Development of TIMSS-Type Mathematical Test Instruments to Train Higher Order Thinking Skills (HOTS) Grade V Students

Rafi Syafitri 1,*, Yuniawatika 1, Sutansi 1

1 Department of Elementary School Teacher Education, Faculty of Education, Universitas Negeri Malang, Malang, Indonesia
*Corresponding author. Email: rafi.syafitri@gmail.com

ABSTRACT
TIMSS is a large-scale assessment of mathematics and science that demands high levels of thinking skills in students in solving existing problems. The results of Indonesian students’ participation in TIMSS show that there is still a low level of thinking ability in elementary school students. This is because students are not used to receiving non-routine problems that in their completion require high levels of thinking skills. To overcome this, a valid, practical and qualified TIMSS-type mathematical test instrument was developed to train the ability of high-level students to think. This research was conducted with Borg and Gall development model from the first step to step five. The development steps are data collection, planning, product development, expert validation, revision, and product trials. This TIMSS-type mathematical test instrument is declared valid, practical and qualified and can be used to train students to conduct high order thinking skills (HOTS).

Keywords: test instrument, TIMSS, HOTS

1. INTRODUCTION
Based on the standard content of the 2013 curriculum, mathematics is one of the subjects that students must learn at the elementary school level. In mathematics subjects the materials studied are mutually sustainable. The statement is in line with Yuniawatika opinion [1] stating that mathematics is an interconnected and connect discipline Learning mathematics not only teaches about concepts, knowledge and counting skills, but also teaches students to do reasoning in solving problems encountered. According to Hanifah & Manoy [2] the questions contained in TIMSS (Trends in International Mathematics and Science Study) can be used to develop problem solving skills in students with varying degrees of difficulty.

TIMSS is an international institution tasked with measuring students’ level of education in the realm of mathematics and science. Mullis [3] stated that Indonesia has followed the assessment held by TIMSS since 1999 until now. Based on this, Indonesia’s participation in TIMSS is one form of effort to measure the quality of Indonesian education compared to other countries in the world. Based on the results of the TIMSS survey, conducted by the International Association for the Evaluation of Education Achievement (IAE) [3] stated that the mathematics achievement of Indonesian students is relatively low. The low achievement of Indonesian students in the TIMSS program is due to students being poorly trained in working on TIMSS model questions.

Based on the results of interviews with VA classroom teachers at MIN 13 Blitar found that the teacher did not know about the TIMSS model test. Teachers state that the need for questions that can stimulate students’ high-level thinking skills is urgently needed, but in reality, these questions are rarely encountered in learning or in textbooks. In line with this, the results of interviews with students show that students have never had a problem of this type of TIMSS evidenced when students are shown about TIMSS in 2015 students say that they have never done a problem like this and students find it difficult when working on the TIMSS question.

To determine the success rate of students in learning, teachers use test instruments as evaluation tools in learning. The definition of a test instrument according to
Wulan [4] is an assessment tool in the form of writing to observe the level of progress of student achievement in learning. Each test instrument made has certain conditions in order to be declared a good test instrument. According to Akbar [5] the question of a good test must qualify as follows (a) validity, which can be interpreted as the validity, truth or validity of a measured thing, in this case meaning the validity of the test instrument used. A test instrument can be said to be valid if it has a validity index of 0.401-1.00 according to validity criteria, to determine the success rate of students in learning, teachers use test instruments as evaluation tools in learning. The definition of a test instrument according to Wulan [4] is an assessment tool in the form of writing to observe the level of progress of student achievement in learning.

Each test instrument made has certain conditions in order to be declared a good test instrument. According to Akbar [5] the question of a good test must qualify as follows: (a) validity, which can be interpreted as the validity, truth or validity of a measured thing, in this case meaning the validity of the test instrument used. A test instrument can be said to be valid if it has a validity index of 0.401-1.00 according to validity criteria, (b) reliability, a test instrument can be declared to have high reliability if the test results are fixed, meaning not the results on each test should be the same, but the results follow changes, and (c) objectivity, if in the implementation of the work of the test instrument there are no subjective factors affecting, according to Rahayu [6] there are two factors that affect subjectivity namely the form of tests and assessors, (d) practicality, test instruments must have practical properties, meaning test instruments should be easy to use, easy to examine, and equipped with working instructions, (e) economical, good test instruments do not have to cost a lot of money, a lot of power and a long time in the manufacture.

