VALIDITY AND RELIABILITY OF HIGHER ORDER THINKING TEST INSTRUMENT DEVELOPMENT SKILLS : SYSTEMATIC LITERATURE REVIEWS

Teguh Febri Sudarma[1], Irfandi[1], Festiyed[2], Yuni Adha[2], Yerimadesi[2], Hefi Alberida[2]

[1]*Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Jalan Williem Iskandar Pasar V Medan Estate 20221, Sumatera Utara, Indonesia*
[2]*Universitas Negeri Padang, Jalan Prof. Dr. Hamka, Air Tawar Padang, Sumatera Barat, Indonesia*

*Corresponding Author: teguhfebri@unimed.ac.id

ARTICLE INFO:

ABSTRACT

Higher Order Thinking Skill (HOTS) refers to the ability to think beyond simple recall, restatement, or direct reference. HOTS questions in the context of assessments measure the capacity to transfer concepts, process and apply information, identify connections among various pieces of information, utilize information to solve problems, and critically evaluate ideas and information. For this study, a total of 17 articles were utilized, comprising 14 national articles and 3 international articles. These seventeen articles revolved around the development of the Higher Order Thinking Skill Test Instrument. The test instrument that was developed incorporates Higher Order Thinking Skills. Based on the findings of the meta-analysis research, it can be concluded that the validity value of the seventeen journals is 1.85, falling into the high category. The reliability value ranges from 0.615 to 0.94, also in the high category. The development of the instrument, in terms of the level of study, primarily focuses on junior and senior high school levels.

INTRODUCTION

The 21st century was marked by rapid developments in various sectors, both in the economic, communication, education, and technology sectors. Among these several sectors, the education sector is a sector that needs attention, because it is the initial process of forming a person's personality so that it is able to bring about a change (Fatimah & Pahlevi, 2020). (Faridah & others, 2019) stated that education is a global challenge to create quality human resources, to prepare a creative Indonesia in 2045. According to (Pratiwi & Fasha, 2015) 21st century learning requires students to have the ability to think creatively, think critically and solve problems. One of the government's efforts to improve the quality of education in Indonesia is to revise the curriculum that has been implemented. Apart from that, the improvement of the curriculum in Indonesia was carried out as an effort to develop Indonesian education at an international level (Sara, Suhendar, & Pauzi, 2020). Currently, the curriculum used in Indonesia. is the revised 2013 curriculum which was revised in 2017 ago. The revised 2013 curriculum focuses on improving assessment standards and content standards. The content standards are carried out by reducing irrelevant material, while the assessment standards are carried out by adapting various international standard assessment models (Hanifah, 2019). Learning assessment applied in learning focuses on students' higher-order thinking skills (Higher Order Thinking Skills). High-level thinking skills HOTS are high-level thinking skills that not only require the ability to remember but also require other higher abilities (Hamidah, 2018). Meanwhile (Lewis & Smith, 1993) states that HOTS is a skill or ability possessed by students in receiving new information which can then produce new information. (Dinni, 2018) also stated that HOTS is the ability to solve problems in new situations by connecting, manipulating, changing knowledge and experience that has been owned critically and creatively.

Higher Order Thinking Skill (HOTS) is the ability to think that does not just remember (recall), restate (restate), or refer without processing (recite). HOTS questions in the context of assessment measure the ability to transfer one concept to another, process and apply information, find connections from different types of information, use the information to solve problems, and examine ideas and information critically.

Even so, the HOTS-based questions do not mean the questions are more difficult than the recall questions. Judging from the scientific dimension, HOTS questions generally measure the metacognitive dimension, not just the factual, conceptual, and procedural dimensions. The metacognitive dimension describes the ability to connect several different concepts, interpret, solve problems (Problem-Solving), choose problem-solving strategies, find (discover) new methods, argue (Reasoning), and make the right decisions.

