How is the quality of 11th grade midterm exam in Indonesia? Is it important?

Ibrohim Aji Kusuma[1], Heri Retnawati[2]
{ibrohimaji.2018@student.uny.ac.id[1], heri_retnawati@uny.ac.id[2]}

Graduate School Program, Universitas Negeri Yogyakarta, Jl. Colombo No. 1 Yogyakarta[1], Faculty of Mathematics and Science, Universitas Negeri Yogyakarta, Jl. Colombo No. 1 Yogyakarta [2]

Abstract. This study aims to analyze the characteristics of the item test of mathematics midterm exam of 11th grade secondary students. This study was quantitative research using a descriptive explorative approach. The midterm script and its answer key ware validated by three expert and students answer sheet was analyzed using classical test theory method using QUEST. The midtest test script and its answer key were valid with score 0.93 using Aiken index and the internal consistency score was 0.74. It means that the midtest script was feasible in learning but cannot be used to reflect students' knowledge and skills. Fit Pt-biserial score was 52.5% and Low Pt-biserial score was 47.5%. The difficult items consist of the arithmetic series related to real problems, infinite geometric series, algebraic limits with substitution of fraction forms containing roots and factorization of fraction forms containing exponential.

Keywords: Item test characteristics, midterm exam, classical test theory

1 Introductory

Assessment is one of the most prominent aspects of education. Assessment is a process of collecting and organizing information to measure students' learning outcomes [1, 2]. This assessment is used to measure the extent to which competencies developed in the classroom have been learned by students [3]. Further, the results will not only used as evaluation in order to improve the learning quality in the classroom but also indeed affecting educational policies in the future [4].

In Indonesia, curriculum 2013 states that there is three types of assessment which include assessment of teachers, assessment of schools or educational institutions and assessment of government. Assessment of teacher consists of daily tests and midterm exams. Assessment of schools consists of final semester tests, final exams, school exams, and national standard school exams. Moreover, the assessment of the government consists of a national standard exam.

The daily test aims to measure students' outcomes of particular topics or competencies intended in learning. In general, it measures only one-topic or basic competence. Besides, the midterm exam aims to describe students' achievement of several topics or competencies in a half-one semester of learning. This assessment of teacher is used to measure and describe students' outcomes, create remedial and enrichment programs based on their mastery level of competences, improve the learning process, and report the students' learning outcomes. Moreover, the final semester test aims to measure students' outcomes at the end of the semester.

MSCEIS 2019, October 12, Bandung, Indonesia

DOI 10.4108/eai.12-10-2019.2296528

The average of a daily test, midterm exam and final semester test will be used to determine the students' final score. This final score will be reported in students' report as a score and category that represent students' knowledge and skills related to the topic that has been studied in one-semester of the academic year.

The final score is essential for students. This score will be used as a primary consideration in the process of decision making in their future [4]. Thus, the score must be able to truly represent the students' knowledge and skills. Therefore, students' learning tests must be able to measure their knowledge and skills precisely. Good students' learning test consists of proper quality items. Otherwise, unfeasible items result from error measurement that causes detrimental to students. Hence, good quality items are needed to measure students' outcomes truthfully [5]

Analysis of items characteristic is a method to analyze and develop good quality items test [6-9]. This analysis is used to describe the quality test items empirically [8]. There is two types of analyses such that classical test and item response theory. Classical test theory assumed that students' score is obtained from the total of students real score and error measurement score. Besides, it is required some assumptions that are: 1) no interaction and correlation between error measurement scores and the real score, 2) error measurement scores in any test are dissimilar and 3) average of error measurement score is equal to zero [8].

Based on the assumptions, therefore students learning test is validated by experts and also calculate the internal consistency of the test. Thus, this theory has weakness that is: 1) group dependant, the analysis depends on the samples, 2) the scores cannot be generalized, 3) internal consistency is difficult to fulfill, 4) the results of the analysis cannot provide a basis for determining the response of students to certain items, 5) error measurements are assumed equal each students, and 6) test item testing procedures are less practical [8].