TIMSS is a large-scale assessment designed to inform educational policies and practices by providing an international perspective on teaching and learning in mathematics and science. According to Provasnik [7] the goal of TIMSS is to measure the achievement of 4th and 8th graders in math and science materials. In TIMSS there are 2 domains that are testing frameworks that are cognitive domains and content domains. Mullis [3] states that the cognitive domain in TIMSS is based on 3 sub domains namely knowledge, application and reasoning. The domain of content in TIMSS is a test in the realm of mathematical materials. Mullis [8] states that in the content domain there are 3 topics of grade IV SD/MI mathematical material namely numbers, geometry shape and measures, and data display presentation. Based on Highlights from TIMSS and TIMSS Advanced 2015 Mullis [9] states that students’ abilities are categorized into 4 categories namely low, medium (intermediate), high, and very high (advance).

High-level thinking ability is important for the future of students, given that it prepares students to face many challenges that will arise in their lives, careers and at the level of their personal obligations and responsibilities. According to Rofiah [10] HOTS is a deeper and broader use of the mind to discover new challenges. Similar to that opinion Heong [11] states that HOTS or high-level thinking ability is thinking at a higher level, not just memorizing facts or saying something to someone exactly as heard but is a mental activity in an effort to explore complex experiences. Based on these two opinions it can be concluded that a high level of thinking ability is a person’s ability to think in solving problems at a higher level and in many ways’ completion. In the 2013 curriculum there is a HOTS stage that is stated by Ariyana [12] namely: (1) transfer of knowledge which is the stage at which students can identify and analyze a material, (2) problem solving that students can do problem solving by using a certain way, and (3) critical and creative thinking that students can think critically and string one material into another material that is interconnected.

2. METHOD

This research uses this type of research and development or research and development (R&D). The development model used is the Borg and Gall development model. According to Borg and Gall [13] there are ten steps to implementing research and development strategies. However, in research is only used up to the fifth step with the first stage of research data collection, planning, product development, field trials, and product revisions.

In the first step, researchers conducted data collection through observations, interviews, identification of basic competencies, as well as library studies on TIMSS, HOTS, and mathematical instruments. The second step is for researchers to plan the development of instruments in the form of establishing material coverage, test indicators, and setting the type and number of questions. In the third step, researcher began to develop products that have been designed to include the creation of test work instructions, TIMSS-type math problems, key answers and discussions, incorporating questions into wonder share applications, as well as product validation in media, materials, and language experts. The fourth step is a field trial divided into 2, namely one-to-one trials and small-scale trials. In the fifth step, the researcher revises the product based on criticism and advice given by the validator.

The data obtained in this development is quantitative and qualitative data. The instruments used as data collectors in the development of this module are validation instruments of material experts, media experts, users, namely teachers in the form of assessment questionnaires and questionnaires working on
mathematical test instruments of TIMSS type by students. Data analysis techniques used to analyze validation results data by experts or experts and readability tests by students are with average analysis techniques.

3. RESULT

In the validation results by material experts in Table 1 obtained the results of the calculation of validity shows that the product gets a percentage of 95.8%. Thus, based on the validity criteria according to Akbar [5] the intrusion of the K-13 mathematical test of the TIMSS type is declared valid by the material expert. Data validation results from material experts namely test instruments must be adjusted to the basic competencies contained in the curriculum, re-search hots indicators in accordance with the curriculum, improvements in question writing and scrutiny of the content of basic competencies, indicators and questions.

