct validity showed that 16 out of 25 items were valid after revision. Reliability testing using the split-half technique produced coefficients of Cronbach's Alpha, Spearman-Brown, and Guttman above 0.98, indicating strong internal consistency. The difficulty index ranged from 0.42 to 0.51 (moderate), while the discrimination index showed that 16 items had values equal to or greater than 0.755. The instrument is considered valid and feasible for use.

1. Introduction

In the 21st century, mastering three core competencies is essential for students: creative thinking, critical thinking, and problem-solving skills. Problem-solving is the ability to understand a problem in order to provide an appropriate solution based on one's existing knowledge. According to (Kurniawati et al., 2019) the development of problem-solving skills is one of the essential competencies that must be instilled in students to shape a generation that is competitive and adaptive to the challenges of the 21st century. There are four indicators related to problem-solving skills: 1) problem identification, 2) problem formulation, 3) implementation of solution strategies, and 4) verification of the obtained solution (Chabibah et al., 2019).

In the context of Natural and Social Sciences (IPAS) learning, students are often faced with various challenges, particularly related to their ability to solve problems presented by the teacher. IPAS is an educational framework that integrates the content of Natural Sciences (IPA) and Social Sciences (IPS) within a thematic learning process (Suhelayanti et al., 2023). Furthermore (Azzahra et al., 2023) also emphasize that IPAS includes scientific studies on the interactions among organisms within an ecosystem.

The integration of Natural Sciences (IPA) and Social Sciences (IPS) in the Merdeka Curriculum provides students with the opportunity to understand the interrelationships between natural phenomena, social contexts, and human interactions (Ni Kadek Sinta Ulandari et al., 2024). Problem-solving presented as challenges in IPAS learning demands higher-level cognitive abilities, commonly known as Higher Order Thinking Skills (HOTS). These higher-order cognitive skills are critically important in the 21st century, given the increasingly complex and dynamic development of scientific disciplines in line with changes in contemporary society (Higher et al., 2023).

Higher Order Thinking Skills (HOTS) encompass the capacity for critical thinking, particularly in managing information that leads to comprehensive problem-solving. The indicators of Higher Order Thinking Skills (HOTS) consist of three main components: analysis (C4), evaluation (C5), and creation (C6) (Suci et al., 2021). Numerous studies have been conducted extensively to investigate Higher Order Thinking Skills (HOTS). One such study was conducted by (Nuraini & Julianto, 2022) which investigated the factors causing difficulties experienced by fourth-grade elementary students in solving science problems related to higher order thinking skills.

The findings of this study indicate that students face significant challenges in solving questions categorized as Higher Order Thinking Skills (HOTS). However, previous studies have not specifically focused their investigations on problem-solving abilities. For example, a study conducted by (Marvia Afrita & Rahmawati Darussyamsu, 2020) focused on assessing the validity of HOTS test instruments in the context of respiratory system material at the secondary school level. The study employed descriptive statistical analysis using Microsoft Excel to interpret data from validation questionnaires completed by expert validators. Nevertheless, the study did

Unlike previous studies, this research explicitly focuses on analyzing content validity, construct validity, reliability, difficulty level, and item discrimination of HOTS questions as instruments to evaluate problem-solving abilities in IPAS learning among fifth-grade students. The developed HOTS questions play a crucial role and therefore must align with the competencies being measured, correspond to problem-solving indicators, and reflect the cognitive levels that represent higher-order thinking skills. Additionally, the clarity of language in the test items must be considered. All these aspects are evaluated through a validation process of the developed instrument. The validation of the HOTS items was conducted by two

validators with expertise in IPAS and the development of higher-order thinking skill assessments.

(Marvia Afrita & Rahmawati Darussyamsu, 2020) stated that assessment instruments for higher-order thinking skills must be developed by considering the competencies being measured, the use of language that meets established standards, and cognitive levels that align with higher-order thinking abilities. Based on this explanation, the purpose of this study is to evaluate the content validity, construct validity, reliability, difficulty level, and item discrimination of Higher Order Thinking Skills (HOTS) test items as instruments to measure the problem-solving abilities of fifth-grade students in the Natural and Social Sciences (IPAS) subject.

