

Quality of Tests on Mathematics Questions for Year-End Assessment (PAT) Class VIII at SMPN 2 Bangkinang Kota

Yola Putri Anggraini¹, Yenita Roza², Maimunah³

¹ Mahasiswi Magister FKIP Pendidikan Matematika Universitas Riau

Email: yola.anggraini06@gmail.com

²³ FKIP Pendidikan Matematika Universitas Riau

Email: ¹ yenita.roza@lecturer.unri.ac.id, ² maimunah@lecturer.unri.ac.id

(Received: 4-11-2020; Reviewed: 25-11-2020; Revised: 29-11-2020; Accepted: 2-12-2020; Published: 21-12-2020)



©2020 –Daya matematis: Jurnal inovasi pendidikan matematika. Ini adalah artikel dengan akses

terbuka dibawah licensi CC BY-NC-4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>)

Abstract

This study aims to see the quality of the test on multiple choice questions in mathematics for Year End Assessment (PAT) class VIII SMPN 2 Bangkinang Kota for the 2019/2020 academic year based on the level of question validity, reliability level, difficulty level, differentiation power, and option effectiveness. This research is an evaluation study with a quantitative approach. The research was conducted with a sample of 30 student answer sheets. Techniques for collecting data are documentation techniques. After the data is collected, the data will be analyzed with the help of Excel. Judging from the results of this analysis, it can be seen that the End of Year Assessment (PAT) questions for class VIII SMPN 2 Bangkinang Kota for the 2019/2020 academic year were not carried out by quantitative reviewers. Judging from the level of validity, it consists of 14 valid questions, while the other 16 questions are said to be invalid. The PAT questions have a high level of reliability. As for the difficulty level, there were 7 items (23,33%) difficult, 18 items (60%) moderate, and 5 items (16,66%) easy. The distinguishing power is 8 questions that are bad (26.66%), 7 questions with sufficient criteria (23.33%), 7 questions with good criteria (23.33%), and 8 questions with very good criteria (26.66%) For the effectiveness of the option, 4 items (13,33%) are very good, 15 items (50%) are good, 9 items (30%) are bad, and 2 items (6,66%) are very bad. questions are not included in good quality for Year-End Assessment (PAT) class VIII for the 2019/2020 school year.

Keywords: validity level; reliability; level of difficulty; distinguishing power; effectiveness of options (copter)

INTRODUCTION

Permendikbud Number 104 of 2014 states that the assessment of learning outcomes (evaluation) is a process carried out in gathering information or evidence regarding the learning achievements of students in the competence of spiritual attitudes, social attitudes, knowledge competencies, and skills competencies that are carried out systematically. So that learning evaluation must be mastered by every teacher (Arifin in Lestari and Siregar, 2019). The success or failure of the evaluation activities carried out will be influenced by the success of the evaluators who are educators in planning, compiling and producing good test instruments based on things that are considered in the development of learning outcomes tests (Magdalena, Soleha, Rukoyah, Hidayah, and Nufus; 2020)

As evaluators, teachers not only make evaluation tools (tests) but also can understand the making of evaluation tools because of the importance of evaluation in the learning system. Measurements and assessments in the field of education can be obtained using tests (Kadir, 2015). A test will have a very important meaning if it can test important objectives and represent all the tested materials in a representative manner (Lestari, Wardhana, and Rahmawati, 2019). Evaluation tools must also be of good quality, so that there are no errors in measuring students' abilities. Nana Sudjana (2014) also said that a

good quality test has met the validity and reliability requirements. In line with this, high quality test items can also be seen from the level of difficulty, distinguishing power and effectiveness function (distractor function). So, the test questions that are made must be relevant to the learning objectives to be achieved. So, the learning outcome test can be said to be of good quality if it is through item analysis activities (Ariandi in Lestari and Siregar, 2019). So, to find out which questions are good, bad, or bad, it can be done by analyzing the items.

