



## Development of Test Instruments Based on Revision of Bloom's Taxonomy to Measure the Students' Higher Order Thinking Skills

Dwi Oktaviana, Utin Desy Susiaty\*

IKIP PGRI Pontianak, Indonesia.

\*[d3or4f4ty4@gmail.com](mailto:d3or4f4ty4@gmail.com)

© 2020 JIPM (Jurnal Ilmiah Pendidikan Matematika)

This is an open access article under the CC-BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>) ISSN 2337-9049 (print), ISSN 2502-4671 (online)

**Abstract:** The objectives of this study were to determine: (1) the appropriateness of the comparative material test instrument based on the Revised Bloom Taxonomy to measure HOTS student; (2) the quality of the comparative material test instrument based on the Revised Bloom Taxonomy to measure HOTS student; and (3) analysis of potential effects on the use of test instruments for comparison material based on Revised Bloom's Taxonomy to measure HOTS students' developed. This research is a development research with the Tessmer model development design. The subjects of this research were all students of class VIII SMP Koperasi Pontianak. The research data was obtained based on the expert's / expert's assessment of the test instrument, the results of the test small group which will be analyzed in terms of validity, difficulty index, distinguishing power, and item reliability, as well as the results of the tests field test which will be analyzed related to potential effects. The results showed that: (1) the feasibility of the comparative material test instrument based on the Revised Bloom Taxonomy to measure HOTS students' reached the valid criteria level; (2) the quality of the comparative material test instrument based on the Revised Bloom Taxonomy to measure the HOTS of students reaching the criteria level is valid, difficult, has poor distinguishing power, and is declared to have low feasibility as a research instrument; (3) analysis of potential effects on the use of test instruments for comparison material based on Revised Bloom's Taxonomy to measure HOTS to student achieve an effectiveness level of 68.96% with the criteria for positive potential effects

**Keywords :** Development; Test Instrument; Revised Bloom's Taxonomy; HOTS.

### Introduction

Mathematics is taught at every level of education, from elementary school to high school, even to college. In every mathematics learning process at any level there will always be an evaluation or assessment process. The evaluation process is very important for the development of the quality of education in all countries, especially in Indonesia. This is made clear in (Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, n.d.) states that evaluation is carried out in the context of controlling the quality of education nationally as a form of accountability for providers education to interested parties. In addition, (Sudijono,

2009) stated that the purpose of holding an evaluation is to obtain evidentiary data which will be a guide to where the results of the ability level and success level of students in achieving learning goals after they take the learning process. The evaluation process carried out at the secondary school level ends with a test at the end of each learning material.

The evaluation process cannot be separated from the use of instruments. An instrument plays an important role in capturing learning outcomes (Al-Tabany, 2014). In addition, (Arifin, 2012) states that the instrument has a very important function and role in order to determine the effectiveness of the learning process. The instrument used to determine the effectiveness of the learning process was a test.

(Mulyadi, 2010) argues that the evaluation process includes two things, namely measurement and testing. When conducting evaluations, the educator must take measurements in it and must also use a tool called a test. The test is an instrument for collecting participant data that responds to questions so that participants can demonstrate their maximum ability and mastery. (Nursalam, 2016) states that to determine the level of knowledge, skills, intelligence, talents or abilities of a person can be tested by asking a number of questions. Therefore, giving a test is very important in the evaluation process.

The test is an instrument for collecting participant data that responds to questions so that participants can demonstrate their maximum ability and mastery (Purwanto, 2014). In addition, Hasan in (Arifin, 2012) explained that the test is a specially designed data collection tool. Thus, an educator can design a test instrument with the aim of measuring the abilities of his students.

Making the test cannot be separated from the taxonomic component of bloom. Taksonomi bloom outlines six response rate in thought processes namely: (1) knowledge, (2) comprehension, (3) application, (4) analysis, (5) synthesis, 6) evaluation. The levels in the taxonomy have been used for nearly half a century as the basis for the preparation of educational goals, the preparation of tests and curricula. Revisions were made to the Bloom Taxonomy, where changes in the standards of objects (in Bloom Taxonomy) become the basis for work (in the Revision of the Bloom Taxonomy). This change is made for the purpose of education. verb) with something (noun).