2. Methodology

Developmental research, also known as Research and Development (R&D). Model pengembangan yang digunakan adalah Borg & Gall (Assyauqi, 2020). The development research stages proposed by Borg and Gall consist of ten steps; however, in this study, the researcher focuses on seven stages: 1) collecting information and data, 2) planning, 3) developing the initial product, 4) conducting preliminary field testing, 5) revising the initial product, 6) conducting main field testing, and 7) revising the final product. The data collection process involved interviewing IPAS teachers. The planning stage began with designing a HOTS test blueprint, which included IPAS material, problem-solving indicators, HOTS cognitive levels (C4, C5, and C6), question formats, and scoring guidelines. Next, the initial product development involved drafting HOTS questions, scoring rubrics, and answer keys, followed by validation conducted by two experts in HOTS item development who served as validators.

Following the validation process, an initial field trial was conducted involving 10 students. The results from this trial were utilized to revise the preliminary product. Subsequently, the main field trial was carried out with 100 students to evaluate their problem-solving skills. Based on the outcomes of the main trial, the final product was revised accordingly. Data analysis employed both qualitative and quantitative approaches. The qualitative analysis was descriptive, aiming to describe the results of the previously collected data. The quantitative analysis included content validity, construct validity, reliability, item discrimination, and item difficulty level. Content validity was analyzed using Aiken's V formula in Excel, with the formula presented below and the Aiken's V criteria shown in Table 1.

Formula

$$V = \frac{\sum S}{(\mathrm{n}(\mathrm{c}-1))}$$

V: Aiken's content validity coefficient
S: Expert score adjusted by subtracting the lowest score on the scale (s = r – 1)
r: Score given by the expert for item 1
n: Number of validators.

Table 1. Criteria for Aiken's V

Aiken's V	Category
$V \leq 0,4$	Low
$0,4 < V \leq 0,8$	Medium
$0,8 < V \leq 1,0$	High

(Utami et al., 2024)

Construct validity analysis aims to determine the extent to which test items measure the intended construct. Construct validity was analyzed using SPSS by applying the product-moment correlation formula. The criteria for assessing construct validity are presented in Table 2 below.

Table 2. Construct Validity Criteria

Construct validity	Category
$0,7 < D \leq 1,0$	Very high
$0,4 < D \leq 0,7$	High
$0,2 < D \leq 0,4$	Currently
$0 < D \leq 0,2$	Low

Gregory, 2007 as cited in (Fina Sholeha et al., 2023)

Reliability was analyzed using SPSS software with the Guttman Split-Half coefficient technique. The criteria for the reliability test are presented in Table 3 below.

Table 3. Instrument Reliability Criteria

Reliability	Category
0,81 - 1,00	Very high
0,61 - 0,80	High
0,41 - 0,60	Currently
0,21 - 0,40	Low
0,00 - 0,20	Very low

(Nu'man, 2023)

Item difficulty level was analyzed using SPSS with descriptive statistical techniques. The criteria for item difficulty levels are presented in Table 4 below.

Table 4. Criteria for Item Difficulty Level

Item Difficulty Level	Category
0,71 – 1,00	Difficult
0,31 – 0,70	Medium
0, 00 – 0,30	Easy

(Nu'man, 2023)

Item discrimination was analyzed using the SPSS application. The criteria for item discrimination for each test item are presented in Table 5 below.

Table 5. Criteria for Item Discrimination

Item Discrimination	Category
$D \leq 0{,}00$	Very bad
$0{,}00 < D \leq 0{,}20$	Bad
$0{,}20 < D \leq 0{,}40$	Enough
$0{,}40 < D \leq 0{,}70$	Good
$0{,}70 < D \leq 1{,}00$	Very good

(Nu'man, 2023)

3. Results and Discussion

The development of HOTS-based questions in the IPAS subject aims to produce an instrument that is valid, reliable, and possesses appropriate levels of difficulty and discrimination to assess the problem-solving abilities of fifth-grade elementary students. The developed instrument consists of 25 multiple-choice items. Following the development process, content validity was assessed by two expert validators who are university lecturers specializing in HOTS and IPAS. Content validity analysis was conducted using Aiken's V formula, as presented in Table 6.

Based on the results of the content validity analysis, the Aiken's V score obtained was 0.81, indicating that the instrument falls into the "very valid" category. This content validity was determined based on the assessments of two validators across several aspects, including: the relevance of images to the questions, alignment of questions with the material, clarity of language, completeness of answer keys, and clarity of scoring guidelines. The evaluation was conducted using a 1–5 scale (very low to very high). Therefore, the developed HOTS items are considered appropriate for use as an instrument to assess the problem-solving abilities of fifth-grade students.