Based on information obtained from an interview with a grade VIII mathematics teacher at SMPN 2 Bangkinang Kota, that the teacher still has not conducted an item analysis of the PAT Even questions for class VIII SMPN 2 Bangkinang Kota for the 2019/2020 academic year. So that the quality of the questions is not yet known whether they are of good or bad quality. This is due to the lack of attention of the teacher in assessing the test questions to be used. So, the item analysis needs to be done to see the quality of the PAT test questions whether the questions are of good quality or not. In line with the research conducted by Raafi and Ndia, (2015), namely "The Quality of Teacher-Made Tests on Multiple Choice Questions in Mathematics Class VIII SMP Negeri 4 Kendari Even Semester Academic Year 2013/2014" states that the results of the analysis of test questions The difficulty consists of 5 questions (easy), 14 questions (medium), and 0 questions (difficult). In addition, seen from the distinguishing power it also shows that the items are classified as good, namely consisting of 10 questions (enough) and 9 questions (good). Furthermore, from the perspective of the effectiveness option, 19 questions consisted of 19 questions that had good distracting options. Then, the items also have a high level of reliability. So it can be concluded that the math test questions for even semester VIII grade SMP Negeri 4 Kendari in the academic year 2013/2014 have good quality.

A teacher must be able to design and develop questions for PAT, so that in addition to developing his own abilities in making questions, the teacher must have special abilities to produce quality test questions. Good questions will have criteria for the level of validity, level of difficulty, differentiation, and effectiveness of options (distractors) that are good (Nur, Mania, and Halik, 2019). According to Salmina and Adyansyah (2017) also stated that validity was carried out to see to what extent the test could measure what it wanted to measure. The quality of the test in relation to measuring what should be measured can be determined by doing the validity of the test. In line with this, Arikunto (in Salmina and Adyansyah, 2019) also states that the level or degree of consistency of an instrument is called reliability. Errors in the measurement process can be seen from the level of reliability, which can lead to changes in the composition of the group. Alpha Cronbach formula can be used to calculate the level of reliability. Furthermore, the level of difficulty can be seen from the calculation of the proportion between the number of students who answered the questions correctly and the total number of test takers (Rahayu and Djazari, 2016). The number that expresses the difficulty of a problem is called the difficulty index. According to Zainal (in Rahayu and Djazari, 2016) states that the questions that can see which students master the material and students who do not master the material can be seen from their distinguishing power. In line with this, by looking at the size of a discrimination index, the items can determine their discriminating power. Then, to measure the distractor index of an item, it can be measured by its effectiveness (Rahayu and Djazari, 2016).

Based on this explanation, the things that will be examined in this study are determining the quality of the math problem test for PAT class VIII SMPN 2 Bangkinang Kota for the 2019/2020 school year. So that from this research it is hoped that it can provide benefits to readers, especially teachers so that they can find out the quality of the questions made and have a reference to follow up on the questions made.

METHOD

This research was conducted in Kampar Regency from 02 to 30 October 2020, with the subjects of 30 junior high school mathematics teachers randomly selected from 21 sub-districts in Kampar district, both those who have or not. This type of research is categorized as a qualitative descriptive study. Qualitative research is a research approach that reveals certain social situations by describing reality

correctly, formed by words based on relevant data collection and analysis techniques obtained from natural situations without any treatment by the researcher on the subject. Descriptive research is research that aims to determine the circumstances and conditions in which the results are explained in the form of a research report.

This research was conducted in October 2020 at SMPN 2 Bangkinang Kota. The analysis conducted in this research is descriptive analysis using a quantitative approach. Researchers can analyze data by describing or describing the data that has been collected. The type of data used in this research is quantitative data. The data collected were PAT multiple choice question sheets, syllabus, question grids, student answer sheets, and answer keys to PAT questions in mathematics for the 2019/2020 school year. The research variable is the quality of the items seen from the level of validity, reliability, difficulty, differentiation, and effectiveness of options (distracting). The population in this study were the question sheets and answer sheets for all eighth grade students. Then, the sampling technique was carried out by purposive sampling, namely 30 student answers. The technique used to collect research data is documentation technique. Previously, the researchers prepared the material that would be used to analyze the questions. Researchers will take class VIII PAT questions and student answers according to the agreement.

