Journal of Educational Research and Evaluation
Volume 6, Issue 3, 2022, pp. 407-414
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://doi.org/10.23887/jere.v6i3.48656

Test Instruments to Measure Non-Routine Mathematics Problem Solving Ability Grade IV Elementary School Students

Parwati Dewi Jatiningtyas[1*], Kartono[2], Budi Naini Mindyarto[3]
[1,2,3] Educational Research and Evaluation, Pascasarjana, Universitas Negeri Semarang, Semarang, Indonesia

ARTICLE INFO

Article history:
Received June 22, 2022
Revised June 27, 2022
Accepted July 29, 2022
Available online August 25, 2022

Kata Kunci :
Non Rutin, R&D, kemampuan matematika, Rasch

Keywords:
Non Routine, R&D, Mathematics Ability, Rasch

ABSTRAK

Guru memiliki peran yang sangat vital dalam terselenggaranya pendidikan yang berkualitas. Berbagai praktik kelas yang digunakan dalam berinteraksi dengan siswa memainkan peran penting dalam memahami konsep matematika dan kinerja keseluruhan dalam matematika. Penelitian ini bertujuan untuk menghasilkan prototipe instrumen tes matematika yang valid dan reliabel untuk siswa kelas IV sekolah dasar. Penelitian ini merupakan penelitian pengembangan. Analisis butir soal instrumen tes menggunakan model Rasch meliputi sepuluh langkah strategi pelaksanaan penelitian dan pengembangan. Responden penelitian terdiri dari 5 orang pada uji ahli, 28 orang pada uji skala kecil, dan 86 orang pada uji skala besar. Pengumpulan data dilakukan dengan teknik wawancara, dokumentasi dan angket. Analisis dilakukan melalui data kuantitatif dan kualitatif. Instrumen pengembangan prototipe tes ini menghasilkan tes Matematika yang lengkap. Reliabilitas yang diuji meliputi reliabilitas antar penilai yaitu 0,613, reliabilitas Cronbach Alpha untuk uji skala kecil 0,81 dan hasil uji skala besar pada reliabilitas Cronbach Alpha 0,78. Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa instrumen tes matematis untuk mengukur keterampilan pemecahan masalah nonrutin yang dikembangkan adalah valid dan reliabel. Instrumen yang ada dapat membantu guru untuk menilai kemampuan matematika siswa kelas IV SD.

ABSTRACT

Teachers have a very vital role in the availability of quality education. Various classroom practices used in interacting with students play an important role in understanding mathematical concepts and overall performance in mathematics. This study aims to produce a prototype of a valid and reliable mathematical test instrument for fourth grade elementary school students. This research is a development research. Item analysis of the test instrument using the Rasch model includes ten steps of research and development implementation strategy. Research respondents consist of 5 people on the expert test, 28 people on the small-scale test, and 86 people on the large-scale test. Data was collected by using interview, documentation and questionnaire techniques. Analysis conducted through quantitative and qualitative data. This test prototype development instrument produces a complete Mathematics test. The reliability tested includes inter-rater reliability, namely 0.613, Cronbach Alpha reliability for small- scale tests 0.81 and large-scale test results on Cronbach Alpha reliability 0.78. Based on the results of research and discussion, it can be concluded that the mathematical test instrument to measure non-routine problem solving skills developed is valid and reliable. Existing instruments can help teachers to assess the mathematics ability of fourth grade elementary school students.

1. INTRODUCTION

The Regulation of the Minister of Education and Culture concerning Standards of Content for Primary and Secondary Education states that the competencies in learning mathematics at the Elementary/Elementary School (SD) level that are expected in each material are: (1) Demonstrate a positive attitude in mathematics: logical, careful and thorough, honest, responsible, and not easily give up in solving problems, as a form of implementation of habits in mathematical inquiry and exploration. (2) Have curiosity, enthusiasm for continuous learning, self-confidence, and interest in mathematics, which is formed through learning experiences. (3) Understand the addition and subtraction of natural numbers. (4) Grouping objects according to the appearance of their shape. (5) Understand the effects of addition and subtraction of a collection of objects. (6) Identify all and parts in daily life. (7) Using pictures or photos to express information and answer questions about it (8) Using concrete models in problem solving (Kolar & Hodnik, 2021; Madrazo & Dio, 2020).