In Permendikbud number 21 of 2016 concerning content standards for primary and secondary education it is explicitly stated that learning outcomes in the realm of knowledge follow Bloom's taxonomy which has been revised by Lorin Anderson and David Krathwohl (2001) consisting of abilities: knowing (C1), understanding (Understanding -C2), apply (applying-C3), analyze (analyzing-C4), evaluate (evaluating-C5), and create (Creating-C6). The thought process can be seen in the following figure.
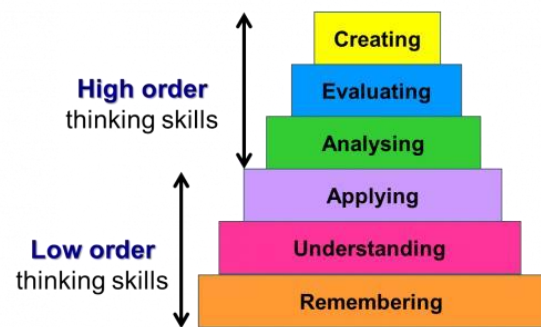

Figure 1. Cognitive Thinking Process in Bloom 's Taxonomy

Then the dimensions of the cognitive process are grouped by Puspendik into 3 levels. Level 1(LOTS): C1 knows and C2 understands, Level 2(MOTS): C3(implements), and Level 3(HOTS): C4(Analyzes), C5(evaluates), and C6(creates).

In the context of writing HOTS questions, the verb "to determine" could be in the realm of evaluating) if planning is preceded by a thought process analyzing the information presented in the stimulus and then students are asked to make the best decision. Even the verb "to determine" can be classified as C6 (to create) if the question requires the ability to develop new problem-solving strategies. So, the realm of operational verbs (KKO) is strongly influenced by what thought processes are needed to answer the questions given.

Thus, HOTS questions are educational evaluation models that test higher-order thinking skills. HOTS questions will hone students' logic, critical thinking, and creativity. HOTS questions are

able to invite students to connect the dots, linking one material to another to construct new knowledge.

METHOD

The research method used in this study is meta-analysis. The method in meta-analytic research examines articles of both national and international research studies. The meta-analytic research method was first pioneered by Glass (1976). The meta-analysis research method is quantitative. This method combines research results through data analysis by calculating numbers and statistics. The data collection technique in this study is an observation technique in research study articles that will be examined through Google Scholar. The instrument used in this study in collecting data was an observation sheet equipped with data coding (coding). Coding is an important requirement to facilitate researchers in collecting and analyzing data. The sample of this research is seventeen article titles from https://scholar.google.co.id/. Data collection steps include (1) identification of research variables. Once found, enter the appropriate variable column, (2) identify the mean content validity for each subject/sub-study, (3) identify the mean validity, (4) identify the mean differential power, (5) identify the average level of difficulty, (7) identify the average reliability, and (8) calculating the final average level of validity, discriminating power, level of difficulty, and reliability of the journal and the average.

The word valid is often interpreted correctly, correctly, validly, and validly. Valid means that the instrument (in Module development research using the Guided Inquiry module) can be used to measure what should be measured. Analysis of suggestions and validation sheets from experts and practitioners is used as a basis for improvement or revision. This aims to get a valid product. In this study, what is seen is the result of the validity that has been carried out by the researcher who is the subject of the study. The results of the validity were analyzed and then obtained results that were in accordance with the research objectives. Next is power difference. The calculation of the discriminating power of a problem is a measurement of the extent to which a question can distinguish students who already understand the material well from students who still do not or lack mastery of the material. If a question has a proportional level of difficulty, it means that the question is good because it is not too difficult but not too easy (Zenal, 2013). One of them is to calculate the level of difficulty of the questions by using the proportion of correct answers (proportion correct).

RESULTS AND DISCUSSION

In this study, 17 articles were used which included 14 national articles and 3 international articles. The seventeen articles discussed the development of the Higher Order Thinking Test Instrument Development instrument skills. The developed test instrument contains Higher Order Thinking skills. Based on the research objectives, the research results obtained are the value of validity, discriminatory power, level of difficulty, and reliability. The following is the distribution of the Higher Order Thinking research sample skills.