Then, the criteria for the level of difficulty of a test item are carried out by looking at the comparison between the average score obtained and the maximum score of an item. Meanwhile, the size of an item discriminant index is known to see the distinguishing power of a question. Then, the effectiveness of an item's options can be seen based on a scale ranging from very good to not good.

RESULTS AND DISCUSSION

Results

The results of the analysis of multiple choice PAT math problems class VIII for the 2019/2020 school year of SMPN 2 Bangkinang Kota with the help of Exel based on the level of validity stated that the questions were said to be valid if r counted from the r table (0.361), so that of the 30 items there were 14 questions valid and 16 questions are invalid.

The results of the analysis of the level of reliability of the Year End Assessment (PAT) questions of class VIII in the mathematics subject for the 2019/2020 school year of SMPN 2 Bangkinang Kota with the help of Exel obtained that the reliability test results of a test, then the test has high test reliability. If the difficulty index of the questions exceeds 0.70 then the questions are categorized as easy, while the difficulty index of the questions is between 0.31-0.70 then the questions are in the medium category, while the questions with the difficult category are those with the difficulty index less than 0.30. So, the difficulty level of multiple choice PAT math problems class VIII for the 2019/2020 academic year of SMPN 2 Bangkinang Kota with the help of Exel obtained that of the 30 questions consisting of 7 (difficult) questions, 18 (moderate) questions, and 5 (easy) questions.

The test with a good category are questions with a distinguishing power index of more than 0.40. Meanwhile, the index of distinguishing power below 0.41 was categorized as less good. So, the results of the analysis of the distinguishing power of PAT class VIII multiple choice math problems for the 2019/2020 school year of SMPN 2 Bangkinang Kota using Exel's assistance, it was found that out of 30 questions there were 8 questions (bad), 7 questions (enough), 7 questions (good), and 8 questions (very good).

As for the effectiveness of the options (distractor) multiple choice math problems in PAT class VIII for the 2019/2020 school year at SMPN 2 Bangkinang Kota with the help of Exel, it shows that the questions with the good option effectiveness category are questions that have at most one distractor that does not function properly, while the questions those with more than one non-functioning delinquent were included in the poor option effectiveness category. So, of the 30 questions, there were 4 questions in the

very good category, 15 questions in the good category, 9 questions in the bad category and 2 items in the very bad category.

Discussion

Multiple choice math problems for PAT class VIII 2019/2020 SMPN 2 Bangkinang Kota is a bad question, because when viewed from the level of validity of the questions there are 14 questions that are categorized as valid, namely questions number 7 and 18 have very high criteria with a percentage of 6.66 %, question number 1,15,16,17,21,25,28,30 has high criteria with a percentage of 26.66%, and question number 4,23,24,29 has medium criteria with a percentage of 13.33%. While the other 16 questions are categorized as invalid, namely question number 2,3,26,14 has low criteria with a percentage of 13.33%, question number 5,6,8,9,11,13,27 has very low criteria with a percentage of 23, 33%, while question number 10,12,19,20,22 has invalid criteria with a percentage of 16.66%. One of the questions presented is item number 1:

"It is known that the length of the hypotenuse of a right triangle is 25 cm. One of the perpendicular sides is 24 cm long. The length of one side of the other is"

The competency standard of the question is to explain and prove the Pythagorean theorem and the pythagorean triple. The indicator is that students can determine the length of the other side if it is known the length of the hypotenuse and the length of the perpendicular side. This question has a very high validity category, because it is in accordance with the learning material contained in the 2013 curriculum.