According to Bloom's Revised Taxonomy, cognitive thinking skills can be classified into six categories. The Bloom's Revised Taxonomy (Anderson & Kratwohl, 2010) consist of remember, understand, apply, analyze, evaluate, and create. Therefore, If the teacher's assessment is only quantitative, it cannot be known the extent of the student's thinking process. Meanwhile, to find out the achievement of learning outcomes in students' cognitive processes, it can be used by using the Revised Blom Taxonomy as already mentioned, so that if it is examined further, the matter can also be used as material for consideration to further optimize the learning activities that are taking place in class.

In the Revised Bloom's Taxonomy, the ability to think to analyze, evaluate, and create is a cognitive domain that must be measured in the evaluation process. Mathematics is a subject to form a way of thinking at a high level (analyzing, evaluating, and creating) or Higher Order Thinking. In Mathematics Learning, it is expected that students will be careful in their work, critical in thinking, consistent in being and honest in various situations (Tiro, 2010). This is confirmed in Permediknas No. 22 of 2006, mathematics subjects need to be given to all students starting from elementary school to equip students with the ability to think logically, analytically, systematically, critically, and creatively, and the ability to work together. However, the reality is that the students have not fully achieved the objectives of learning mathematics.

Based on research conducted by researchers at one of the schools, namely SMP Koperasi Pontianak, the results showed that the classification of daily test questions for class VII B SMP

Koperasi Pontianak on the comparison material based on the cognitive domain of Revised Bloom Taxonomy already contains questions with an indicator of understanding or understanding by 20% and indicators of applying or apply for 80% (Oktaviana & Prihatin, 2018). This is also seen in several other studies, namely (Giani, Zulkardi, & Hiltrimatrin, 2012) which examined the cognitive level of mathematics textbook problems. The results showed that the percentage of questions for each cognitive level were: C1 (3.23%), C2 (30.97%), C3 (61.93%), C4 (3.87%), C5 (0%). ), C6 (0%). In addition, research by Amelia, dkk (2015) with the results of research on question criteria on daily tests on the subject of the set with a percentage of 13.3% for the cognitive level of knowledge (C1), 46.7% for the cognitive level of understanding (C2), and 40%. for the application cognitive level (C3). It can be seen from several studies conducted that the learning process of mathematics so far is only giving questions to students with a low level of thinking ability (remembering, understanding, and applying). Therefore, the researcher continues the research that has been done by developing a test instrument based on the revised bloom taxonomy to measure higher order thinking skills or Higher Order Thinking Skills. The instrument developed is an instrument that is completely new as a result of the researcher's own development.

Researchers realize that with the development of this test instrument helps all teachers in compiling and developing math problems at the level of analyzing, evaluating, and creating which include higher order thinking questions. Therefore, this research and development takes the title Development of Comparative Material Test Instruments Based on Revised Bloom's Taxonomy to Measure Students' High-Level Thinking Ability.

## **Method**

The research method used in this research is the method of research and development (R&D). The research design was in atype model formative research by Tessmer. The design developed by Tessmer is a formative evaluation development design which consists of 2 stages, namely stage I: thestage, whichpreliminaryis the initial step in this research where the researcher begins this research by conducting a preparatory analysis by determining the place and research subject by contacting the head. school and teacher of mathematics at the school who will be the location of the research as well as holding other preparations, stage II: the stage of self-evaluation(self-evaluation)is the stage of curriculum analysis and preparation of design covers, stage formative evaluation(formativeevaluation),expert judgment / expert(expertreviews)is the result of designing issues higher level thinking and guide the interview as a prototype I consulted with specialists / experts for validation which includes content validity, evaluation of one-on-one(onetoone)that trials conducted one-on-one with provides a comparative test for measuring HOTS,the evaluation group(smallgroup)that the tests on the student small groups(smallgroup)in the eighth grade students in junior cooperatives by providing comparison tests, and field tests(field oftest)the trials conducted in the eighth grade students besides the student group small by providing a comparative test in measuring HOTS. This study was used to develop a comparative test to measure HOTS. The subjects of this research were all students of class VIII SMP Koperasi Pontianak. The data collection technique was done using expert validation sheets, tests with research instruments in the form oftests comparisonsto measure HOTS and interviews.