On the contrary, item response theory estimates the probability of items using the mathematical model. The probability means the change of students' correct answers which depends on the students' competencies and items characteristics. There is three assumptions to fulfill that are unidimensional, local independence and parameter variations. Unidimensi means that each item only measures one indicator of competences. However, this assumption is difficult to fulfill absolutely because of other factors that can influence students in answering items test. Local independence means students' responses in completing any items will be independent of each other statistically when the influencing factors are constant. In other words, the answers of students to an item will not affect other item's answers. Variable parameter means that the characteristics of the items do not depend on the distribution of the parameters of students 'competencies and vise versa [10].

Several studies conducted to determine the characteristics of the daily test have been carried out [11], as well as the analysis of items characteristics of final semester test, school exam and national standard [12, 13]. Yet, research that aims to analyze the characteristics of the item test of the midterm exam is not carried out. In fact, the midterm exam also contributes to determine students' final scores. Therefore, this study aims to analyze the characteristics of items tests of the midterm exam that have been used recently.

2 Methods

This research was a quantitative research approach with descriptive-exploratory methods. It aimed to describe the characteristics of mathematics midterm exam test items of senior high

school in Sleman, Yogyakarta. This item characteristic analysis is a way to establish a good quality items test. Further, it describes the quality of test items empirically Fourty items of multiple choice that consist of two basic competencies and a hundred-forty five students' answer sheets of 11th grade was collected.

There was three procedure to analyze the data that are documentation, validation by experts, classical test theory analysis. Midterm exam script and students answer sheet are collected using documentation. Moreover, those are validated by three experts. The instruments are validated using Aiken Indeks. If the score is more than 0.4 or 0.8, it is categorized valid or very valid [8]. Reliability or internal consistency, level of difficulty and pt-biserial are analyzed using QUEST. Students' learning tests can be used as consideration for students in the process of decision making should have minimum reliability coefficient score 0.85 [14]. Items are categorized as easy items with a difficulty level of less than 0.3. It is categorized as medium items with a score between 0.3 and 0.7 and as difficult items with a score more than 0.7 [15]. Good quality item has more than or equal to 0.3 level of difficulty score [8].

3 Results

Midterm exam script and a hundred-fourty five students' answer sheets of 11th grade were collected. The script consists of forty items of multiple-choice with a correct answer and four distractors. Students' answers were tabulated using the 0-1 scale where 0 for an incorrect answer and 1 for the correct answer.

Moreover, the midterm exam script was validated by three experts. The validity score of the midterm exam script was 0.93 that categorized as very valid. It means that the midterm exam was feasible to measure student's learning outcomes. After that, we examined the reliability of the midterm exam script using QUEST. Reliability used in this research was the coefficient of internal consistency. Internal consistency was defined as the constancy of items that are mutually independent and interchangeable [16]. The coefficient of internal consistency was 0.74. Thus, it cannot be used as a primary consideration in the process of decision making for student's future

The next stage was the analysis of the item characteristics of the midterm exam script and its answers. This stage was aimed to analyze the level of difficulty, pt-biserial and its interpretation. The result of the analysis is in Table 1.

It shows that two items were categorized as easy items, thirty-four items were categorized as medium items, and four items were categorized as difficult items such that items 10. 19. 26 and 32. Moreover, items with a good pt-biserial score were 52.5% and the low pt-biserial score was 47.5%. The difficult items and low pt-biserial were discussed as follows:

Table 1. Result of items analysis.