The validity of the content was further supported by findings from several previous studies. (Aristiyo et al., 2021) stated that the developed items met the criteria for being valid and highly valid. A study conducted by (Kusuma & Nurmawanti, 2023) also showed that the developed HOTS items were considered valid, with an Aiken's V score of 0.88. Furthermore, (Risdiana et al., 2022) stated that the developed items were categorized as highly valid, with a feasibility percentage of 90.7%.

Thus, based on the validators' assessments and data analysis using Aiken's V formula, it can be concluded that the developed instrument meets the criteria for high content validity. Following this stage, the instrument will be administered in a limited trial involving 10 fifth-grade students to be further analyzed for construct validity, reliability, difficulty level, and item discrimination using SPSS software, in order to obtain a comprehensive overview of the quality of the developed HOTS items. Based on the results of the construct validity analysis presented in Table 7 below, out of a total of 25 test items analyzed, 19 items (numbers 1, 2, 4, 5, 6, 8, 10, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25) were declared valid, as they had calculated r-values (r-count) greater than or equal to the critical r-value (r-table). These findings indicate that the majority of the test items accurately

measured the intended construct, in accordance with the learning objectives and predetermined indicators, can be seen in Table 6.

Table 6. Results of Content Validity Analysis of Test Items

Item	Validators		S1	S2	∑S	n(c-1)	V	Category
	I	II						
Item 1	4	5	3	4	7	8	0,87	Very High
Item 2	3	5	2	4	6	8	0,75	High
Item 3	4	4	3	3	6	8	0,75	High
Item 4	5	4	4	3	7	8	0,87	Very High
Total Score	16	18	12	14	26	32	0,81	Very High

Meanwhile, six items (numbers 3, 7, 9, 11, 13, and 15) were found to be invalid, as their calculated r-values were lower than the critical r-value, indicating no significant correlation with the total score. This suggests that these six items did not adequately represent the intended construct and therefore require revision, either in terms of language, content, or clarity of context.

The findings of this study are consistent with (Fina Sholeha et al., 2023) who reported that the multiple-choice test instrument consisting of 30 items demonstrated good quality based on construct validity analysis, with 50% of the items categorized as valid.

Table 7. Results of Construct Validity Analysis from the Limited Tr

Criteria	Item Number
Valid	1, 2, 4, 5, 6, 8, 10, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
Invalid	3, 7, 9, 11, 13, 15

Based on the reliability test results presented in Table 8, the Cronbach's Alpha values for each section of the test ranged from 0.990 to 0.995, indicating that each part of the instrument possesses a very high level of internal consistency. The correlation value between the two sections of the test was 0.995, indicating a very strong relationship and consistent measurement of the same construct. The Spearman-Brown coefficients, both under equal and unequal item lengths, were 0.997. This value is near perfect and suggests that the reliability of the instrument is not affected by differences in the number of items in each section. Additionally, the Guttman Split-Half Coefficient value of 0.996 further confirms that the instrument has very high reliability, as the coefficient approaches the maximum value of 1.00.

Based on these results, the researcher concluded that the developed test instrument meets the criteria for very high reliability, with a value range of $0.80 < r \leq 1.00$. These findings are consistent with the results of a study conducted by (Rohmah et al., 2023) in which the analyzed instrument demonstrated a reliability score of 0.91, which is also classified as very high. Similar results were also found in a study conducted by (Luthfi et al., 2019) which employed the same approach and analysis technique, namely the SPSS application with the Guttman Split-Half formula, and obtained a reliability score of 0.90. This demonstrates consistent findings, indicating that the developed instrument is categorized as highly reliable.

Table 8. Reliability Analysis Results Using the Guttman Split-Half Coefficient Formula

Reliability Statistics			
Cronbach's Alpha	Part 1	Value	.990
		N of Items	10[a]
	Part 2	Value	.995
		N of Items	9[b]
	Total N of Items		19
Correlation Between Forms			.995
Spearman-Brown Coefficient	Equal Length		.997
	Unequal Length		.997
Guttman Split-Half Coefficient			.996

Based on the item difficulty analysis results presented in Table 9, it was found that 94.7% of the test items fell into the moderate category, indicating that the instrument has a proportional difficulty level suitable for measuring students' abilities. Only one item (Item 5) was classified as difficult, with a mean score of 0.30 (5.3%), and no items were categorized as easy (mean ≥ 0.71). This indicates that none of the items were answered correctly by all students, suggesting that the items were effective in distinguishing varying levels of content mastery. These findings are consistent with the study conducted by (Fina Sholeha et al., 2023) which showed that most of the developed items had a moderate level of difficulty and were appropriate for use in learning evaluation.