Based on the level of reliability, it shows that the math problems (PAT) class VIII of SMPN 2 Bangkinang Kota has a reliability coefficient (r_{11}) of 0.7292 with a high category. According to Raafi and Ndia (2015) states that a question is said to be reliable if an item can be tested repeatedly with the same object, but gets consistent and relatively the same results.

One of the questions presented is question number 11:

"It is known that the length of the tangent to the outer association of circles C and D is 12 cm. The radii of the circles C and D are 7.5 cm and 4 cm, respectively. The distance between the two center circles is... cm".

The competency standard of the question is about explaining the tangent of the outer partnership and the partnership in two circles and how to paint it. The indicator is to determine the distance between the two center circles. This question has a level of difficulty with a bad category, because it can only be answered by one student, so the question is considered a difficult question.

Based on the results of the analysis, it can be seen that there are 60% of the questions with a good level of difficulty, and 40% with a bad level of difficulty, namely 23.33% including difficult questions and 16.66% including easy questions. The difficulty level of the questions (medium) is a good quality question which can be seen from the difficulty index between 0.31-0.70 meaning it is neither easy nor too difficult. Because if the questions are too easy it cannot make students think actively in solving a given problem, but questions that are too difficult or difficult will also make students despair, so that there is no desire and enthusiasm to solve the problems given. The good items consist of 18 questions, namely question number 1 (index 0.67), 4 (index 0.57), 5 (index 0.37), 6 (index 0.4), 7 (index 0.67), 8 (index 0.57), 13 (index 0.7), 14 (index 0.7), 15 (index 0.67), 17 (index 0.67), 18 (index 0.7), 21 (index 0.4), 22 (index

0.3), 23 (index 0.43), 24 (index 0.7), 25 (index 0.57), 28 (index 0.7), 29 (index 0.67). Then these questions can be entered into the question bank. In line with this, questions that have a difficulty index between 0.71-1.00 or 0.00-0.30 (easy or difficult) are categorized as poor quality questions (Rahayu and Djazari, 2016). That is because easy questions come from easy material. Then, questions that are difficult or difficult come from material with a high level of complexity, so that accuracy is needed in this case. Thus, the questions that are not good consist of 7 (difficult) questions, namely questions with numbers 3 (index 0.23), 10 (index 0.2), 11 (index 0.03), 19 (index 0.1), 20 (index 0.03), 26 (index 0.23), 27 (index 0.23) and 5 items (easy), namely 2 (index 0.8), 9 (index 0.77), 12 (index 0.83), 16 (index 0.77), 30 (index 0.77). Questions that are in the bad category must first revise them before being submitted to the question bank.

Based on the results of the analysis, seen from the distinguishing power there were 8 questions in the bad category (26.66%), 7 questions in the good category (23.33%), 7 questions in the good category (23.33%), and 8 questions in the good category. very good (26.66%). One of the questions presented is question number 19 as follows:

"A block is 12 cm long, 6 cm wide, and 4 cm high, then the surface area of the block is...."

Problem number 19 with competency standards on distinguishing and determining the surface area and volume of flat side spaces (cubes, blocks, prisms, and pyramids). While the indicator is that students can determine the surface area of the block. This question is included in the bad question category, because they cannot distinguish which students can answer the questions correctly because they have mastered the material with students who do not master the material.

A good question can be seen from the level of distinguishing power either or very well with an index of 0.41-0.70 or 0.71-1.00 (Rahayu and Djazari, 2016). The good questions consist of 15 questions, namely questions number 1 (index 0.6), 7 (index 0.53), 8 (index 0.33), 13 (index 0.33), 14 (index 0.4), 15 (index 0.4), 16 (index 0.33), 18 (index 0.47), 21 (index 0.53), 23 (index 0.33), 24 (index 0.33), 25 (0.33), 28 (index 0.53), 29 (index 0.33), 30 (index 0.4). Questions that can be entered in the question bank if they are in a good category. Meanwhile, questions that are good enough can also see students who do not master the material who happen to be able to answer the questions correctly. So, the questions that are not good consist of 7 questions, namely questions with numbers 2 (index 0.27), 4 (index 0.27), 5 (index 0.27), 6 (index 0.2), 9 (index 0), 17 (index 0.27), 26 (index 0.2). So, these questions must be revised before being included in the question bank. While the bad items consisted of 8 items, namely questions with the number 3 (index 0), 10 (index -0.27), 11 (index 0.07), 12 (index -0.07), 19 (index -0.2), 20 (index -0.07), 22 (index -0.2), 27 (index 0). So, these questions must be deleted and replaced with new questions in order to be included in the question bank.