Table 1 Material expert validation results

| No. | Assessment Indicators                        | Amount | Percentage |
|-----|----------------------------------------------|--------|------------|
| 1.  | Material suitability with grids              | 16     | 100        |
|     | Average                                      | 4      | 100        |
|     | Criteria                                     |        | Very Valid |
| 2.  | Material accuracy                            | 14     | 350        |
|     | Average                                      | 3.5    | 87.5       |
|     | Criteria                                     |        | Very Valid |
| 3.  | Higher Order Thinking Skills (HOTS) Presentation Techniques | 11     | 275        |
|     | Average                                      | 3.7    | 91.7       |
|     | Criteria                                     |        | Very Valid |
| 4.  | Presentation techniques                       | 8      | 200        |
|     | Average                                      | 4      | 100        |
|     | Criteria                                     |        | Very Valid |
| 5.  | Linguistics                                  | 16     | 400        |
|     | Average                                      | 4      | 100        |
|     | Criteria                                     |        | Very Valid |
|     | Overall Amount                               | 65     | 1625       |
|     | Overall Average                              | 3.84   | 95.8       |
|     | Overall Criteria                             |        | Very Valid |

In the validation results by media experts obtained the results of the calculation of validity in Table 2 shows that the percentage is 95.9%. Thus, based on the validity criteria according to Akbar [5] the intrusion of the K-13 mathematical test of TIMSS type is declared very valid by media experts. Note the validation results from media experts i.e. the use of font letters should be equalized and also note the density of writing in 1 slide.

Table 2 Media expert validation results

| No. | Assessment Indicators                        | Amount | Percentage |
|-----|----------------------------------------------|--------|------------|
| 1.  | Presentation of text or writing              | 11     | 275        |
|     | Average                                      | 3.7    | 91.7       |
|     | Criteria                                     |        | Very Valid |
| 2.  | Linguistics                                  | 16     | 400        |
|     | Average                                      | 4      | 100        |
|     | Criteria                                     |        | Very Valid |
|     | Overall Amount                               | 27     | 675        |
|     | Overall Average                              | 3.85   | 95.9       |
|     | Overall Criteria                             |        | Very Valid |

In the validation results by the user get the results of the calculation of validity in Table 3 shows that the percentage is 100%. Thus, based on the validity criteria according to Akbar [5] the intrusion of the K-13 math test of TIMSS type is declared very valid by the teacher as a user. The teacher’s validation of the k-13 math test instrument of TIMSS type is very good for stimulating high-level thinking in students, for further development should be adjusted between the material and the ability of the student in general.

From the Figure 1 poll results in the one to one test is known there is 1 indicator that has not reached 100%. The results are then used as a reference to revise the product to better suit the students’ wishes. But from the above poll it is known that both students are challenged with K-13 math test instruments of TIMSS type and can be used to study independently with the intrusion of the math test. After the revision, a small-scale trial was conducted to see how large a K-13-type mathematical test instrument of the TIMSS type could train students to think at a high level. After the initial field test the small scale was known to be the average percentage of students who agreed with the statement of 95.3%. A percentage of 95.3% if interpreted according to Akbar [5] validity criteria then the K-13 mathematical test instrument product of TIMSS type is declared very valid and usable.
4. DISCUSSION

Based on the validation results of the material experts, it can be noted that the product of mathematical test instruments of TIMSS type has been valid and worth using. In the validation results are processed the percentage achieved by 95.8%. Based on these results, the product of mathematical test instrument of TIMSS type is declared valid and can be used. This is in accordance with the validity criteria of Akbar [5] which states that if the validity result is in the range of 0.801 – 1.00 or 80% - 100% then the validity rate is very high. The validation of this material is good enough, because previously Muslich [15] developed hots assessment instruments only get validity of 80%, Siti S. [16] developed a high level of thinking ability tests package is also still validated 87.5%, validated by media experts.

Based on the analysis of data from media experts, the product of mathematical test instruments is worth using reviewed from the presentation of text or writing and language. In the validation results of media experts processed the percentage achieved by 95.9%. Based on these results, the product of mathematical test instruments of TIMSS type is declared valid and can be used by media experts. The statement is in accordance with the opinion of Soyah [16] the question can be said to be valid if the coefficient of validity must be more than 0.600 ($r_{xy} > 0.600$) or can be expressed in percent up to 60%. So, the question of getting $r_{xy}$ results below 0.600 can be expressed as an invalid question. The results of this media validation have been quite good, seeing previously Nurmadinah [17] develop hots test instruments with a media validity achieve of only 80.66%.