No Item	Level of Difficulty	Category	Pt-Biserial	Category	No Item	Level of Difficulty	Category	Pt-Biserial	Category
1	0.54	Medium	0.25	Less	**21**	0.37	Medium	0.02	Less
2	0.71	Easy	0.50	Good	**22**	0.60	Medium	0.26	Less
3	0.66	Medium	0.30	Less	**23**	0.60	Medium	0.39	Good
4	0.62	Medium	0.25	Less	**24**	0.57	Medium	0.33	Good
5	0.44	Medium	0.28	Less	**25**	0.44	Medium	0.36	Good
6	0.77	Easy	0.44	Good	**26**	0.21	Difficult	0.13	Less

No Item	Level of Difficulty	Category	Pt-Biserial	Category	No Item	Level of Difficulty	Category	Pt-Biserial	Category
7	0.31	Medium	0.32	Good	**27**	0.31	Medium	0.26	Less
8	0.55	Medium	0.43	Good	**28**	0.35	Medium	0.48	Good
9	0.69	Medium	0.34	Good	**29**	0.49	Medium	0.16	Less
10	0.19	Difficult	0.28	Less	**30**	0.42	Medium	0.28	Less
11	0.49	Medium	0.27	Less	**31**	0.53	Medium	0.47	Good
12	0.56	Medium	0.34	Good	**32**	0.30	Difficult	0.08	Less
13	0.34	Medium	0.34	Good	**33**	0.39	Medium	0.39	Good
14	0.68	Medium	0.42	Good	**34**	0.39	Medium	0.29	Less
15	0.51	Medium	0.27	Less	**35**	0.46	Medium	0.35	Good
16	0.69	Medium	0.37	Good	**36**	0.35	Medium	0.33	Good
17	0.65	Medium	0.34	Good	**37**	0.65	Medium	0.27	Less
18	0.30	Medium	0.25	Less	**38**	0.62	Medium	0.36	Good
19	0.22	Difficult	0.03	Less	**39**	0.50	Medium	0.28	Less
20	0.43	Medium	0.26	Less	**40**	0.45	Medium	0.18	Less

Item 10: Father distributes IDR 100.000.00 to 4 of his children. The younger child earns less money. If the difference received by each of the two children who are close to age is IDR 5.000.00 and the eldest receives the greater money, then the money received by the youngest is

A. Rp15.000.00 B. Rp17.500.00 C. Rp19.750.00 D. Rp20.000.00 E. Rp22.500.00

The percentage of students' correct answer for item 10 was 19% and the pt-biserial score was 0.28. Students need to create a mathematical model of the real-world problems in solving this item. Indeed, they should be able to master the concept of arithmetic series that is $s_n = \frac{n}{2}(2a + (n-1)b)$. Thus, we get $100000 = \frac{4}{2}(2a + 3.5000) \rightarrow 4a = 70000 \rightarrow a = 17500$ rupiah. The correct answer was B. Incorrect answer may occur caused by students' misunderstanding of word problems, unable to create a model of mathematics and using inappropriate concepts of arithmetic series.

Item 19: The sum of infinite geometric series of $1 + \frac{2}{3} + \frac{4}{9} + \frac{8}{27} + \ldots$ is ….

A. 2 B. 3 C. 4 D. 6 E. 8

The percentage of students' correct answer to item 19 was 22% and the pt-biserial score was 0.28. Students should be able to master the concept of geometric series that is $S_\infty = \frac{a}{1-r}$. Therefore, we get $S_\infty = \frac{1}{1-\frac{2}{3}} = 3$. The correct answer was B. Incorrect answer may occur caused by using the inappropriate concept of geometric series or the sum of geometric series.

Item 26: Determine $\lim_{x \to 2} \frac{4-x^2}{3+\sqrt{x^2+5}} = \ldots$

A. – 1 B. 0 C. 2 D.6 E. 8

The percentage of students' correct answer to item 26 was 21% and the pt-biserial score was 0.13. Students should be able to master the concept of limit. They need to solve the problem using substitution rule in the form of the fraction that consists of the root. We substituted $x = 2$ to the limit function so that we get the answer is 0. The correct answer was B. Incorrect answer may occur caused by using the inappropriate concept of limit using substitution infraction with root and mathematical error calculation.