Based on the results of the analysis, it shows that of the 30 questions there are 4 items (13.33%) with very good option effectiveness, 15 items (50%) with good option effectiveness, 9 items (30%) with poor option effectiveness, and 2 items (6.66%) with very poor effectiveness of the options. Questions that have at most 1 option that do not work are in the good or very good category (Rahayu and Djazari, 2016). In line with this, the questions were said to be good if at least 5% of the test takers did not choose one cheater. There are 19 good questions (13.33% very good and 50% good), namely questions with numbers 2,3,5,6,7,8,9,11,14,15,16,18,20,21, 24,25,26,28,29. So, questions can be entered into the question bank if they have a good category. Whereas in the question where there are 2 options that do not work, then the question is categorized as quite good. This can also be seen from the number of test takers at least 5% who did not choose the distractor option, so it is a fairly good question. The questions in the fairly good category consisted of 9 questions, namely questions with the numbers

1,4,10,13,17,19,22,23,30. So, these questions must be revised before being put into the question bank. Furthermore, questions with three options that do not work are considered very bad criteria. It can also be seen that at least 5% of the test takers did not choose the three distracting options. So, there are 2 items that are categorized as very bad, namely with question numbers 12 and 27. Therefore, questions in the bad or very bad category should be deleted and replaced with new questions.

CONCLUSIONS AND SUGGESTIONS

Based on the results of the analysis of the items on the Final Semester Assessment (PAT) questions in mathematics class VIII SMPN 2 Bangkinang Kota for the 2019/2020 academic year seen from the level of question validity, reliability, difficulty level, distinguishing power, and effectiveness of options, it shows that of the 30 PAT questions there are only 14 questions that are said to be valid with a high level of reliability. Judging from the level of difficulty of the items, including the questions that are not good, this is because there are still questions that are difficult or easy. If it is seen from the distinguishing power of PAT questions, it is also a bad question, because there are still 15 questions that are categorized as bad, namely 7 items with sufficient distinguishing power and 8 items with bad distinguishing power. Then, seen from the effectiveness of the options, it is also included in the bad category, this is because there are still 11 items with the effectiveness of the option (distractor) which are in the bad category, namely 9 items with bad option effectiveness and 2 items with very bad option effectiveness bad.

So, it can be concluded that the multiple choice questions in mathematics subject for Year End Assessment (PAT) class VIII SMPN 2 Bangkinang Kota did not analyze the questions well. Based on the results of the quantitative test, it is known that most of the items cannot be used for class VIII PAT for the 2019/2020 school year. This is because the requirements for validity, reliability, difficulty level, distinguishing power, and the effectiveness of deceiving a question are not fulfilled. So, only a few questions can be used. Therefore, in this study it is hoped that the teachers can analyze the items properly. If the teacher has analyzed the questions well, then the quality of the items used to measure students' abilities is expected to be more accurate.

