



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



Validity and Reliability of Students' Mathematical Process Rubric (Prom3) based on many-Facet Rasch Model (MRFM).

Normarina Abd Rahman, Siti Eshah Mokshein, Hishamuddin Ahmad

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v11-i2/9207> DOI:10.6007/IJARBSS/v11-i2/9207

Received: 14 December 2020, **Revised:** 16 January 2021, **Accepted:** 31 January 2021

Published Online: 25 February 2021

In-Text Citation: (Rahman et al., 2021)

To Cite this Article: Rahman, N. A., Mokshein, S. E., & Ahmad, H. (2021). Validity and Reliability of Students' Mathematical Process Rubric (Prom3) based on many-Facet Rasch Model (MRFM). *International Journal of Academic Research in Business and Social Science*, 11(2), 1315-1331.

Copyright: © 2021 The Author(s)

Published by Human Resource Management Academic Research Society (www.hrmars.com)

This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: <http://creativecommons.org/licenses/by/4.0/legalcode>

Vol. 11, No. 2, 2021, Pg. 1315 - 1331

<http://hrmars.com/index.php/pages/detail/IJARBSS>

JOURNAL HOMEPAGE

Full Terms & Conditions of access and use can be found at
<http://hrmars.com/index.php/pages/detail/publication-ethics>



INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES



www.hrmar.com

ISSN: 2222-6990

Validity and Reliability of Students' Mathematical Process Rubric (Prom3) based on many-Facet Rasch Model (MRFM).

Normarina Abd Rahman, Siti Eshah Mokshein, Hishamuddin Ahmad

Faculty of Human Development, Universiti Pendidikan Sultan Idris, 35900, Tanjong Malim, Perak, Malaysia

Abstract

The objective of this research was to determine the validity of Students' Mathematical Process Rubric (ProM3) using Many-Facet Rasch Model (MFRM). The data were gathered from 7 raters marking 188 scripts of Form 1 students from a number of boarding schools in the middle and northern zone of Malaysia. ProM3 Rubric was used to analyse students' responses in problem solving and reflective writing tasks for 29 criteria, that is in five dimensions of the mathematical process, namely connection, representation, communication, reasoning and problem solving. MFRM was used to analyse the data to look into three facets; student ability, rater severity and item difficulty. The findings indicated that the accurate index measuring students' ability facet was between -3.48 until 4.71 logit, raters' severity facet between -0.59 until 0.74 logit, and items/criteria's difficulty facet at -2.38 (problem identifying) until 1.45 (quantitative reasoning) logit. The high validity and reliability construct based on MFRM signified that the model can measure the accuracy of each facet score.

Keywords: Mathematical Process, Problem Solving, Rater, Many-Facet Rasch Model (MFRM), Rubric Validation, Performance Assessment, Alternative Assessment

Introduction

The main objective of teaching mathematics is to develop thinking. Students who think mathematically, give reasons and justifications before they do something (Sundstrom, 2014). They examine something from the grounds for a thoughtful decision, and do not arbitrarily guess or repeat what is taught without assessing its relevance (Akdemir, 2018). PISA which is OECD's Programme for International Student Assessment drafted 2021's mathematics framework to assess mathematical literacy as an individual's capacity to reason mathematically and to formulate, employ, and interpret mathematics to solve problems in a variety of real world contexts. It includes concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It helps individuals know the role that mathematics plays in the world and make the well-founded judgements and decisions needed by constructive, engaged and reflective 21st Century citizens (OECD, 2018).

While in Malaysia, in order to accomplish the second wave in Malaysian National Education Blueprint (2003-2025), Malaysian Education Ministry introduced Standard Curriculum for Secondary School (KSSM) to be implemented since 2017, which replaced the Integrated Curriculum for Secondary School (KBSM) that was used since 1989. The curriculum transformation through KSSM focuses on the students' intelligence development in favour of the 21st century learning needs that is to develop a society, which is intelligent, creative and innovative (Malaysian Education Ministry, 2016)

Standard Curriculum for Secondary School (KSSM) for Mathematics aims to shape individuals to think mathematically, creatively and innovatively, as well as competent in effectively and responsibly applying mathematical knowledge and skills in solving problems and making decisions, based on values and attitudes as to be able to potentially solve and deal with challenges in daily life, in favour of the advancement in science and technology and the 21st century's challenges.