Indonesian students have not been able to apply the basic knowledge they have to solve problems, understand and apply knowledge in complex problems, and make conclusions and make

*Corresponding author.
E-mail addresses: jatiningtyas.pd@gmail.com (Parwati Dewi Jatiningtyas)

generalizations (Ekawati et al., 2020; Nurtanto et al., 2020; Rachmadtullah et al., 2020). The low yield is certainly caused by several factors. One of the factors that occur is that students in Indonesia are less successful in solving contextual problems, because they require reasoning, argumentation, and creativity in solving them (Chong et al., 2019; Kolar & Hodnik, 2021; Pereira et al., 2020). Teachers have a very vital role in the availability of quality education (Dewi & Alam, 2020; Muppudathi et al., 2014; Webb, 2009). Various classroom practices used in interacting with students play an important role in understanding mathematical concepts and overall performance in mathematics. Teacher preparation in teaching certain topics, professionalism, and teacher qualifications greatly affect the improvement of students' abilities in working on problem-solving model questions (Dobber et al., 2017; Nkhoma et al., 2020; Soucy McCrone, 2005). The reality that occurs in schools as written the questions tend to test students' memory a lot with minimal content of high thinking skills (HOTS) (Antara & Dewantara, 2022; Budiman & Jailani, 2014). Basic Competencies (KD) in these mathematics subjects can be used as a reference for developing HOTS in related subjects. Previous research mentions that the use of learning models that train higher-order thinking skills can improve analytical, evaluation, and sustainable creation skills in the student learning process (Diputera et al., 2018).

The teacher's lack of habit of practicing questions that require a high level of reasoning ability is due to several factors, including the teacher has not been able to make questions that are relevant to reasoning indicators, while not many reference to reasoning questions that can be taken and used by the teacher directly (Nurlaily et al., 2019; Pramitasari et al., 2019; Rizta et al., 2013). The difficulty of students in learning mathematics is related to the way the teacher asks questions or provides practice questions in the class that are less varied (Istiyani et al., 2018; Utami & Wutsqa, 2017). Even though the learning process of mathematics should be an active process such as investigating, justifying, exploring, drawing, constructing, using, explaining, developing and proving that takes place socially interactive and reflective (Barnes, 2005; Sukriadi et al., 2015; Szabo et al., 2020).

The development of mathematical test instruments has been carried out several times by previous researcher. One of them developed a math problem to measure the math problem solving ability of class VIII students (Angriani et al., 2018). The material made by the researcher is in the form of comparative learning topics, social arithmetic, lines and angles, triangles and quadrilaterals and data presentation. This study resulted in 15 items of the test instrument developed that met the valid and reliable criteria, then when viewed from the level of difficulty and distinguishing power. Previous research also developed questions to measure math problem solving abilities in elementary school grade V SD N 8 Banyuning (Wiguna et al., 2019). This study resulted in a problem- solving test instrument of 5 items. Based on data analysis and discussion of research results, it can be concluded that the application of brain-based learning can improve the mathematical problem solving ability of fifth grade students of SD N 8 Banyuning (Ariska et al., 2020). Based on the problems and findings of previous researchers, producing a prototype of a valid and reliable mathematical test instrument becomes an urgency that must be done. Therefore this study aims to produce a prototype of a valid and reliable mathematical test instrument for fourth grade elementary school students, analyze content validity and inter-rater reliability, and construct validity analysis.

2. METHODS

The research development procedure is carried out following the development model (Gall et al., 2003) which includes ten steps of research and development implementation strategies as follows: (1) preliminary study, (2) planning, (3) development initial model or product, (4) study of hypothetical models, (5) revision, (6) small scale test, (7) revision of trial results, (8) large-scale testing, (9) revision of the final model, (10) dissemination. Research respondents were 5 people on the expert test, 28 people on the small-scale test, and 86 people on the large-scale test. Data was collected by using interview, documentation and questionnaire techniques. Analysis conducted using quantitative and qualitative data.