Based on the validation results of teachers and students, it is known that the product of a mathematical test instrument of TIMSS type is valid and worth using. From validation results by teachers obtained a percentage of achievability of 100%. Meanwhile, student validation results through one-to-one tests and small-scale trials achieved an achievable percentage of 94.5%. Based on the validation results, the product of mathematical test instrument of TIMSS type is declared very valid and can be used. The validity rating of this user is already good to see when Hapsari [18] when developing hots assessment instrument skilled practitioners (users) give a value of 90%.

Validity tests are carried out aimed at testing the validity or correctness of the designed test instruments. According to Akbar [5] states that a test instrument can be said to be of very high validity when it is obtained values in the range of 0.801 to 1.00 in its validity test. And it is said to have a very low validity when obtaining values in the range of 0.00 to 0.200. From the validity test
results of mathematical test instruments of type TIMSS shows the result that there are 9 question numbers that get high validity scores and 11 numbers get very high validity scores. The results of this validity test are good enough, looking at previously Anggrianin [19] the results of the validation of hots assessment instrument products stating that all questions are in the value range of 4 ≤ Va ≤ 5 with a valid interpretation.

Reliability tests are carried out aimed at testing the reliability or determination of designed test instruments. According to Matondang [20] states that the criteria used in assessing the reliability of an instrument can use validity criteria. The reliability criterion is that a test instrument can be said to have very high reliability if it is obtained values in the range of 0.801 to 1.00 in its reliability test. And it is said to have very low reliability when obtaining values in the range of 0.00 to 0.200. From the reliability test results of mathematical test instruments of TIMSS type produce a value of 0.895. The value is in the range of 0.801 to 1.00 and it can be stated that mathematical test instruments of TIMSS type have very high reliability, the result is quite good when compared to the previous Muslich [15] only got a degree of reliability of 0.713 with high interpretation.

Calculation of difficulty level is done to measure the level of difficulty in the work of question items by students. according to Arikunto [5] when the question item is worth 1.00 then n = the question is very easy to work with, and when the question item gets a value close to 0.00 then the question is difficult to work with. From the calculation of difficulty levels on mathematical test instruments obtained results that there are 4 questions that have a high/difficulty index, 13 questions that have a moderate difficulty index, and 3 questions that have a low/easy difficulty index. According to Mullis [5] it states that the problems found in TIMSS have levels consistent with the description of cognitive domains in table 2.1 about the percentage of cognitive domains in TIMSS. The result has been quite good, with the previous Sofyah [16] there are 5 questions with difficulty index and 4 questions with medium difficulty index no question with easy difficulty index.

The calculation of test differentiating power aims to test the ability of test instrument in distinguishing highly skilled students and low-skilled students. According to Akbar [5] stated that the question that can be answered correctly by highly skilled and low-skilled students then the question has a low or punishing differentiation power. Therefore, a question can be said to have a high differentiating power / very good if it is in the range of values 0.71 to 1.00, and is said to have a low differentiating power / bad if it is in the range of values 0.00 to 0.20. From the calculation of the differentiating power of mathematical test instruments of TIMSS type shows the result that there are 5 question numbers that get a very low/bad differentiating power index and 4 numbers get a very high/very high differentiating power index. However, the results of the differentiating power are better than Sofyah [16] when developing a high level of thinking ability test package there is 1 problem with good differentiating power, 4 questions with sufficient differentiating power, and 4 questions with bad differentiating power.

The validation of mathematical test instruments of TIMSS type shows the criteria are very valid so that it can be used without revision. It can be interpreted that mathematical test instruments deserve to be used as a tool to train students’ high-level thinking skills, especially in math subjects. The feasibility of the test instrument is in accordance with the requirements of a good test according to Akbar [5] namely validity, reliability, objectivity, predictivity, and economy. The six conditions have been spelled out in the validity instrument and have been analyzed the quality of each question item contained in the test instrument.