Table 1. Data Validation and Reliability

No	Name	validity	Rehabilitation
1	(Utama, Nurkamto, & others, 2020)	0.8	0.75
3	(Rintayati, Lukitasari, & Syawaludin, 2021)	0.92	0.801
4	(Effendi & Andromeda, 2022)	0.8	0.85
5	(Noryandi, Sitompul, & ..., 2022)	0.905	0.72529
6	(Kristiansi, Susilaningsih, Sumarni, & Priatmoko, 2022)	0.73	0.94
7	(Ananda, Kartono, & Ghasya, n.d.)	0.6	0.77
8	(Baddu & Arsyad, 2022)	4.08	0.83
9	(Sari, Pasani, & Hidayanto, n.d.)	3.49	
10	(Hidayah, Tureni, Shamdas, & others, 2022)	0.85	0.9
12	(Indrawati & Elawati, 2022)	4.08	0.83
13	(Rangkuti, Nasution, Hasibuan, Afrida, & Ritonga, 2022)	3,9	
14	(Asri, Sukmawati, & ..., 2022)	3,444	0.615
15	(Kosasih, Asmawati, & Suhendar, 2023)	0.95	0.717
16	(Maulina, 2022)	0.361	0.83
17	(Aryadi & Margunayasa, n.d.)		0.686

From the table, the results of the first study were obtained, namely the validity value of each journal. According to Santyasa (2014), validity is important for developing learning outcomes tests because a score is lacking or does not even reflect learning outcomes if the instrument is unable to measure what students have learned. In addition, validity is useful for seeing the suitability of the instrument with the competence of learning outcomes, accurate material content, and appropriate material in terms of modernity (Wedyawati & Lisa, 2018). Validity values include content, construct, language and external validity. The validity value of each journal ranges from 0.39 to 4.08. The average value of the validity of the twelve journals is 1.85 in the high category. Thus, it can be concluded that the HOTs test instrument developed is valid with a high category. The research results can be seen in Figure 1.
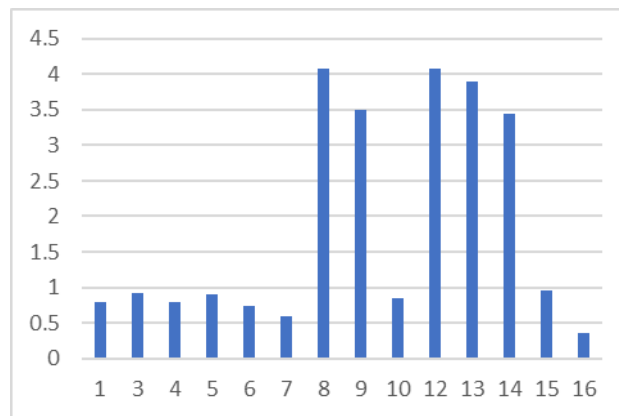

Figure 1. Graph validity of each each journal

The results of the research on the reliability value of each journal. Reliability according to Supardi (2016) is defined as the determination that if the items of the assessment instrument are used to conduct multiple assessments the results are relatively the same or constant, meaning that after the results of the first test with the next test are correlated there is a significant correlation. The reliability value of each journal ranges from 0.615 to 0.94.
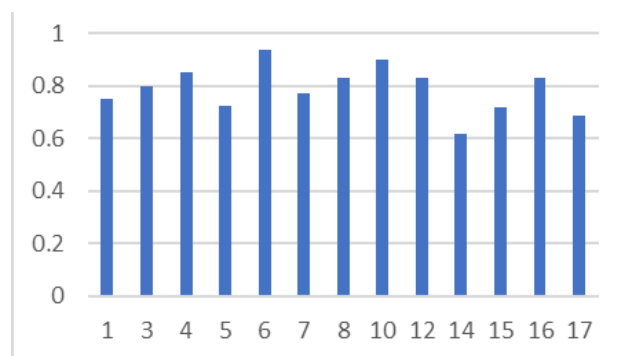

Figure 2. Reliability Graph respectively each journal

T

he results of research on educational levels and subjects in the table 2.