Content validity was assessed by 5 professional raters. Reliability was assessed by the Inter-Class Correlation Coefficient (ICC) because the rater was more than 3. Reliability was analyzed using IMB SPSS 25. Construct validity using construct validity in the Rasch model can be seen through the output of Item Polarity. A positive Point Measure Coorelation (Pt. Mea-Corr) value indicates that there is no conflict between the item and the construct being measured. The Mean Square Outfit that is smaller than 1.5 indicates that the measurement value is productive. Quantitative data analysis was carried out by calculating validity and reliability. In addition to calculating validity and reliability, the researchers also analyzed the level of suitability of the items (item fit), the level of difficulty of the items (item measure), the level of individual suitability (person fit), and DIF using the Rasch Model analysis on small & large scale tests in class. IV Elementary Schools in all Negara Districts, Jembrana Regency.

3. RESULT AND DISCUSSION

Results

A preliminary study in the form of a needs analysis was carried out by conducting interviews with the fourth grade mathematics teacher. Interviews were conducted to determine the implementation of mathematics learning and the process in making questions used to assess students' mathematical abilities.

Planning for the mathematical test instrument with non-routine problem solving ability questions that will be developed has been carried out. Planning is done by making a grid of mathematical test instruments with developed non-routine problem solving skills. Researchers prepared as many as 16 questions that have been divided according to the proportion of cognitive domains that have been adjusted to KD. The items that are made are a type of polytomy. The items made are in the form of descriptions (essays). The researcher has prepared 16 questions that have been developed which have been adapted to the Mathematics Basic Competence of Grade IV Elementary School. The proportion of non-routine items that have been adapted to the basic competencies developed in the early stages is described in Table 1.

Table 1. Basic Competencies of Mathematics Test Items in Non-Routine Problem Solving Ability

No	Basic Competence Non-Routine Mathematics Test		Items Description
1	Analyzing the properties of regular polygons and irregular polygons	37,50%	6
2	Explaining and determining the perimeter and area of squares, rectangles and triangles	50%	8
3	Explaining the relationship between lines (parallel, intersecting, coincide) using the model concrete	12,50%	2
	Total		**16**

Based on Table 1, mathematical test instruments in non-routine problem solving abilities that were developed were tested for validity to produce the right test measuring the desired ability. Content validity and reliability tests were carried out on the assessment of the experts. The non-routine problem-solving ability mathematical test instrument developed for fourth grade students was tested for content validity by several experts. The content validity test was carried out by 5 experts to see the suitability of the material, construction, and language of the developed instrument. The non-routine problem solving ability mathematical test instrument validators developed were academics and practitioners. The results of the instrument expert assessment test are then calculated using the Aiken V formula that has been designed in Microsoft Excel calculations. Content validity calculation data using the Aiken's V formula is shown in Table 2.

Table 2. Coefficient of Expert Agreement

Index Item	Aiken's V Index Item	Description	IndexItem	Aiken's V Index Item	Description
1	0.80	Valid	9	0.95	Valid
2	0.85	Valid	10	0.80	Valid
3	0.95	Valid	11	0.80	Valid
4	0.95	Valid	12	0.80	Valid
5	0.95	Valid	13	0.95	Valid
6	0.95	Valid	14	0.95	Valid
7	0.95	Valid	15	0.95	Valid
8	0.80	Valid	16	0.80	Valid

Base on Table 2, coefficient of agreement the experts obtained were then compared with *Validity* Coefficient *Table.* The Aiken version items are said to be valid if the coefficient of expert agreement is > V. Aiken's V table shows that the V value for 5 assessors and 5 scale options is validity > 0.80. On the other hand, the coefficient of expert agreement is said to be invalid or not 0.80 for an error probability of 0.04. The construct validity test proves that the indicators developed are really factors that measure ability variables. Mathematics in problem solving non-routine problem is solving abilities. The data used to test the construct validity are test scores obtained on a small scale test. The analysis was carried out using Winsteps software. The construct validity of the Rasch model can be seen through the Item Polarity. A positive Point Measure Correlation (Pt. Mea Corr) value indicates that there is no conflict between the item and the construct being measured. Undimensionality test is show in Table 3.

Table 3. Unidimensionality Test

Unidimensionality Test	Empirical
Total raw variance in observations	100
Raw variance explained by measures	68
Raw variance explained by persons	11.1
Raw variance explained by items	56.8
Raw variance explained total	32
Unexplned variance in 1st contrast	5.7
Unexplned variance in 2nd contrast	5.6
Unexplned variance in 3rd contrast	4.8
Unexplned variance in 4th contrast	2.9
Unexplned variance in 5th contrast	2.6

Based on Table 3, it can be seen that the raw variance data measurement results are 68.%. The value is not much different from the expected value, which is 64.91%. This shows that the unidimensionality requirement of 20% can be met. If the unidimensionality value is 40% better, and if the unidimensionality value is 60% it is special. Unexplained variance is 5.7%, 5.6%, 4.8%, 2.9%, and 2.6%, respectively. This shows that the variance that cannot be explained by the instrument ideally does not exceed 15% so that the instrument is in the good category.