TIMSS-type mathematical test instruments can be used by teachers as a reference in developing daily problems in learning, especially in mathematics subjects. Based on previous research that mathematical test instruments of TIMSS type can train high-level thinking skills in the realm of critical and creative thinking. Oktavia [18] has been developing a TIMSS-type test instrument with the results of the study stating that the development of mathematics sola of type TIMSS can improve critical thinking skills in students, although students have difficulty in identifying problems given in the question and time constraints, this is because students are wrong in choosing the problem-solving process.

TIMSS-type mathematical test instruments developed have advantages that are the material contained in it in accordance with KI and KD curriculum 2013 revision 2017, materials tested in accordance with the daily life of students, mathematical test instruments have 3 indicators in each question namely from the 2013 curriculum, indicators from TIMSS and HOTS indicators. The sentences in the question are typed and arranged and adjusted to the student’s age, making it easier for students to work. The look of instruments packed with Wonder share app is not boring, it can increase students’ learning motivation. TIMSS-type mathematical test instruments can stimulate students to perform high order thinking skills (HOTS).

In addition to having advantages, the product of mathematical test instruments of TIMSS type also has weaknesses. There is a weakness that in terms of the assessment process cannot be automatically available after completion of the test instrument because in the mathematical test instrument of TIMSS type there are 5 details question that require manual assessment with the indicator stipulated. In the implementation of the work of the test instrument requires paper and pen to write the answer of the description. Mathematical test instruments
developed cannot only be opened with laptops cannot be opened on mobile phones or gadgets.

REFERENCES

[1] Yuniarwati. Penerapan pembelajaran matematika dengan strategi REACT untuk meningkatkan kemampuan koneksi dan representasi matematik siswa sekolah dasar, 2011. http://jurnal.uptp.ac.id/penelitian-pendidikan/view/639/penerapan-pembelajaran-matematika-dengan-strategi-react-untuk-meningkatkan-kemampuan-koneksi-dan-representasi-matematik-siswa-sekolah-dasar(studi-kuasi-eksperimen-di-kelas-v-sekolah-dasar-kota-cimahi).html.

[2] Distariana Hanifah & Janet Tineke Manoy, Indetifikasi tipe berpikir dengan soal higher order thinking (HOT) ditinjau berdasarkan kemampuan matematika, 2014, http://journalmahasiswa.unesa.ac.id/index.php/mathedunes/article/view/12715.

[3] InaV.S. Mullis, TIMSS assessment framework. united state: TIMSS& PIRLS international study center, Lynch school of education, boston college and international association for the evaluation of educational achievement (IEA), 2015, https://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Bk.pdf.

[4] Ana Ratna Wulan, pengertian dan esensi konsep evaluasi, esesmen, tes dan pengukuran, 2013, https://s3.amazonaws.com/academia.edu.documents/34534033/pengertian_asesmen.pdf?AWSAccessKeyId=AKIAIWQWWYGGZ5UL3A&Expires=1545703861&Signature=eXZ50pRRURf18c8tZChvkwVIMY%3D&response-content-disposition=inline%3B%20filename%3DPENGERTIAN_DAN_ESENSI_KONSEP_EVALUASI_AS.pdf.

[5] Sa’duh Akbar. Instrumen perangkat pembelajaran. Bandung, PT Remaja Rosdakarya, 2017.

[6] Nanik Dwi Rahayu, Analisis kemampuan siswa kelas V dalam menyelesaikan soal matematika model TIMSS di MIN Gedog Kota Blitar. Skripsi tidak diterbitkan. 

[7] Stephen Provasnik, Highlights from TIMSS and TIMSS advanced, 2015, 2016, https://nces.ed.gov/pubs2017/2017002.pdf.

[8] InaV.S. Mullis, TIMSS.Trend In Mathematics and Science, 2017, http://www.iea.nl/timss.

[9] InaV.S. Mullis, International result in mathematics. United State: TIMSS & PIRLS international study center, Lynch School of Education, Boston College and International Association for the Evaluation of educational achievement (IEA), 2016, https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf.

[10] Emi Rofiah, Penyusunan instrumen tes kemampuan berpikir tingkat tinggi fisika pada siswa SMP, 2013, http://journal.fkip.uns.ac.id/index.php/ptfisika/article/view/2797.