Table 9. Results of Item Difficulty Analysis Using Descriptive Statistics Formula

Statistics			
	N		**Mean**
	Valid	**Missing**	
Item 1	10	0	.60
Item 2	10	0	.60
Item 4	10	0	.60
Item 5	10	0	.30
Item 6	10	0	.60
Item 8	10	0	.60
Item 10	10	0	.60
Item 12	10	0	.60
Item 14	10	0	.60

Item 16	10	0	.60
Item 17	10	0	.60
Item 18	10	0	.60
Item 19	10	0	.60
Item 20	10	0	.70
Item 21	10	0	.60
Item 22	10	0	.60
Item 23	10	0	.60
Item 24	10	0	.60
Item 25	10	0	.60

Based on the item discrimination analysis using the Corrected Item-Total Correlation indicator, as presented in Table 10, the reliability analysis using Cronbach's Alpha yielded a score of 0.996 for the 19 test items. This value falls within the very high category ($0.80 < r \leq 1.00$), indicating that the instrument possesses excellent internal consistency and is capable of reliably measuring the intended construct. Therefore, the test instrument is considered reliable and suitable for use in learning evaluation.

Table 10. Results of Item Discrimination Analysis Using Cronbach's Alpha Formula

Reliability Statistics	
Cronbach's Alpha	N of Items
.996	19

Based on the item discrimination analysis using Cronbach's Alpha, as presented in table 11 below, it was found that 18 out of 19 items had a Corrected Item-Total Correlation value of ≥ 0.801 and a Cronbach's Alpha if Item Deleted value of 0.996, indicating excellent discrimination and high internal consistency. Only one item, Item 5, had a correlation value of 0.534 and a Cronbach's Alpha if Item Deleted value of 0.999. Although this value is comparatively lower, the item is still categorized as good and suitable for use. Overall, the instrument demonstrates excellent quality in distinguishing students' levels of mastery and is considered reliable.

Table 11. Results of Item Discrimination Analysis

	Item-Total Statistics			
	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
Item 1	10.60	79.600	.998	.996
Item 2	10.60	79.600	.998	.996
Item 4	10.60	79.600	.998	.996
Item 5	10.90	84.100	.534	.999
Item 6	10.60	79.600	.998	.996
Item 8	10.60	79.600	.998	.996
Item 10	10.60	79.600	.998	.996
Item 12	10.60	79.600	.998	.996
Item 14	10.60	79.600	.998	.996
Item 16	10.60	79.600	.998	.996

Item 17	10.60	79.600	.998	.996
Item 18	10.60	79.600	.998	.996
Item 19	10.60	79.600	.998	.996
Item 20	10.50	81.833	.801	.997
Item 21	10.60	79.600	.998	.996
Item 22	10.60	79.600	.998	.996
Item 23	10.60	79.600	.998	.996
Item 24	10.60	79.600	.998	.996
Item 25	10.60	79.600	.998	.996

After the analysis, the test items that were found to be invalid were revised in order to obtain valid items suitable for large-scale trials. The large-scale trial involved 100 students and aimed to gather more comprehensive empirical data regarding the quality of the instrument. Students' responses to all test items were analyzed using construct validity analysis, instrument reliability, item difficulty, and item discrimination techniques. The revised instrument results are presented in Table 12 below.

Table 12. Results of the Revised HOTS Instrument

Before revision	**After revision**
Question 3. The development of biological science helps scientists understand the relationships among living organisms within an ecosystem. One important concept is the food web, which is formed from several … a. Food chains that are not interconnected b. Food chains that occur sequentially c. Interdependent food chains d. Organisms that share the same habitat	Advances in biological sciences enable scientists to understand interactions among living organisms within an ecosystem. One concept arising from this understanding is the food web, which is composed of several … a. Independent food chains that are not connected b. Food chains arranged sequentially c. Interconnected and interdependent food chains d. Organisms that occupy the same habitat
Question 7. Based on the image above, if the rat becomes extinct, the populations of owls and snakes will decrease, causing the ecosystem within the food web to become unstable. From this statement, it can be concluded that... a. Every component in the food web preys on each other and benefits one another. b. Every organism in the food web is connected but harms one another. c. Every component in the food web is interconnected and interdependent.	Based on the image above, if the rat becomes extinct, the populations of owls and snakes will decrease. This causes the food web to become unbalanced and the ecosystem to be disrupted. From this statement, it can be concluded that... a. Every living organism in the food web preys on and benefits one another. b. Every organism in the food web is interconnected but harms one another. c. Every component in the food web is connected and mutually dependent.

d. All of the above are correct.