The test referred to in this study is a comparative test in measuring HOTS to find out data on Higher ability Order Thinking Skills (HOTS students) 'to comparison material where the form of the test used is a comparative test in the form of a description with a total of 5 questions

containing aspects of thinking ability high level. The validation sheet is used to obtain data about the validity of the test instrument developed. Where there are several aspects that will be validated on the expert validation sheet including: validation of content, constructs, and language. Unstructured Interviews were used to obtain data about the potential effects of the test on abilities Higher Order Thinking Skills (HOTS students<sup>1</sup>) on comparison material.

## **Result and Discussion**

The process of developing a test instrument based on the revised bloom taxonomy to measure HOTS students' uses the development of the Tessmer model. The model development consists of two phases, namely Phase I: preliminary, stage II: the stage of formative evaluation that includes a self-evaluation, expert reviews, evaluation of one to one, small group evaluation, and field test. This refers to the initial aim of this study is to develop a test instrument that is feasible, of quality and has potential effects.

The preliminary stage of preparation begins by conducting a preparatory analysis by determining the place and subject of research by contacting the principal and teachers of mathematics subjects in schools that will be used as research locations and making other preparations, such as arranging research schedules and procedures for collaboration with teachers mathematics which will be used as a place of research. After doing all the preparatory analysis, a research place was obtained, namely the SMP Koperasi Pontianak because there were many problems, especially in problem solving based on HOTS. Based on interviews with mathematics teachers at the school, this was due to students who were not familiar with questions HOTS. Never mind solving the problem HOTS, for problem solving problems within the scope of the lowest 3 aspects of the bloom taxonomy (C1, C2, and C3), students still have difficulties. For this reason, researchers feel the need to develop a test instrument based on the revised bloom taxonomy to measure higher order thinking skills. That way, the researcher will find out the extent of the HOTS student's in order to improve students' problem-solving abilities to a higher level.

Furthermore, there are five steps that must be performed on the stage of the formative evaluation (formative evaluation), namely, the self-evaluation (self-evaluation), the valuation expert / specialist (expert reviews), evaluation of one-on-one (one to one), the evaluation group (small group), and field test (field test). Phase evaluation (self-evaluation) begins by analyzing curriculum that aims to assess the core competencies and core competencies which refers to the syllabus has been prepared which will be used as the basis for determining the amount of the item or items in making lattice is composed test. Materi Tests based on the 2013 Curriculum. For the development of the test HOTS, researchers chose a material, namely comparison. Comparative material test instruments compiled can be used to measure abilities HOTS students'. The researcher compiled the test by referring to the three aspects HOTS, namely analyzing, evaluating and creating. Researchers also designed test grids, comparative questions to measure HOTS, and interview guides. The test grid design includes writing basic competencies, subject matter, indicators, time allocation, and test forms based on higher order thinking criteria. The interview guide design includes interview problems, interview objectives, interview implementation steps, and interview questions.

The next stage is the expert reviews. At this stage the design results of higher order thinking questions and interview guidelines as prototype I were consulted with experts / experts to be validated which included content validity. In this study, the validation carried out was content validation. The content validation was carried out by three mathematics lecturers at

the IKIP PGRI Pontianak. The three experts provide an assessment based on an assessment sheet prepared according to the BSNP assessment guide. Regarding content validation, there are four criteria that are assessed, including: content feasibility, presentation feasibility, language assessment, assessment HOTS. The validation results are matched with the validity criteria can be seen in Table 1 below.

**Table 1.** Validation Criteria for Test Instruments

<b>Average</b>	<b>Decision</b>
$85\% < \bar{X} \leq 100\%$	Very valid, or can be used without revision
$70\% < \bar{X} \leq 85\%$	Quite valid, or can be used but needs minor revision
$50\% < \bar{X} \leq 70\%$	Less valid, it is recommended not to be used because it needs major revision
$0\% < \bar{X} \leq 50\%$	Not valid, or may not be used

Adapted from (Hidayati, 2016)

The results from the validation of the experts are presented in Table 2.