[11] Yee Mei Heong, The level of marzano higher order thinking skills among technical education student, 2011, https://pdfs.semanticscholar.org/6e5a/76e3994e6a23df0e3296da8b9c944d119b87.pdf.

[12] Yoki Arayana, Buku Pegangan Pembelajaran Berorientasi Pada Keterampilan Berpikir Tingkat Tinggi. Direktorat Jendral Guru dan Tenaga Kependidikan. Kementerian Pendidikan dan Kebudayaan, 2018.

[13] Sugiyono, Metode penelitian kuantitatif, kualitatif dan R&D. Bandung, Alfabeta, 2015

[14] Febriawanwati Yusup, Uji validitas dan reliabilitas instrumen penelitian kuantitatif, 2018, https://www.researchgate.net/publication/327699726_Uji_V aliditas_dan_Reliabletis_Instrumen_Penelitian_Kuantitatif /fulltext/5b9b09ea6fdec3cb5ed353/327699726_Uji_V aliditas_dan_Reliabletis_Instrumen_Penelitian_Kuantitatif.pdf?origin=publication_detail.

[15] Muslich, Pengembangan instrumen asesmen HOTs pada mata pelajaran matematika SMP kelas VIII semester I, 2010, muslich@un.ac.id.

[16] Siti Sofyiah, Pengembangan paket tes kemampuan berpikir tingkat tinggi matematika berdasarkan revisi taxonomy bloom pada siswa kelas V SD, 2015, https://repository.unej.ac.id/bitstream/handle/123456789/64108/SITI%20SOFYAH.pdf?sequence=1.

[17] Nurmadinah, Pengembangan instrumen tes higher order thinking skills (HOTS) pokok bahasan hitung aljabar serta persamaan dan perbandingan linear satu variabel kelas VII MTs Guppi Samata, 2017, https://respositori.uii-alauddin.ac.id/8559/1/NURMADINAH.Pdf.

[18] Oktavia Hapsari, Pengembangan soal serupa TIMSS untuk mengukur kemampuan berpikir kritis dan pemecahan masalah pada konten aljabar kelas VIII, 2016, eprints.ums.ac.id/48159/2/HALAMAN%20DEPAN.pdf.

[19] Andi Dia Anggiranin, Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Koneksi Matematis. 2018, http://ejurnal.iainmataram.as.id/index.php/.

[20] Zuulkifi Matondang, Validitas dan reliabilitas suatu instrumen penelitian, 2009, http://digilib.unimed.ac.id/id705/.

[21] Imron, A., Wiyono, B. B., Hadi, S., Gunawan, I., Abbas, A., Saputra, B. R., & Perdana, D. B. (2020, November). Teacher Professional Development to Increase Teacher Commitment in the Era of the Asean Economic Community. In 2nd Early Childhood and Primary Childhood Education (ECPE 2020) (pp. 339-343). Atlantis Press.

[22] Sulton, S., Gunawan, I., & Pratiwi, F. D. (2018). Perbedaan Motivasi Belajar Mahasiswa antara Sebelum dan Sesudah Mengikuti Pelatihan Motivational. Ilmu Pendidikan: Jurnal Kajian Teori dan Praktik Kependidikan, 3(1), 115-119.

[23] Bafadal, I., Nurabadi, A., Soepriyanto, Y., & Gunawan, I. (2020, November). Primary School Principal Performance Measurement. In 2nd Early Childhood and Primary Childhood Education (ECPE 2020) (pp. 19-23). Atlantis Press.

[24] Gunawan, I., Triiwianto, T., Kusumaningrump, D. E., Romady, M., Alfarina, M., & Widiana, R. A. (2018). Pemberdayaan Tenaga Administrasi Sekolah Menengah Pertama Kota Batu: Studi Deskriptif. JAMP: Jurnal Administrasi dan Manajemen Pendidikan, 1(4), 467-471.

[25] Wardani, A. D., Gunawan, I., Kusumaningrump, D. E., Benty, D. N. N., Sumarsono, R. B., Nurabadi, A., & Handayani, L. (2020, November). Student Learning Motivation: A Conceptual Paper. In 2nd Early Childhood and Primary Childhood Education (ECPE 2020) (pp. 275-278). Atlantis Press.