Question 9. Observe the image of the food web in the paddy field ecosystem below.!

Based on the image above, it can be concluded that...

a. Rice serves as producer 1 and the tree as producer 2.
b. The frog primarily feeds on caterpillars.
c. There is a mutually beneficial relationship between producers and consumers.
d. A group of animals relies on plants as their main source of food.

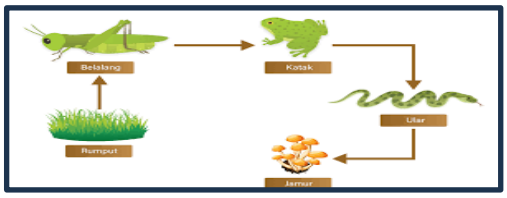
d. All of the above options are correct.

Observe the following food web diagram in a paddy field ecosystem.

Based on the image, the most accurate conclusion is…

a. Rice plants act as the primary producers, while trees serve as secondary producers.
b. Frogs obtain their main food source from caterpillars within the scope of the food web.
c. There is a mutualistic relationship between producers and consumers in the ecosystem.
d. Several animal species depend on plants as their primary food source.

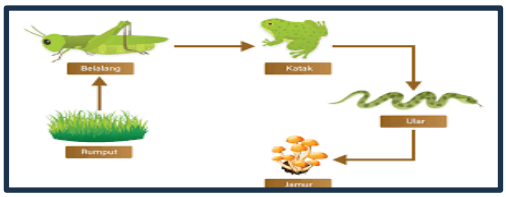
Question 11. Observe the food chain diagram below.!



The impact if the frog population becomes extinct is…

a. The populations of snakes and grasshoppers will increase.
b. The snake population will remain stable because grasshoppers are still available as their food source.
c. The grasshopper population will increase, thereby keeping the ecosystem stable.
d. The snake population will decrease, and the grasshopper population will increase.

Observe the food chain diagram below.!

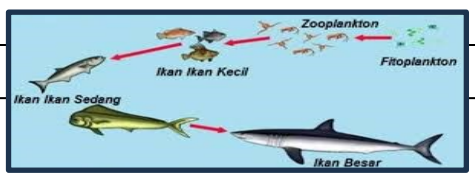


If the frog population becomes extinct, the most likely impact will be...

a. The populations of snakes and grasshoppers will increase simultaneously.
b. The snake population remains stable because it can still feed on grasshoppers.
c. The grasshopper population increases, but the ecosystem balance remains maintained.
d. The snake population decreases due to loss of food source, and the grasshopper population increases due to absence of predators.

Question 13. Observe the following food chain diagram!

Perhatikan gambar rantai makanan berikut !



Based on the image above, if the primary consumer becomes extinct, the resulting impact would be…

a. The populations of secondary, tertiary, quaternary, and quinary consumers will decrease.
b. The producer population will decrease.
c. The secondary consumer population will increase, while the tertiary consumer population will decrease.
d. The quaternary consumer population will incre

Question 15. Observe the following food chain:
Grass → Grasshopper → Frog → Snake

If the grasshopper population in this food chain becomes extinct, the impact would be…

a. The grass population will decrease because there are no herbivores to consume it.
b. The frog population will decrease due to the loss of a food source.
c. The snake population will increase due to an abundance of food.
d. The snake population will decrease, maintaining ecosystem stability.

Based on the diagram, if the primary consumer becomes extinct, the most likely impact would be…

a. The populations of secondary, tertiary, quaternary, and quinary consumers will decline.
b. The producer population will decrease due to the absence of herbivores.
c. The secondary consumer population will increase, while the tertiary consumer population will decrease.
d. The quaternary consumer population will increase due to reduced competition for food.

Observe the following food chain:
Grass → Grasshopper → Frog → Snake

If the grasshopper population in this food chain becomes extinct, the most likely impact would be…

a. The snake population will decrease, thus maintaining ecosystem stability.
b. The frog population will decline due to the loss of its food source.
c. The snake population will increase because it has more food available.
d. The grass population will decrease because there are no herbivores to consume it.

Based on the construct validity analysis presented in Table 13 below, it was found that 16 items met the criteria for validity, while 9 items were deemed invalid. The valid items include numbers: 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 17, 18, 20, 22, and 24. Construct validity is demonstrated by the statistically significant correlation between individual item scores and the total score, where the correlation coefficients exceed the critical value ($r_{calculated} \geq r_{table}$). This indicates that these items consistently and representatively measure the intended construct or cognitive aspect. In contrast, nine items were considered invalid due to correlation values below the threshold ($r_{calculated} < r_{table}$), specifically items numbered 7, 9, 11, 13, 15, 19, 21, 23, and 25. Therefore, the valid items can be retained in the instrument, while the invalid items should be either eliminated or revised for future implementation.