Item Characteristics of Test Instruments Analyzed with Item Response Theory or Rasch

The analysis of the Rasch model was carried out on the results of small-scale test data. In the small-scale test, there are 16 questions. The item analysis carried out is an analysis of the level of difficulty of the items (item measure), analysis of the level of suitability of items (item fit), analysis of individual capabilities (person measure), and analysis of the level of individual suitability (person fit). Item Fit explains whether the item functions normally to take measurements or not. *Outfit means- square* (MNSQ), *z-standard* (ZSTD), and *point measure correlation* (Pt. Mean Corr) are the criteria used to see the level of item suitability. If the item does not meet the three criteria, then the item is not good enough so it needs to be repaired or replaced. The results of the Item Fit output are described in Table 4.

Table 4. Output Item Fit

Item	Outfit		Pt. Measure Corr	Item	Outfit		Pt. Measure Corr
	MNSQ	ZSTD			MNSQ	MNSQ	
9	3.30	1.7	-0.0.6	4	0.81	-0.7	0.62
14	3.22	1.6	-0.03	3	0.74	-0.9	0.80
16	3.01	2.9	-0.02	7	0.75	-0.6	0.70
1	1.97	3.0	0.41	10	0.72	-1.1	0.81
2	1.92	1.7	0.11	12	0.60	-1.3	0.61
8	1.91	1.5	0.00	5	0.50	-2.0	0.60
13	1.15	0.7	0.73	15	0.48	-1.9	0.68
11	0.95	-0.2	0.74	6	0.48	-1.8	0.68

Base on Table 4, the accepted MNSQ value is 0.5 MNSQ 1.5. The accepted ZSTD value is -2.0 < ZSTD < 2.0. Pt value. Measure Corr received is 0.4 < Pt. Measure Corr < 0.85 Items are considered *fit* or accepted if they meet one of the three criteria. The results of the Item Fit showed that of the 16 items tested, 10 items were accepted, and 6 items were discarded. The researcher discarded item numbers 9, 14, 16, 1, 2, and 8 because the items were considered misfits.

In Item Measure, the standard deviation is 1.80. Value > 0.0 logit + SD is a difficult question. Values between 0.0 logit + SD to 0.0 logit – SD can be categorized as items in the medium category. Value < 0.0 logit – SD is categorized as an easy question. The results of Item Measure and difficulty level classification are shown in Table 5.

Table 5. Output Item Measure

Item	*Measure*	Conclusion	Item	*Measure*	Conclusion
9	3.08	Difficult	15	0.24	Moderate
14	3.08	Difficult	5	-0.18	Moderate
8	1.56	Moderate	1	-0.79	Moderate
2	1.37	Moderate	4	-1.37	Moderate
16	1.37	Moderate	13	-2.16	Easy
7	0.61	Moderate	10	-2.30	Easy
6	0.36	Moderate	11	-2.36	Easy
12	0.36	Moderate	3	-2.85	Easy

Based on Table 5 on the small scale test there are 2 items with difficult categories, namely items number 9 and 14. A total of 10 items with a moderate level of difficulty category. The remaining 4 items in the easy category, namely items number 10, 11, and 3. The reliability of the mathematics test instrument to measure non-routine problem solving abilities for fourth grade elementary school. The tested reliability is a) interrater reliability; b) small-scale test reliability; c) large-scale test reliability. Consecutive tests on the reliability value in each successive trial is 0.613; 0.81; 0.78. Based on the analysis, it is obtained in Table 6.

Table 6. Estimated Of Reliability

The Experimental Phase	Reliability	Number of Items
Judges	0.613	16
Small-Scale Test	0.810	16
Large-Scale Test	0.780	10

Interrater Reliability

The results of the Intraclass Correlation Coefficient (ICC) is show in Table 7.