There are a few changes made to ensure that the education can empower students to think and solve different and challenging situations. Explicit focus was given to the students' mathematical processes, including problem solving, reasoning, mathematical communication, making connection, and representation which requires teachers to conduct teaching and learning to enable them understand the concepts in depth and meaningfully. This shift is crucial so mathematics would not be viewed as a mechanical subject by students. They require a balance between procedural and conceptual knowledge of mathematics.

Students who mastering the syllabus without understanding the mathematical process can make them perceive the subject as nothing more than arithmetical structures which they have to learn, memorize, and express in examination (Fuson, Kalchman, & Bransford, 2005). This contradicts to the goals of KSSM that is to produce students who have mathematical thinking. Students need to realise the relationship between mathematics and everyday life. Previous research has shown that the ability of students to make such connection, represent information in the form of symbols and visuals, communicate mathematically in explaining information that are reasoned, hence solve the problems being faced will be meaningful and enhance students' learning (Edelen & Bush, 2020; Dollah, Saad, Abdullah, & Yusof, 2016; Rasiman, Prasetyowati, & Kartinah, 2020; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2014).

Mathematical Process

The development of Mathematical Process rubric (ProM3) is based on the constructivism approach with cognitive theory and other learning theories as its foundation. Analysing the cognitive development and learning theory worked by scholars such as Piaget, Bruner, and Van Hiele shapes the understanding of students' mathematical process in making connection, representation, reasoning via mathematical communication, hence solving problems. The mathematical process is measured to understand on students' ways to process mathematical knowledge and translated through application in solving problems. Looking into the literature allows the criteria of students' mathematical process to be determined, and presented as attributes based on the operational definition of students' mathematical process in developing scoring rubric for mathematical process.

Mathematical process that students undergo is evident to be enhancing their learning when they are able to relate their experience to the knowledge learnt in the classroom (McLeod, 2017). This is on par with the reflective learning theory in which reality is perceived as a part of students' mental development. Learning takes place when students can identify

assimilation and accommodate and modify the scheme that they readily possess (Piaget, 2008). Mathematical process including solving problems takes the subject beyond the context of worksheets and the traditional concept in which each question has a specific solution and working scheme, thus the learning activities focus on students exploring reality (Schoenfeld, 2013). This is rather a new approach which requires students to face uncertainty, risks, and overcome challenges in making connection, representation, mathematical communication, reasoning, hence solving the proposed problems with different strategies with varying possibilities of either achieving the accurate solution, or otherwise (Usta, 2020).

It has been suggested that mathematical process is an important skill that students should acquire by the end of learning mathematics. Studies have found that when students are able to relate between a concept and mathematical knowledge, represent idea and mathematical strategies, communicate well and confidently, make reasoning, and solve problems that are challenging, they are considered to be demonstrating important processes in enhancing and strengthening their understanding of mathematics (Ayllon, Gomez, & Ballesta-Claver, 2016; Kadir & Parman, 2013; Krawec, 2010) The students' mastery in the mathematical process can be proven by looking at the students' ability to work systematically in solving problems, trying different strategies, restarting when the previous strategy did not work, and testing conjecture such as assumption or idea (Sujadi & Masamah, 2017).