**Table 2.** Results of Validation The Instrument HOTS by Material Expert

<b>No</b>	<b>Ahli</b>	<b>Persentase</b>	<b>Keterangan</b>
1	Material Expert I	89,70%	Very Valid
2	Material Expert II	86,67%	Very Valid
3	Material Expert III	78,18%	Enough Valid
<b>Average</b>		84,85 %	Enough Valid

Based on tables 1 and 2, the average value obtained from the results of the assessment by material experts is 84.85% with a fairly valid category. So that the test instrument is categorized as feasible.

Furthermore, the researchers conducted an evaluation stage one to one. At this stage, a one-on-one trial will be carried out by providing a comparative test to measure HOTS. Number of Students The testers are people. At this stage it can be seen that the comments of students as testers are only focused on the problem solving process. However, no substantial comments were made. So it can be concluded that at this stage the instrument of comparison test questions did not experience significant changes and could be continued to the next stage of research, namely the evaluation of the small group. At this stage the tests on a small group of students (small group) in the eighth grade students in junior Cooperative By providing a comparison test. The purpose of giving this test is to see the quality of the test instruments including validity, difficulty index, distinguishing power, and reliability of test items.

Based on the results of the trials that have been carried out, the results of the validity analysis of each question are shown in the following table:

**Table 3.** Results of the Validity Analysis of Test Items

No. Question	$r_{xy}$	Information
1	0,5675	Moderate
2	0,5416	Moderate
3	0,6320	High
4	0,5416	Moderate
5	0,5915	Moderate

Interpretation of the correlation coefficient  $r_{xy}$  used the following criteria.

$0,80 < r_{xy} \leq 1,00$ : very high
$0,60 < r_{xy} \leq 0,80$ : high
$0,40 < r_{xy} \leq 0,60$ : moderate
$0,20 < r_{xy} \leq 0,40$ : low
$r_{xy} \leq 0,20$ : very low

(Jihad & Haris, 2010)

Based on table 3 and the criteria for the correlation coefficient, it shows that all test questions are declared valid. Furthermore, the results of the test difficulty index analysis are shown in the Table 4.

**Table 4.** Results of the Index Analysis of Test Items

No. Question	$\bar{S}$	$S_{maks}$	Index of Difficulty	Information
1	2.6857	12	0,2238	difficult
2	2.8286	12	0,2357	difficult
3	2.7429	12	0,2286	difficult
4	2.7143	12	0,2262	difficult
5	2.7143	12	0,2262	difficult

The criteria for the interpretation of the difficulty level of the items.

- 0,00 – 0,30 : difficult
- 0,31 – 0,70 : moderate
- 0,71 – 1,00 : easy

(Jihad & Haris, 2010)

Based on Table 4 and the item difficulty level criteria, it shows that the analysis of the item difficulty index on the research instrument is classified as difficult.

The results of the calculation of the distinguishing power index from the test results are shown in the Table 5.

**Table 5.** Results of Distinguishing Power Index Analysis of Test Items

No. Question	Power Index differentiator	Description
1	0.0259	Bad
2	0.0313	Bad
3	0.0319	Bad
4	0.0161	Bad
5	0.0305	Bad

The criterion for the interpretation of distinguishing items.

- 0,4 or more : very good
- 0,30 - 0,39 : good enough, may need to be fixed
- 0,20 - 0,29 : minimum needs to be repaired
- 0,19 and below : bad, discarded or overhauled

(Jihad & Haris, 2010)

Based on Table 5 and the criteria for distinguishing between items, all questions are classified as having poor distinguishing power, so the problem needs to be discarded or overhauled. Based on the results of the reliability analysis that has been done, the reliability value is obtained  $r_{11} = 0.4844$ . With the following reliability criteria.