Item 32: Calculate $\lim_{x \to 3} \frac{(x-2)^2-1}{x-3} = \ldots$

A. 0
B. 1
C. 2
D. 4
E. 6

The percentage of students' correct answer for item 32 was 30% and the pt-biserial score was 0.08. Students should be able to master the concept of limit. They need to solve the problem using factorization between numerator and denominator. Before that, students firstly determined the exponential of the numerator. Thus, we get $\lim_{x\to 3} \frac{x^2-4x+4-1}{x-3} = \frac{(x-1)(x-3)}{x-3} = x-1$. Substituted $x = 3$, hence $\lim_{x\to 3} \frac{x^2-4x+4-1}{x-3} = 2$. The correct answer was C. Incorrect answer may occur caused not only by using the inappropriate concept of the limit with factorization and exponential but also mathematical error calculation.

Item 21: Shown that f(x) = $\begin{cases} x^2+1. for\ x>2 \\ 2x-1.\ for\ x\leq 2 \end{cases}$. Calculate $\lim_{x\to 2}\ f(x)\ =..$

A. 10 B. 8 C. 5 D. 3 E. 2

Students should be able to master the concept of limit using substitution. This item was categorized as easy that may involve a low pt-biserial score. Therefore, this item cannot distinguish the student's competencies. Unfortunately, only 37% of students answer the item correctly.

4 Discussion

The validity score of the midterm exam script was 0.93 that categorized as very valid. While the reliability score was 0.74. It can be concluded that the midterm exam script was valid but not reliable. Moss argues that it may happen if we define reliability as the internal consistency of the items that are mutually independent and interchangeable [16].

It implies that the midterm exam script was feasible to measure student's learning outcome but it cannot be used as a primary consideration to determine students' competencies and skills. It involved since reliability is related to measurement errors. A smaller reliability score increases error measurement and vice-versa [8]. Thus, the average of several unreliable daily tests and school tests affected the final score which cannot represent students' knowledge and skills precisely for one-semester. Moreover, it cannot be used as a primary consideration in the process of decision making for student's future [4].

The reliability score can be increased by adding the number of items tested. This can result in excessive costs being needed. We can also separate the items according to the competencies independently and testing them at only one time. In addition, improvements can also be made by creating more specific items and detail scoring guidelines [16].

Teachers need to know the competencies, skills, problems and learning difficulties of students to develop effective and meaningful learning designs [17]. It can be seen that students have difficulty in item 10. 19. 26 and 32. The lack of students' concepts understanding was the most dominant cause in solving these problems. Besides, some concepts that needed to be improved include the sum of arithmetic series, the sum of infinite geometric series, and the limit. The concept of algebraic limits can be detailed into inappropriate concept application infractions with root and exponential and using factorization and substitution procedure.

This was in accordance with Gooding [18], Ismail, Shahrill & Mundia [19], Maarif [20]. Alifiani [21] and Retnawati [13] about the lack of students' conceptual understanding. Specifically, the concept of limit [22]. The lack of conceptual understanding can be caused by misconception [22] and motivation [13].

Concept understanding can be increased with contextual teaching learning [13]. Students were not only solving abstract (symbolic) mathematical problems related to the targeted concept but also tied to realistic contexts [23. 24]. Moreover, concept understanding can be enhanced by facilitating students with contextual problems exercises [25]. Hence, it can be concluded that in order to improve concept understanding, students need to associate targeted concepts with realistic contexts. Although, sometimes the implementation of contextual teaching-learning can also attain obstacles [26. 27].

The role of the teacher, in this case, was to facilitate students by developing an instructional design that enhances students' understanding of concepts such as Alifiani [21] by applying the NHT-TGT learning model. In addition. Nugraheni & Sugiman [28] was applying the realistic mathematics education approach and Lasmiyati & Harta [29] that developed learning modules.