This construct validity analysis represents a critical step in ensuring that each test item accurately reflects the intended dimension of the measured ability, both theoretically and empirically, thereby guaranteeing the quality of the instrument's measurement. These findings are consistent with those reported by (Kartini et al., 2022) which indicated that out of 30 developed multiple-choice items, 50% were classified as valid based on the construct validity analysis.

Table 13. Results of Construct Validity Analysis

Criteria	Item Number
Valid	1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 17, 18, 20, 22, 24
Invalid	7, 9, 11, 13, 15, 19, 21, 23, 25

According to the reliability analysis results presented in Table 14, the instrument demonstrates a very high level of internal consistency. The Cronbach's alpha coefficient for Part 1 was .941, and for Part 2 it was .998, indicating that both sections of the instrument possess excellent individual reliability. Each section contained eight items, resulting in a total of 16 items analyzed. The correlation between the two sections (Correlation Between Forms) was .985, indicating a strong positive relationship between Part 1 and Part 2.

The Spearman-Brown coefficient, calculated for both equal and unequal test lengths, was .992, indicating that the overall reliability of the instrument remains very high when both parts are combined. Furthermore, the Guttman Split-Half coefficient was .985, further supporting the instrument's excellent internal stability. All obtained reliability coefficients were well above the minimum threshold for research instrument reliability (≥ .70). Therefore, it can be concluded that the developed instrument is highly reliable and suitable for consistently and accurately measuring students' abilities. These results are consistent with the findings of (Dasar, 2024) which obtained a reliability coefficient of .76 and was categorized as high. This classification is supported by Guilford (Pertiwi et al., 2024), which states that an r value of .76 falls within the category of high reliability.

Table 14. Reliability Analysis Results Using the Guttman Split-Half Coefficient Formula

Reliability Statistics			
Cronbach's Alpha	Part 1	Value	.941
		N of Items	8[a]
	Part 2	Value	.998
		N of Items	8[b]
	Total N of Items		16
Correlation Between Forms			.985
Spearman-Brown Coefficient	Equal Length		.992
	Unequal Length		.992
Guttman Split-Half Coefficient			.985

Based on the difficulty level analysis presented in table 15 below, it was found that the 16 items had mean values ranging from .42 to .51. These values indicate that most students answered the questions correctly with a moderate proportion.

Referring to Arikunto's criteria, this range is categorized as (Fina Sholeha et al., 2023), Items with difficulty indices ranging from .30 to .70 are categorized as having a moderate level of difficulty. Therefore, all analyzed items namely items 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 17, 18, 20, 22, and 24 fall into the moderate difficulty category.

This balanced level of difficulty indicates that the items are neither too easy nor too difficult, making them appropriate for optimally and objectively measuring students' abilities. Such balance is essential in developing evaluation instruments aimed at identifying variations in student ability proportionally. These findings are consistent with the results of research conducted by (Dasar, 2024) These findings are consistent with the results of a study that reported an average difficulty index of .59, which falls within the moderate category. Therefore, it can be concluded that the Higher Order Thinking Skills (HOTS) test instrument possesses good quality in terms of item difficulty and is appropriate for use in learning evaluation. Subsequently, the items deemed appropriate will be further analyzed for their discriminating power using the Cronbach's Alpha formula.

Table 15. Difficulty Level Analysis Results Using Descriptive Statistics Method

Statistics			
	N		Mean
	Valid	Missing	
Item 1	100	0	.51
Item 2	100	0	.49
Item 3	100	0	.48
Item 4	100	0	.50
Item 5	100	0	.44
Item 6	100	0	.42
Item 8	100	0	.50
Item 10	100	0	.49
Item 12	100	0	.50
Item 14	100	0	.50
Item 16	100	0	.50
]Item 17	100	0	.49
Item 18	100	0	.50
Item 20	100	0	.51
Item 22	100	0	.49
Item 24	100	0	.50

Based on the item discrimination analysis of 16 items using the Cronbach's Alpha indicator, as presented in table 16 below, the obtained reliability coefficient was .987. This value indicates that the instrument has a very high level of internal consistency, as it is well above the minimum recommended threshold for research instruments (≥ .70). The high Cronbach's Alpha value suggests that most items have a strong correlation with the total score, thereby effectively distinguishing between students with high and low levels of understanding. Therefore, the instrument can be considered highly reliable and suitable for use as an evaluation tool capable of consistently and accurately measuring student performance. Overall, the instrument demonstrates excellent discriminating power and supports the reliability of the test. These findings are consistent with the study conducted by

(Fina Sholeha et al., 2023) yang menunjukkan bahwa soal-soal HOTS memiliki daya beda yang baik dan layak digunakan sebagai alat evaluasi pembelajaran.