Table 2. Data analysis of Education Levels and Subject

No	Year	Educational level	Subjects
1	2020	SENIOR HIGH SCHOOL	Biology
3	2021	SD	Indonesian
4	2022	SENIOR HIGH SCHOOL	Chemical
5	2022	JUNIOR HIGH SCHOOL	IPA
6	2022	SENIOR HIGH SCHOOL	CHEMICAL
7	2022	SMP / MTs	Mathematics
8	2022	JUNIOR HIGH SCHOOL	Mathematics
9	2022	JUNIOR HIGH SCHOOL	Mathematics
10	2022	SENIOR HIGH SCHOOL	Biology
11	2022	JUNIOR HIGH SCHOOL	IPA
12	2022	JUNIOR HIGH SCHOOL	Mathematics
13	2022	MTs	Mathematics
14	2022	JUNIOR HIGH SCHOOL	Mathematics
15	2022	JUNIOR HIGH SCHOOL	IPA
16	2022	JUNIOR HIGH SCHOOL	PIE
17	2022	JUNIOR HIGH SCHOOL	IPA

As seen in wtable 2, there is a lot of research on the Development of Higher Order Thinking Test Instruments Skills from 2020 to 2022 obtained the development of the HOTs test at the junior high school level with a larger number of 15 titles followed by high school with 3 titles. Instrument development for mathematics subjects obtained 6 research titles followed by 4 science and chemistry 2 research titles.

CONCLUSION

Based on the results of the meta-analysis research, it can be concluded that the validity value of the seventeen journals is 1.85 in the high category. Reliability value 0.615 to 0.94 with high category. For development based on study level, it still piles up at the junior and senior high school levels. then it is still possible to make a high- order thinking test instrument skills at the elementary school level.

REFERENCES

Ananda, D., Kartono, K., & Ghasya, D. A. V. (n.d.). PENGEMBANGAN INSTRUMEN SOAL HIGH ORDER THINKING SKILLS PADA PEMBELAJARAN TEMATIK KELAS V SEKOLAH DASAR. Jurnal Pendidikan Dan Pembelajaran Khatulistiwa, 11(2).

Aryadi, K. S., & Margunayasa, I. G. (n.d.). Instrumen Penilaian High Order Thinking Skills (HOTS) pada Pembelajaran IPA.

Asri, A., Sukmawati, S., & … (2022). PENGEMBANGAN TES SUBJEKTIF BERBASIS HIGHER ORDER THINKING SKILL (HOTS) PADA MATERI POLA BILANGAN. SIGMA: JURNAL …. Retrieved from https://journal.unismuh.ac.id/index.php/sigma/article/view/9323

Baddu, N. R., & Arsyad, N. (2022). PENGEMBANGAN INSTRUMEN TES HIGHER ORDER THINKING SKILL (HOTS) UNTUK MATERI STATISTIKA DI KELAS IX SMP. Infinity: Jurnal Matematika Dan …. Retrieved from https://science.e-journal.my.id/ijma/article/view/117

Dinni, H. N. (2018). HOTS (High Order Thinking Skills) dan kaitannya dengan kemampuan literasi matematika. PRISMA, Prosiding Seminar Nasional Matematika, 1, 170–176.

Effendi, Z., & Andromeda, A. (2022). Pengembangan Instrumen Tes Two-Tier Higher Order Thinking Skill (HOTS) Berbasis IBT Pada Materi Kesetimbangan Kimia untuk Siswa SMA/MA. Entalpi Pendidikan Kimia. Retrieved from http://entalpipendidikan.ppj.unp.ac.id/index.php/epk/article/view/264

Faridah, E. M. I., & others. (2019). Analisis Kemampuan Berpikir Kritis Siswa Melalui Soal-Soal HOTS (Higher Order Thinking Skills) Mata Pelajaran Sejarah Kelas X-IPS SMAN 2 SIDOARJO. Avatara, 8(2).

Fatimah, S., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Berbasis HOTS (Higher Order Thinking Skills) Pada Kompetensi Dasar Menerapkan Sistem Penyimpanan Arsip Sistem Abjad, Kronologis, Geografis, Nomor, dan Subjek di Jurusan OTKP SMKN 1 Bojonegoro. Jurnal Pendidikan Administrasi Perkantoran (JPAP), 8(2), 318–328.