REFERENCE

- Anas Sudijono. 2015. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada
- Halik, Andi Suharman,. Mania, Sitti., dan Nur, Fitriani. 2019. Analisis Butir Soal Ujian Akhir Sekolah (UAS) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 SMP Negeri 36 Makassar. *Journal of Islamic Education*. Vol 1 No 1. Lembaga Penelitian Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Alauddin Makassar
- Kadir, Abdul. 2015. Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al-Ta'dib*. Vol 8 No 2. Lembaga Penelitian Fakultas Tarbiyah dan Ilmu Keguruan IAIN Kendari
- Lestari, Ayu Dwi, dan Siregar, Hita Paulina. 2019. Analisis Butir Soal Matematika Buatan Guru SMP Negeri 1 Mempura. *Mathematics Education Journal*. Vol 2 No 1. Lembaga Penelitian Program Studi Magister Pendidikan Matematika Universitas Riau Pekanbaru
- Lestari, Shintia, Wardana, M. Yusuf Seti, dan Rahmawati, Intan. 2019. Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran Matematika Kelas III SD Kecamatan Petarukan Tahun 2018/2019. *Elementary School Journal*. Vol 9 No 2. Lembaga Penelitian Universitas PGRI Semarang
- Magdalena, Ina., dkk. 2020. Evaluasi Kemampuan Guru Dalam Pengembangan Tes Pilihan Ganda di SDN Rawa Buaya 05 Petang Jakarta Barat. *Jurnal Halaqah*. Vol 2 No 4. Oktober 2020. Lembaga Penelitian Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Muhammadiyah Tangerang

- Nana Sudjana. 2014. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya
Permendikbud Nomor 104 Tahun 2014 *Penilaian Hasil Belajar*.
- Raafi, Wa Ode Nirwana dan Ndia, La. 2015. Kualitas Tes Buatan Guru Pada Soal Pilihan Ganda Mata Pelajaran Matematika Kelas VIII SMP Negeri 4 Kendari Semester Genap Tahun Pelajaran 2013/2014. *Jurnal Penelitian Pendidikan Matematika*. Vol 3 No 1. Lembaga Penelitian Program Studi PMIPA SKIP UHO
- Rahayu, Rahmatika. 2016. Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*. Vol 14 No 1. Lembaga Penelitian Prodi Akuntansi Universitas Negeri Yogyakarta
- Salmina, Mik., dan Adyansyah, Fadlillah. 2017. Analisis Kualitas Soal Ujian Matematika Semester Genap Kelas XI sma Inshafuddin Kota Banda Aceh. *Jurnal Mik Salmina*. Vol 4 No 1. Lembaga Penelitian STKIP Bina Bangsa Getsempena
- Sary, Yessy Nur Endah. 2018. *Evaluasi Pendidikan*. Yogyakarta: CV Budi Utama
- Solichin, Mujianto. 2017. Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes Interpretasi Hasil Tes dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Jurnal Manajemen dan Pendidikan Islam*. Vol 2 NO 2. Lembaga Penelitian Program Pascasarjana UNIPDU Jombang
- Sriyanti, Ika. 2019. *Evaluasi Pembelajaran Matematika*. Jawa Timur: Uwais Inspirasi Indonesia
- Rahaju, E.B., Dini K.F., Pradnyo W., Ismail. 2020. Kemampuan Guru-guru Matematika SMP Kabupaten Ponorogo dalam mengembangkan soal Berpikir Tingkat Tinggi. *JPMR*, Vol.05
- Risnawati, 2008. *Peningkatan Pengembangan Matematika*, Suska Press, Pekanbaru
- Riadi, Arifin. 2016. Problem-Based Learning Meningkatkan Higher-Order Thinking Skills Siswa Kelas VIII SMPN 1 Daha Utara Dan SMPN 2 Daha Utara. *Math Didactic Jurnal Pendidikan Matematika*
- Sani, Ridwan Abdullah. 2019. *Pembelajaran Berbasis HOTS (Higher Order Thinking Skills)*. Tangerang : Tira Smart
- Setiawati, Wiwik dkk. 2019. *Buku penilaian Berorientasi Higher Order Thinking Skills* Dirjen GTK Kemendikbud
- Tajudin, N. M., & Chinnappan, M. 2017. Role of higher order thinking skillss in enhancing mathematical problem solving, *Man in India*, 97(17), 209-214