The students' competencies to solve word problems or modeling task was also needed to be improved. This was in line with Wijaya, Heuvel, Doorman, & Robitzsch [30] which stated that 38% of students made a failure in executing the mathematical procedure and 3% of students were unable to present the mathematical results. In addition, modeling tasks were difficult to solve caused to the cognitive burden needed which derives from relating mathematical concepts and real-world problems continuously [31]. Novferma [32] found that students' difficulty caused by remembering facts and concepts, understanding facts, applying concepts and procedures, analyzing procedures, evaluating facts, concepts, and procedures, and metacognitive. It also stated that the students' difficulties were due to insufficient time, quitter, inaccuracy, forgetting, anxious, and hastily.

Modeling competencies can be developed by using some activities that were 1) activities to understand real problems and to create a model based on reality, 2) activities to create a mathematical model from real models, 3) activities to solve problems in mathematical models, 4) activities to interpret mathematical results in real situations, and 5) activities to validate solutions to real contexts [33]. Teachers can give guided to students in solving the modeling task and encouraging metacognitive activities [31]. Some studies have conducted to develop modeling capabilities such that using the scaffolding with a solution plan strategy [34] and the implementation of realistic mathematics education approach [35]

Mathematical error calculation also needed to be considered in learning. Wijaya [30] categorized the calculation of the mathematical errors in the students' mistakes in applying mathematical procedures. It was also in line with the work of Novferma [32]. However, mathematical error calculation should be minimized in problem-solving. One of the solutions was students' activities to look back as well as problem-solving procedures in Polya [36].

The 21 of 40 items have a low pt-biserial score. Indeed, those items can be used to distinguish between students with high and low performance. Therefore, the assessment which resulted from low pt-biserial items was meaningless. Besides, the weakness of classical test theory also appeared in the characteristics items analysis. In this case, the reliability of the test was very difficult to fulfill [8]. Hence, characteristics items analysis needed to apply item response theory further. In order to escalate pt-biserial score, the teacher needed to develop a variety of difficulties of the different type items based on the competencies such that creating transfer problems [37], higher-order thinking skills tasks [38-40], modeling tasks [41] and high element interactivity [42].

5 Conclusion

The score of the validity of the midterm exam script was 0.93 which is categorized as very valid. It means that the midterm exam was feasible to measure student's learning outcomes. Thus, it cannot be used to determine students' knowledge and skills since the coefficient of internal consistency were 0.74. Besides, it cannot be also used as a primary consideration in the process of decision making for student's future.

Based on the level of difficulty, four items were categorized as difficult items, thirty-four items were categorized as medium items and two items were categorized as easy items. Difficult items were taken from arithmetic series of real-world problems, the sum of infinite geometric series, limit with substitutions in form of fractions that consist of the root, and limit with factorizations of its fractions in form of exponentials. The student's difficulties were caused by a lack of concept understanding, modeling skills and mathematical error in calculation. Items with the fit pt-biserial score were 52.5% and the low pt-biserial score was 47.5%. Therefore, the assessment which resulted from the midterm exam was meaningless.

References

[1] Ebel, R. L. & Frisbie, D.A.: Educational Measurement. New Delhi. Prentice-Hall of India (1991)

[2] Bichi, A. A. & Musa, A.: Assessing the correlation between continuous assessment and examination scores of education courses. American International Journal of Research in Humanities. Arts and Social Sciences. 10 (3). 290-294. Retrieved from http://iasir.net/AIJRHASSpapers/ AIJRHASS15-391.pdf (2015)

[3] Popham, W. J.: Instruction that measures up: Successful teaching in the age of accountability. Alexandria. VA ASCD (2009)

[4] Nitko, A. J. & Brookhart, S. M.: Educational assessment of student (6th ed.). Boston. MA. Pearson (2011)

[5] Kartowagiran, B., Munadi, S., Retnawati, H., & Apino. E.: The equating battery test packages of mathematics national examination 2013-2016. SHS Web of Conference. 42. 1-6. doi: 10.1051/shsconf/20184200022 (2018)