Malaysia students' performance in TIMSS 2015 indicated that less than 3% of the participating students were able to apply their mathematical knowledge in solving a non-routine based problem (Yee, Tze, & Abdullah, 2017). Teachers are expected to help improve these students' skills. In order to do so, it is important for teachers to know the current level of their students' ability. How can teachers determine the level of students' mastery in solving non-routine problems, making reason and communicate to explain their findings? Mathematical process cannot be measured only by referring to the method and mathematical calculation. Li and Schoenfeld (2019) identified that mathematics in schools are taught by giving more emphasis on the procedures to solve problems as practiced. However, in reality, problems can occur in daily activities, different field and as well as the field of mathematics itself. The key to being successful in mathematics in school is to abide by the taught procedures and contexts to answer questions, while the key to effectively process mathematical thinking is the confidence and willingness to try out of the box (Klerlein & Hervey, 2000).

Krawec (2010) suggested that one working scheme pertaining mathematical understanding from a constructivism perspective is by making connection between the intrinsic and extrinsic representation of the mathematical idea. The scheme indicated that in order for one to think and communicate mathematically, he or she should represent the idea in some manner. Communication requires good extrinsic modelling such as verbal, symbol, drawing and concrete object. To measure students' mathematical process, their process of mathematical thinking should be translated into words and visuals. The mathematical idea will be easier to comprehend when it can be explained. When students are able to explain their idea, they will better appreciate the concept and knowledge being introduced to them (Krawec, 2010).

A research pertaining the ability of academically excellent Form 4 students in solving non-routine Algebraic problems by Adnan and Jalil (2016) indicated that a structured problem solving activity allows students to translate their understanding on the situation through verbal and visual representation, hence plan and execute solving, and explain the process and justification as to why such decision is made and conclude the finding.

In this present study, a performance task was developed to allow students demonstrate their mathematical processes' skills in solving the given problem. Students needed to process the information based on the situation with provided steps to allow them truly understand the problem statement in the form of communication and visual, hence encourage them to strategize solution, solve the problem, explain the steps to solve, justify the steps and finally state the accurate answer. Each step allows different mathematical process to be demonstrated by the students and scored by the teacher using the scoring rubric which was developed to measure the students' mathematical process.

In order to score the students' mathematical process, the final answer was not the main measure, but the steps demonstrated via their thinking skills and process in order to solve the given problem statement (Brookhart, 2010). The situations were designed to be relevant to the students' lifestyle so they could utilise their full potential and concentration in solving the problem. This also will allow teacher to gather sufficient data to measure the students' mathematical process, thus be used to make a better decision in relation to teaching and learning process of the students.

Conceptual Definition

Mathematical process based on curriculum is the process that supports the effective learning of mathematics including the process of connection, representation, mathematical communication, reasoning, and problem solving. All the five mathematical process are interrelated and should be conducted via integration across the curriculum.

Operational Definition

Students' mastery of mathematical process in this research was measured based on their response in solving given problem statements and reflective writing. The students' mathematical process construct was then scored based on a developed scoring rubric consisting of 29 criteria as below:

- a. Connection; making relation with current knowledge, based on contexts, with the subject, and everyday activities.
- b. Representation; using representation, listing the importance of representation, using different representation (figures, tables, sketches), using appropriate representation, represent thoughts, planning and execution in solving problems and interpret representation.
- c. Mathematical communication; writing structured answers, using mathematical terms, using mathematical symbols, explain mathematical idea and analyse mathematical thinking and other students' strategies.
- d. Reasoning; using data to test idea, using data to reason quantitatively, abstract reasoning, explain based on the needs of problem and support the solution, identify the trend and use different reasoning methods (inductive, deductive, abstract, and quantitative).
- e. Problem solving; identify problem, extract information, arrange information, strategize, verify strategy, construct solution, interpret solution, as well as recheck and reflect.

Research Questions

The research was conducted to develop an instrument to measure students' mathematical process with good psychometric qualities, has validity, reliability and contain items with

appropriate statistics based on Rasch measurement model. Specifically, this research was conducted to find the answers for the following questions:

- i. What are the values of reliability for each facet, including students, raters, and mathematical process criteria?
- ii. What are the values of validity for each facet based on Infit MNSQ and Outfit MNSQ values for:
 - a) Students; Can the measurement used differentiate students based on their abilities?
 - b) Raters; (i) What are the values of severity among the raters? (ii) Are there any differences? (iii) Is there any consistency among the raters?
 - c) Criteria of Mathematical process; (i) What is the positioning of difficulty for each of the measured criteria? (ii) What is the accuracy of the measurement conducted? (iii) Does the response of each criterion correlate with the expected measurement model of Rasch?