Table 16. Discriminating Power Analysis Results Using the Cronbach's Alpha Formula

Reliability Statistics	
Cronbach's Alpha	**N of Items**
.987	16

4. Conclusion

Based on the entire development process of the HOTS IPAS instrument for fifth-grade elementary school students, which was conducted using a modified version of the Borg & Gall R&D model, it can be concluded that the resulting instrument possesses a very strong level of content validity. All aspects including content, visuals, language, and scoring system were evaluated to meet expert reliability standards. Furthermore, the construct validity analysis showed that the majority of the items were deemed valid and appropriate for use, with item selection based on pilot testing and statistical analysis. Internal reliability testing indicated a very high level of consistency, with Cronbach's Alpha, Spearman-Brown, and Guttman Split-Half estimates collectively confirming the robustness of the measurement. The difficulty level analysis revealed that all items fell within a moderate range, indicating the instrument's ability to proportionally measure problem-solving skills without being overly easy or difficult. Additionally, nearly all items demonstrated adequate to excellent discriminating power, with most items meeting commonly accepted item-total correlation standards used as benchmarks for test construction quality; only one item did not meet the criteria and is therefore recommended for elimination. In conclusion, the developed HOTS IPAS instrument demonstrates strong psychometric properties being highly valid, highly reliable, moderately difficult, and effectively discriminative. These characteristics support its use as an objective, proportional, and dependable evaluation tool for assessing fifth-grade students' problem-solving abilities.

References

Aristiyo, D. N., Triastuti, I. Y., & Fasha, E. F. (2021). Pengembangan Instrumen Soal Hots Matematika Tingkat Sma/Smk Untuk Menunjang Kemampuan Literasi Matematis. *Delta: Jurnal Ilmiah Pendidikan Matematika*, *9*(1), 89. https://doi.org/10.31941/delta.v9i1.1262

Assyauqi, M. I. (2020). Model Pengembangan Borg and Gall. *Institut Agama Islam Negeriegeri*, *December*, 2–8. https://www.taufiq.net/2019/09/model-penelitian-pengembangan-borg-and.html

Azzahra, I., Aan Nurhasanah, & Eli Hermawati. (2023). Implementasi Kurikulum Merdeka Pada Pembelajaran IPAS di SDN 4 Purwawinangun. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, *9*(2), 6230–6238. https://doi.org/10.36989/didaktik.v9i2.1270

Chabibah, L. N., Siswanah, E., & Tsani, D. F. (2019). Analisis kemampuan pemecahan masalah siswa dalam menyelesaikan soal cerita barisan ditinjau dari adversity quotient. *Pythagoras: Jurnal Pendidikan Matematika*, *14*(2), 199–210. https://doi.org/10.21831/pg.v14i2.29024

Fina Sholeha, D., Kusasi, M., Bakti, I., Brigjend Hasan Basry Banjarmasin, J. H., & Selatan Indonesia, K. (2023). Pengembangan Instrumen Soal Higher-Order Thinking Skills (Hots) Berbasis Google Form Untuk Mengukur Hasil Belajar Peserta Didik Pada Materi Hidrolisis Garam Development of Higher-Order Thinking Skills (HOTS) Questions Based on Google Form to Measure Students' Learning Outcomes on Salt Hydrolysis Materials. *Journal of Chemistry And Education)*, *7*(1), 32–48. http://jtam.ulm.ac.id/index.php/jcae

Higher, I., Thinking, O., Dalam, S., Pembelajaran, P., Penilaian, D. A. N., Sma, D. I., Ponorogo, B., Pendidikan, J., Islam, A., Tarbiyah, F., & Ilmu, D. A. N. (2023). *Implementasi Higher Order Thinking Skills Dalam*.