[6] Gronlund, N. E.: Assessment of student achievement (6th ed.). Boston. MA: Allyn and Bacon (1998)

[7] Siri, A. & Freddano, M.: The use of item analysis for the improvement of objective examinations. Procedia-Social and Behavioral Sciences. 29. 188-197 (2011)

[8] Retnawati, H.: Analisis kuantitatif instrument penelitian: Panduan untuk peneliti. mahapeserta didik. dan psikometrian. Yogyakarta: Parama Publising (2016)

[9] Quaigrain, K. & Arhin, A. K.: Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education. 4. 1-11. doi:10.1080/2331186X.2017.1301013 (2017)

[10] Retnawati, H.: Teori respon butir dan penerapannya: Untuk peneliti. praktisi pengukuran dan pengujian. mahapeserta didik pascasarjana. Yogyakarta: Parama Publishing (2014)

[11] Santoso, A.: Karakteristik butir tes pengantar statistika sosial berdasarkan teori respon butir. Jurnal Pendidikan Matematika dan Sains. VI(2). 158-168 (2018)

[12] Retnawati, H., Arlinwibowo, J., Sulistyaningsih, E.: The student's difficulties in completing geometry items of national examination. International Journal of Instruction. X(3). 257-276 (2017)

[13] Retnawati, H., Katowagiran, Wibowo, A., & Sulistyaningsih, E.: Why are the mathematics national examination items difficult and what is teachers' strategy to overcome it? International Journal on Trends in Education and Their Implication. Vol 8. Issue 4(3) (2017)

[14] Mehrens, W. A. & Lehmann, I. J.: Measurement and evaluation in education and psychology. New York: Hold. Rinehart and Wilson. Inc. (1973)
[15] Allen, M. J. & Yen, W. M.: Introduction to measurement theory. Monterey. CA; Brooks/Cole Publishing Company (1979)
[16] Moss, P.: Can there be validity without reliability? Educational Researcher. 23(2). 5-12. Retrieved from http://www.jstor.org/stable/1176218 (1994)
[17] Tambychik, T. & Meerah, T. S. M.: Students' difficulties in mathematics problem-solving: What do they say? Procedia Social and Behavioral Sciences. 8. 142151. doi: 10.1016/j.sbspro.2010.12.020 (2010)
[18] Gooding, S.: Children's difficulties with mathematical word problems. Proceedings of the British Society for Research into Learning Mathematics held at The Loughborough University. Saturday 14th November 2009. 31-36. (2009)
[19] Ismail, S. F. Z. H., Shahrill, M., Mundia, L.: Factor contributing to effective mathematics teaching in secondary schools in Brunei Darussalam. Procedia-Social and Behavioral Science. 186. 474-481 (2015)
[20] Maarif, S.: Improving junior high school students' mathematical analogical ability using discovery learning method. International Journal of Research in Education and Science. 2(1). 114-124 (2016)
[21] Alifiani, A.: Penerapan model pembelajaran NHT-TGT untuk meningkatkan motivasi dan pemahaman konsep materi matematika SMA. Jurnal Riset Pendidikan Matematika. 4(1). 11-20. doi:https://doi.org/10.21831/jrpm.v4i1.13100 (2017)
[22] Winarso, W. & Toheri, T.: A case study of misconceptions students in the learning of mathematics; The concept limit function in high school. Jurnal Riset Pendidikan Matematika. 4(1). 120-127. doi:https://doi.org/10.21831/jrpm.v4i1.12060 (2017)
[23] Bradley, J. R., Notar, C. E., Herring, D. F., Eady, C.K.: Teaching mathematics to elementary school students using a variety of tools. Asian Social Science. 4(4). 60-65. Available in http://www.ccsenet.org/journal/index.php/ass/article/view/1600/1514. (2008)
[24] Antony, G. & Walshaw, M.: Characteristics of effective teaching of mathematics: A view from the west. Journal of Mathematics Education. 2(2). 147-164 (2009)
[25] Ali, T.: Exploring students' learning difficulties in secondary mathematics classroom in Gilgit-Baltistan and teachers' effort to help students overcome these difficulties. Bulletin of Education and Research. 33(1). 47-69 (2011)
[26] Palm, T.: Impact of authenticity on sense making in word problem solving. Educational Studies in Mathematics. 67(1). 37–58 (2008)
[27] Verschaffel, L., Greer, B., & De Corte, E.: Making sense of word problems. Lisse: Swets & Zeitlinger (2000)
[28] Nugraheni, E. & Sugiman, S.: Pengaruh pendekatan PMRI terhadap aktivitas dan pemahaman konsep matematika siswa SMP. Pythagoras: Jurnal Pendidikan Matematika. 8(1). 101–108. doi:https://doi.org/10.21831/pg.v8i1.8498 (2013)
[29] Lasmiyati, L. & Harta. I.: Pengembangan modul pembelajaran untuk meningkatkan pemahaman konsep dan minat SMP. Pythagoras: Jurnal Pendidikan Matematika. 9(2). 161-174. doi:https://doi.org/10.21831/pg.v9i2.9077 (2014)
[30] Wijaya, A., Van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A.: Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. The Mathematics Enthusiast. 11(3). 555-584 (2014)
[31] Blum, W.: Can Modelling Be Taught and Learnt? Some Answers from Empirical Research. Trends in Teaching and Learning of Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling. DOI 10.1007/978-94-007-0910-2_3 (2009)
[32] Novferma, N.: Analisis kesulitan dan self-efficacy siswa SMP dalam pemecahan masalah matematika berbentuk soal cerita. Jurnal Riset Pendidikan Matematika. 3(1). 76-87. doi:https://doi.org/10.21831/jrpm.v3i1.10403 (2016)
[33] Maass, K.: What are modelling competencies? ZDM. Vol 38(2) (2006)