Table 7. Interclass Correlation Coefficient

		95% Confidence Interval		F Test with True Value 0			
	Interclass Correlation	Low Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0.241	0.064	0.509	3.897	15	60	0.00
Average Measures	0.613	0.254	0.838	3.897	15	60	0.00

Base on Table 7 analysis obtained reliability rxy of 0.613. According to the reliability criteria are considered moderate in the range of values of 0.4 r < 0.6. Reliability is considered high if it is in the range of 0.6≤ r 8. Based on this classification, the reliability of 0.613 is included in high reliability. Interclass correlation coefficient.

Small Scale Test Reliability

The value of person reliability is 0.79 and item reliability is 0.94, indicating that the consistency of the answers from students is sufficient, but the quality of the items in the instrument has a special reliability aspect. The reliability *Cronbach's Alpha is* 0.81.value *Cronbach's Alpha* in the Rasch Model shows the level of interaction between the person and the items as a whole.

Large-Scale Test Reliability

The value of person reliability 0.78 and item reliability 0.97 indicates that the consistency of the answers from students is sufficient, but the quality of the items in the instrument has a special reliability aspect. The reliability *Cronbach's Alpha is* 0.78.value *Cronbach's Alpha* in the Rasch Model shows the level of interaction between the person and the items as a whole.

Discussion

The instrument produced in this study is a test instrument to measure the non-routine mathematical problem solving ability of fourth grade elementary school students, totaling 10 questions. The form of the test instrument made is a question with a structured response in the form of a description (Rusilowati, 2014; Sumintono & Widhiarso, 2014). The instrument is made based on the basic competencies of the fourth semester II students. The instrument is made based on the cognitive level. The results of the analysis of the characteristics of items using the Rasch model on a large-scale test show that there are 2 items in the difficult category, namely items number 10 and 8. A total of 6 items are in the category of moderate difficulty level. The remaining 2 items in the easy category, namely items number 1 and 6. A total of 10 items tested have met the fit criteria according to the Rasch model.

The resulting instrument has been tested including content validity, interrater reliability, construct validity to item characteristics based on the Rasch model. The instrument has fulfilled content validity as evidenced by the acquisition of an expert agreement index in the range of 0.80-0.95. Based on the small- scale test of the 16 questions tested on the small-scale test, it was found that 10 items were not eliminated to be continued on the large-scale test. The construct validity of the Rasch model shows the variance values of 5.7%, 5.6%, 4.8%, 2.9%, and 2.6%, respectively. This shows that the variance that cannot be explained by the instrument ideally does not exceed 15% so that the instrument is in the good category. The reliability coefficient for each trial in a row is 0.613; 0.81; 0.78.

It is in line with previous research that state This study investigated the influence of reading comprehension skill, Mathematics self-efficacy perception and Mathematics attitude on non-routine Mathematics problem-solving skill (Öztürk et al., 2020). The result of analysis showed that for middle-school students' non-routine Mathematics problem-solving skills, all related factors–reading comprehension, Mathematics self-efficacy perception and Mathematics attitude were significant. Reading comprehension skills and Mathematics self-efficacy perception significantly predicted problem-solving skills, and both predictors explained a total of 22% of the total variance. Moreover other researcher also conducted research in order to investigate two types of strategy flexibility, namely inter-task flexibility and intra-task flexibility (Elia et al., 2009). Findings showed that students rarely applied heuristic strategies in solving the problems. Among these strategies, the trial-and-error strategy was found to have a general potential to lead to success. The two types of flexibility were not displayed to a large extent in students' strategic behavior.

The implication of this research is to provide information related to the development of test instruments to measure non-routine mathematics problem solving abilities of grade IV elementary school students. This information is very useful for educators, especially mathematics teachers, in applying the non-routine mathematics problem solving ability test measure instrument. This study is not perfect and still has many limitations. One of the limitations of this study is the subject of research involving only one school. Therefore, future research is expected to be able to deepen research topics related to test instruments to measure non-routine mathematics problem solving abilities. In addition, researchers are also expected to be able to involve more subjects and consider other factors related to test instruments to measure non-routine mathematics.

4. CONCLUSION

Based on the results of the research and discussion, it can be concluded that the test instrument to measure the non-routine problem solving ability that was developed is valid and reliable. The test results of a valid and reliable test instrument can provide an assessment that is in accordance with the students' abilities in the cognitive domain. Existing instruments can help teachers to assess the mathematical ability of fourth grade elementary school students.