Kartini, N. E., Nurdin, E. S., Hakam, K. A., & Syihabuddin, S. (2022). Telaah Revisi Teori Domain Kognitif Taksonomi Bloom dan Keterkaitannya dalam Kurikulum Pendidikan Agama Islam. *Jurnal Basicedu*, *6*(4), 7292–7302. https://doi.org/10.31004/basicedu.v6i4.3478

Kurniawati, I., Raharjo, T. J., & Khumaedi. (2019). Peningkatan Kemampuan Pemecahan Masalah untuk Mempersiapkan Generasi Unggul Menghadapi Tantangan abad 21. *Seminar Nasinal Pascasarjana*, *21*(2), 702.

Kusuma, A. S., & Nurmawanti, I. (2023). Pengembangan Soal-Soal Literasi dan Numerasi Berbasis High Order Thinking Skills (HOTS) untuk Siswa Sekolah Dasar (SD). *Jurnal Ilmiah Profesi Pendidikan*, *8*(1), 516–523. https://doi.org/10.29303/jipp.v8i1.1313

Luthfi, I. A., Muharomah, D. R., Ristanto, R. H., & Miarsyah, M. (2019). Pengembangan Tes Kemampuan Pemecahan Masalah Pada Isu Pencemaran Lingkungan. *BioEdUIN: Jurnal Program Studi Pendidikan Biologi*, *9*(2), 11–21.

Marvia Afrita, & Rahmawati Darussyamsu. (2020). Validitas Instrumen Tes Berpikir Tingkat Tinggi (HOTS) pada Materi Sistem Respirasi di Kelas XI SMA. *Mangifera Edu*, *4*(2), 129–142. https://doi.org/10.31943/mangiferaedu.v4i2.83

Ni Kadek Sinta Ulandari, Ni Wayan Sri Darmayanti, & I Ketut Dedi Agung Susanto Putra. (2024). Analisis Pelaksanaan Pembelajaran IPA Kelas V SD N 1 Bebalang. *Lencana: Jurnal Inovasi Ilmu Pendidikan*, *2*(3), 117–126. https://doi.org/10.55606/lencana.v2i3.3707

Nu'man, M. (2023). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title. *Aleph*, *87*(1,2), 149–200. https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/Loeblein%2C Lucineia Carla.pdf?sequence=1&isAllowed=y%0Ahttps://antigo.mdr.gov.br/saneamento/proees

Nuraini, T., & Julianto. (2022). Analisis Faktor Penyebab Kesulitan Siswa Sekolah Dasar Kelas Iv Dalam Menyelesaikan Soal Hots (High Order Thinking

Skills) Pada Mata Pelajaran IPA. *Jurnal Pendidikan Guru Sekolah Dasar*, *10*(1), 60–74.

Risdiana, A., Erna, M., & Holiwarni, B. (2022). Pengembangan Soal HOTS (Higher Order Thinking Skills) pada Materi Asam-Basa untuk Kelas XI SMA/MA Sederajat. *Jurnal Inovasi Pendidikan Kimia*, *16*(2), 111–117. https://doi.org/10.15294/jipk.v16i2.32879

Rohmah, F., Rusilowati, A., Supriyadi, S., & Widowati, T. (2023). Validitas dan reliabilitas instrumen penilaian berbasis android untuk menilai kemampuan siswa SMK dalam mendesain busana secara digital. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, *9*(2), 577. https://doi.org/10.29210/020232131

Suci, I. E., Martini, & Purnomo, A. R. (2021). Analisis Muatan Higher Order Thingking Skill (HOTS) dalam Buku IPA Kelas VIII SMP BAB Struktur dan Fungsi Tumbuhan. *Pensa E-Jurnal : Pendidikan Sains*, *9*(3), 316–324. https://ejournal.unesa.ac.id/index.php/pensa

Suhelayanti, Z, S., & Rahmawati, I. (2023). Pembelajaran Ilmu Pengetahuan Alam Sosial (IPAS). In *Penerbit Yayasan Kita Menulis*.

Utami, L., Festiyed, Dian Purnama Ilahi, Arista Ratih, Elvi yenti, & Lazulva. (2024). Analisis Indeks Aiken Untuk Mengetahui Validitas Isi Instrumen Scinetific Habbits of Mind. *Journal of Research and Education Chemistry*, *6*(1), 59. https://doi.org/10.25299/jrec.2024.vol6(1).17430

How to cite this article:

Juita, M., Lastuti, S., Hairunisa., Muslim., & Hidayad, A. (2025). Development of Higher Order Thinking Skills (HOTS) Test Items in IPAS to Evaluate Fifth-Grade Students' Problem-Solving Abilities. *Journal of Educational Sciences*, *9(5)*, 3336-3351.