[34] Nuryadi, Ardi, Budi, Indaryanti: Kemampuan Pemodelan Matematika Siswa Dengan Strategi Scaffolding with A Solution Plan Pada Materi Trigonometri Di Kelas X SMAN 2 Palembang. Jurnal Gantang. 3. 73-81. 10.31629/jg.v3i2.468 (2018)
[35] Pitriani: Kemampuan pemodelan matematika dalam realistic mathematics education (RME). JES-MAT (Jurnal Edukasi dan Sains Matematika). 2. 10.25134/jes-mat.v2i1.282 (2016)
[36] Polya, G.: How to solve it: A new aspect of mathematical method (2nd ed.). Princeton. New Jersey: Princeton University Press (1973)
[37] Barnett, S. M. & Ceci, S. J.: When and where do we apply what we learn?: A taxonomy for far transfer. Psychological Bulletin. 128(4). 612-637. http://dx.doi.org/10.1037/0033-2909.128.4.612 (2002)
[38] Senk, S. L., Beckmann, C. E., & Thompson, D. R.: Assessment and grading in high school mathematics. Journal for Research in Mathematics Education. 28(2). 187-215. (1997)
[39] Resnick, L. B.: Education and learning to think. Washington. DC: National Academy (1987)
[40] Kantowski, M. G.: Processes involved in mathematical problem-solving. Journal for Research in Mathematics Education (1977)
[41] Doerr, H. M. & Lesh, R.: Models and Modelling Perspectives on Teaching and Learning Mathematics in the Twenty-First Century. Trends in Teaching and Learning of Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling. DOI 10.1007/978-94-007-0910-2_3 (2009)
[42] Paas, F., Renkl, A., & Sweller, J.: Cognitive load theory and instructional design: Recent developments. Educational Psychologist. 38(1). 1–4. https://doi.org/10.1207/S15326985EP3801_1 (2003)