5. REFERENCES

Angriani, A. D., Nursalam, N., Fuadah, N., & Baharuddin, B. (2018). Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Pemecahan Masalah Matematika Siswa. *AULADUNA: Jurnal Pendidikan Dasar Islam*, *5*(2), 211. https://doi.org/10.24252/auladuna.v5i2a9.2018.

Antara, I. G. W. S., & Dewantara, K. A. K. (2022). E-Scrapbook: The Needs of HOTS Oriented Digital Learning Media in Elementary Schools. *Journal for Lesson and Learning Studies*, *5*(1), 71–76. https://doi.org/10.23887/jlls.v5i1.48533.

Ariska, N., Sariyasa, & Putrayasa, I. (2020). Pengembangan Instrumen Kemampuan Pemecahan Masalah Matematika dan Self-Efficacy Pada Siswa Kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *10*(1), 11–20. https://doi.org/10.23887/jpepi.v10i1.3502.

Barnes, H. (2005). The theory of Realistic Mathematics Education as a theoretical framework for teaching low attainers in mathematics. *Pythagoras*, *0*(61), 42–57. https://doi.org/10.4102/pythagoras.v0i61.120.

Budiman, A., & Jailani, J. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots) Pada Mata Pelajaran Matematika SMP Kelas Viii Semester 1. *Jurnal Riset Pendidikan Matematika*, *1*(2), 139. https://doi.org/10.21831/jrpm.v1i2.2671.

Chong, M. S. F., Shahrill, M., & Li, H. C. (2019). The integration of a problem-solving framework for Brunei high school mathematics curriculum in increasing student's affective competency. *Journal on Mathematics Education*, *10*(2), 215–228. https://doi.org/10.22342/jme.10.2.7265.215-228.

Dewi, E. R., & Alam, A. A. (2020). Transformation Model for Character Education of Students. *Cypriot Journal of Educational Sciences*, *15*(5), 1228–1237. https://doi.org/10.18844/CJES.V15I5.5155.

Diputera, A. M., Setyowati, D. L., & Susilaningsih, E. (2018). Higher-order thinking skills of junior high school students. *The Online Journal of New Horizons in Education*, *8*(3), 61–67. www.tojned.net/journals/tojned/articles/v08i03/v08i03-07.pdf.

Dobber, M., Zwart, R., Tanis, M., & van Oers, B. (2017). Literature review: The role of the teacher in inquiry-based education. *Educational Research Review*, *22*, 194–214. https://doi.org/10.1016/j.edurev.2017.09.002.

Ekawati, R., Susanti, S., & Chen, J.-C. (2020). Primary students' mathematical literacy: A case study. *Infinity Journal*, *9*(1), 49–58. https://doi.org/10.22460/infinity.v9i1.p49-58.

Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education*, *41*, 605–618. https://doi.org/10.1007/s11858-009-0184-6.

Gall, M. G., Gall, J. P., & Borg, W. R. (2003). *Educational Research: An Introduction, 7th Edition*. Longman Publishing.

Istiyani, R., Muchyidin, A., & Raharjo, H. (2018). Analisis Miskonsepsi Siswa Pada Konsep Geometri Menggunakan Three-Tier Diagnostic Test. *Cakrawala Pendidikan*, *37*(2), 223–236. https://doi.org/10.15713/ins.mmj.3.

Kolar, V. M., & Hodnik, T. (2021). Mathematical literacy from the perspective of solving contextual problems. *European Journal of Educational Research*, *10*(1), 467–483. https://doi.org/10.12973/EU-JER.10.1.467.

Madrazo, A. L., & Dio, R. V. (2020). Contextualized learning modules in bridging students' learning gaps in calculus with analytic geometry through independent learning. *Journal on Mathematics Education*, *11*(3), 457–476. https://doi.org/10.22342/jme.11.3.12456.457-476.

Muppudathi, G., Education, A., John, S., & College, D. B. (2014). Role of Teachers on Helping Slow Learners to Bring Out Their Hidden Skills. *Ijsr -International Journal of Scientific Research*, *3*(2277), 98–99. https://www.academia.edu/download/34288348/Role_of_Teachers_on_Helping_Slow_Learners_to_Bring_out_their_Hidden_Skills.pdf.