Hamidah, L. (2018). Higher Order Thinking Skills: Seni Melatih Kemampuan Berpikir Tingkat Tinggi. Temanggung: Desa Pustaka Indonesia.

Hanifah, N. (2019). Pengembangan instrumen penilaian Higher Order Thinking Skill (HOTS) di sekolah dasar. Current Research in Education: Conference Series Journal, 1(1), 5.

Hidayah, N., Tureni, D., Shamdas, G. B. N., & others. (2022). Analisis Soal Ujian Biologi Berbasis HOTS (Higher Order Thinking Skill) di SMAN 9 Palu. Journal of Biology Science and Education, 10(1), 59–65.

Indrawati, N., & Elawati, E. (2022). Pengembangan Instrumen Tes Higher Order Thinking Skill (HOTS) Pada Materi Statistika SMP. Jurnal …. Retrieved from http://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/8994

Kosasih, K., Asmawati, L., & Suhendar, S. (2023). PENGEMBANGAN PENILAIAN MATERI IPA SMP BERBASIS HIGHER ORDER THINKING SKILL (HOTS). JTPPm (Jurnal Teknologi Pendidikan Dan Pembelajaran): Edutech and Intructional Research Journal, 9(2).

Kristiansi, F., Susilaningsih, E., Sumarni, W., & Priatmoko, S. (2022). Pengembangan Butir Soal Tes Tingkat Makroskopis, Mikroskopis, dan Simbolis Untuk Identifikasi Kemampuan Analisis Siswa yang Dianalisis dengan Model Rasch. Chemistry in Education, 11(1), 49–56.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. Theory into Practice, 32(3), 131–137.

Maulina, W. (2022). Pengembangan Instrumen Asesmen Berbasis Hots pada Mata Pelajaran Pendidikan Agama Islam dan Budi Pekerti di SMAN 6 Padang. Jurnal Pendidikan Dan Konseling (JPDK), 4(6), 7317–7326.

Noryandi, N., Sitompul, S. S., & ... (2022). PENGEMBANGAN TES HIGHER ORDER THINKING SKILL (HOTS) TENTANG PESAWAT SEDERHANA DI SMP PONTIANAK. Jurnal Pendidikan Dan …. Retrieved from https://jurnal.untan.ac.id/index.php/jpdpb/article/view/55476

Pratiwi, U., & Fasha, E. F. (2015). Pengembangan instrumen penilaian hots berbasis kurikulum 2013 terhadap sikap disiplin. Jurnal Penelitian Dan Pembelajaran IPA, 1(1), 123–142.

Rangkuti, R. K., Nasution, A. F., Hasibuan, Z., Afrida, A., & Ritonga, W. A. (2022). Pengembangan Instrumen Tes Kompetensi Sains Madrasah Tsanawiyah Berbasis Hight Order Thingking Skill (HOTS) untuk Meningkatkan Kemampuan Pemahaman Konsep Matematis Siswa. Al-Khawarizmi: Jurnal Pendidikan Matematika, 3(1), 16–20.

Rintayati, P., Lukitasari, H., & Syawaludin, A. (2021). Development of Two-Tier Multiple Choice Test to Assess Indonesian Elementary Students' Higher-Order Thinking Skills. International Journal of Instruction, 14(1), 555–566.

Sara, S., Suhendar, S., & Pauzi, R. Y. (2020). Analisis Higher Order Thinking Skills (HOTS) Siswa Kelas VIII pada Materi Sistem Pernapasan Manusia. Bioedusiana: Jurnal Pendidikan Biologi, 5(1), 52–61.

Sari, R., Pasani, C. F., & Hidayanto, T. (n.d.). Pengembangan Tes Formatif Aritmatika Sosial Berbasis Higher Order Thinking Skills (HOTS) dengan Konteks Lingkungan Lahan Basah. EDU-MAT: Jurnal Pendidikan Matematika. Retrieved from https://ppjp.ulm.ac.id/journal/index.php/edumat/article/view/12927

Utama, C., Nurkamto, J., & others. (2020). The Instrument Development to Measure Higher-Order Thinking Skills for Pre-Service Biology Teacher. International Journal of Instruction, 13(4), 833–848.