- $r_{xy} \leq 0,20$  : reliability very low
- $0,20 < r_{xy} \leq 0,40$  : reliability low
- $0,40 < r_{xy} \leq 0,70$  : reliability moderate
- $0,70 < r_{xy} \leq 0,90$  : reliability high
- $0,90 < r_{xy} \leq 1,00$  : reliability very high

(Jihad & Haris, 2010)

Based on the results and reliability criteria, it can be concluded that the reliability of the questions is in the low category. Based on the results of the item analysis, it shows that all the test items are declared valid, the analysis of the item difficulty index on the research instrument is classified as difficult, all the questions are classified as having poor differentiation, and are

declared to have low feasibility as a research instrument. This can happen because the test instrument contains 3 aspects of higher-order thinking skills, namely analyzing, evaluating and creating. And this shows that the ability is HOTS students' still low, which is shown from the results of the item analysis.

In the final stage of this study, a trial test was conducted on students of class VIII other than small group students by giving this research is to HOTS. The giving comparative tests in measuring aim of obtain data about the potential effects of tests on HOTS student. In addition, interviews were conducted with 3 students in the class field test who represented the field test class. This interview aims to classify and verify the effect of potential on HOTS tests students' and information is obtained that the instrument test questions given enable students to train and develop their abilities HOTS even though they are not aware of it. Even though the questions given were categorized as very difficult questions for these students, the students' enthusiasm in finding and solving the problems presented in the test instrument indicated that they indirectly developed their abilities HOTS.



Figure 1. Implementation of the Test Instrument Trial

## Conclusion

Based on the results of the research and discussion, it can be concluded that: (1) the appropriateness of the comparative material test instrument based on the Revised Bloom Taxonomy to measure HOTS to student reach the valid criteria level; (2) the quality of the comparative material test instrument based on the Revised Bloom Taxonomy to measure the HOTS of students reaching the criteria level is valid, difficult, has poor distinguishing power, and is declared to have low feasibility as a research instrument; (3) analysis of potential effects on the use of comparative material test instruments based on Revised Bloom's Taxonomy to measure HOTS student reaching an effectiveness level of 68.96% with the criteria for positive potential effects.

## References

- Al-Tabany, T. I. B. (2014). *Mendesain Model Pembelajaran Inovatif, Progresif, dan Kontekstual*. Jakarta: Prenadamedia Group.
- Anderson, & Kratwohl. (2010). *Kerangka Landasan untuk Pembelajaran Pengajaran dan Asesmen (Cetakan ke-1)*. Yogyakarta: Pustaka Belajar.
- Arifin, Z. (2012). *Penelitian Pendidikan Metode dan Paradigma Baru*. Bandung: Remaja Rosdakarya.



- Giani, Zulkardi, & Hiltrimatrin, C. (2012). Analisis Tingkat Kognitif Soal-Soal Buku Teks Matematika Kelas VII Berdasarkan Taksonomi Bloom. *Jurnal Pendidikan Matematika*, 66, 37–39.
- Hidayati, N. (2016). Pengembangan Perangkat Pembelajaran Ipa Topik Energi Dalam Sistem Kehidupan Di Madrasah Tsanawiyah. *JINoP (Jurnal Inovasi Pembelajaran)*, 2(2), 389. <https://doi.org/10.22219/jinop.v2i2.3283>
- Jihad, A., & Haris, A. (2010). *Evaluasi Pembelajaran*. Jakarta: Multi Press.
- Mulyadi. (2010). *Pengembangan Model Evaluasi Pendidikan Agama di Sekolah*. Malang: UINMALIKI Press.
- Nursalam, N. (2016). Diagnostik Kesulitan Belajar Matematika: Studi Pada Siswa Sd/Mi Di Kota Makassar. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah Dan Keguruan*, 19(1), 1–15. <https://doi.org/10.24252/lp.2016v19n1a1>
- Oktaviana, D., & Prihatin, I. (2018). Analisis Hasil Belajar Siswa Pada Materi Perbandingan Berdasarkan Ranah Kognitif Revisi Taksonomi Bloom. *Buana Matematika: Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, 8(2), 81–88. [https://doi.org/10.36456/buana\\_matematika.8.2.1732.81-88](https://doi.org/10.36456/buana_matematika.8.2.1732.81-88)
- Purwanto. (2014). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Sudijono, A. (2009). *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Pers.
- Tiro, A. M. (2010). *Cara Efektif Belajar Matematika (Cetakan 1)*. Makassar: Andira Publisher.
- Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional. (n.d.). *Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional*.