Nkhoma, C., Nkhoma, M., Thomas, S., & Le, N. Q. (2020). The Role of Rubrics in Learning and Implementation of Authentic Assessment: A Literature Review. *Proceedings of the 2020 InSITE Conference*, *January*, 237–276. https://doi.org/10.28945/4606.

Nurlaily, V. A., Soegiyanto, H., & Usodo, B. (2019). Elementary school teacher's obstacles in the implementation of problem-based learning model in mathematics learning. *Journal on Mathematics Education*, *10*(2), 229–238. https://doi.org/10.22342/jme.10.2.5386.229-238.

Nurtanto, M., Pardjono, P., Widarto, W., & Ramdani, S. D. (2020). The effect of STEM-EDP in professional learning on automotive engineering competence in vocational high school. *Journal for the Education of Gifted Young Scientists*, *8*(2), 633–649. https://doi.org/10.17478/JEGYS.645047.

Öztürk, M., Akkan, Y., & Kaplan, A. (2020). Reading comprehension, Mathematics self-efficacy perception, and Mathematics attitude as correlates of students' non-routine Mathematics problem-solving skills in Turkey. *International Journal of Mathematical Education in Science and Technology*, *51*(7), 1042–1058. https://doi.org/10.1080/0020739X.2019.1648893.

Pereira, S., Rodrigues, M. J., & Vieira, R. M. (2020). Scientific literacy in the early years–practical work as a teaching and learning strategy. *Early Child Development and Care*, *190*(1). https://doi.org/10.1080/03004430.2019.1653553.

Pramitasari, K., Usodo, B., Subanti, S., Magister, P., Matematika, P., Sebelas, U., & Surakarta, M. (2019). Proses Pembelajaran Matematika Untuk Siswa Slow Learner Di Kelas Inklusi Smp Negeri 7 Klaten Kelas Viii. *Jurnal Elektronik Pendidikan Matematika*, *3*(7), 777–786. https://jurnal.fkip.uns.ac.id/index.php/s2math/article/view/6494.

Rachmadtullah, R., Syofyan, H., & Rasmitadila. (2020). The role of civic education teachers in implementing multicultural education in elementary school students. *Universal Journal of Educational Research*, *8*(2). https://doi.org/10.13189/ujer.2020.080225.

Rizta, A., Zulkardi, Z., & Hartono, Y. (2013). Pengembangan Soal Penalaran Model Timss Matematika Smp. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *17*(2), 230–240. https://doi.org/10.21831/pep.v17i2.1697.

Rusilowati, A. (2014). *Pengembangan Instrumen Penilaian*. UNNES PRESS.

Soucy McCrone, S. (2005). The Development of Mathematical Discussions: An Investigation in a Fifth-Grade Classroom. *Mathematical Thinking and Learning*, *7*(2), 111–133. https://doi.org/10.1207/s15327833mtl0702_2.

Sukriadi, S., Kartono, K., & Wiyanto, W. (2015). Analisis Hasil Penilaian Diagnostik Kemampuan Berpikir Kritis Matematis Siswa Dalam Pembelajaran Pmri Berdasarkan Tingkat Kecerdasan Emosional. *Unnes Journal of Mathematics Education Research*, *4*(2), 139–145. https://journal.unnes.ac.id/sju/index.php/ujmer/article/view/9841.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Szabo, Z. K., Körtesi, P., Guncaga, J., Szabo, D., & Neag, R. (2020). Examples of problem-solving strategies in mathematics education supporting the sustainability of 21st-century skills. *Sustainability (Switzerland)*, *12*(23), 1–28. https://doi.org/10.3390/su122310113.

Utami, R. W., & Wutsqa, D. U. (2017). Analisis kemampuan pemecahan masalah matematika dan self-efficacy siswa SMP negeri di Kabupaten Ciamis. *Jurnal Riset Pendidikan Matematika*, *4*(2), 166. https://doi.org/10.21831/jrpm.v4i2.14897.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, *79*(1), 1–28. https://doi.org/10.1348/000709908X380772.

Wiguna, S. G. A., Widiana, I Wayan, D. N., & Sudana. (2019). Penerapan Model Pembelajaran Berbasis Masalah Untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Kelas V Sekolah Dasar. *Jurnal Pendidikan Guru Sekolah Dasar*, *4*(3), 390–399. https://doi.org/10.17509/jpgsd.v4i3.22948.