Methodology

The methodology for this research was constructed to develop an instrument to measure students' mathematical process in solving problems and reflective writing, via a scoring rubric. The data gathered could be analysed to achieve the important objectives in developing the instrument that was to assess the validity and reliability of the instrument, and to construct a descriptive statistical profile of the research sample to measure students' mathematical process.

Respondents

The research included 7 raters; six mathematics teachers of different levels (2 upper secondary, 2 lower secondary, and 2 primary school) and one teacher-trainee specialising in mathematics. A total of 188 Form 1 students from boarding schools in middle and northern zones were chosen as the respondents. These students were required to achieve a minimum of C grade to be enrolled in those schools. Therefore, the minimum mastery in mathematics could ensure that these respondents could be accessed appropriately to reflect the students' mastery of the mathematical process accordingly and avoid the notion underperformed students.

Instrumentation

In this research, the instrument was developed based on the ADDIE model which consists of five phases to assess the students' performance in given tasks, hence determine the students' mastery of mathematical process (Branch, 2009). In the analysis phase, the issue and main problems in measuring the students' mathematical process were identified by analysing the documents with the focus on students' performance in national and international level, the aspect of students' mastery of mathematical process, standard mathematical process in KSSM, choices of performance task to assess their mathematical process and identify approaches that were used to construct the tasks and scoring rubric.

The second phase in the research focused on the design of scoring rubric. The researcher adapted appropriate steps and approaches to be used to construct the rubric based on previous studies, including Lantz (2004), Moskal (2003), Stergar (2005) also Arter and McTighe (2001). Analysis of the standard document (KSSM) and literature on previous studies on the mathematical process was conducted to identify appropriate measuring

criteria to be developed based on the operational definition of students' mathematical process. Literature review was also conducted to identify the variables involved in measuring mathematical process based on the theory of cognitive development, hence set the aim of developing the instrument and target group being assessed, and define each dimension for the criteria selected.

The third phase was the development phase including the translation based on the design of the scoring rubric. In this phase, the scoring rubric was developed based on the scoring template consisting of 29 criteria covering five main dimensions in mathematical process including connection, representation, mathematical communication, reasoning and problem solving. Each criterion was included with the description which illustrates the level of quality for each observed indicator.

The draft was then shown to a group of experts including lecturers specializing in the field of educational measurement and Mathematics education, as well as field experts (Form 1 Mathematics teachers) to acquire content and face validity. A pilot test was conducted in this phase in a smaller scale among the chosen respondents to gather and assess information pertaining the suitability of language, the duration of task and scoring, and the functionality of the instrument. The validity of the developed rubric was ensured to be at a high level, so it could be used in the actual assessment to measure students' mathematical process.

The fourth phase was the implementation phase, in which the instrument was distributed to the sample groups to determine the items' quality and effectiveness of the assessment scores. In this implementation phase, the students were briefed about the tasks and scoring rubric criteria to give them information on the purpose and aim of the assessment conducted. The students were given the opportunity to demonstrate their mathematical process by solving the given problems and writing reflections.

Their responses were then assessed and scored by a number of raters based on the scoring rubric. The assessing plan was constructed based on the needs of MRFM model. The assessing plan used was nested judging plan (Lunz, 1997). According to Linacre (2018) to construct a good measurement, there must be a good relationship between each facet to ensure each tested parameter was not tested without a structured reference.

The analysis of 29 criteria on 188 students by 7 raters yielded a total of 38 164 ratings. However, to improve the quality of scoring by the raters, the researcher had to consider a few factors including the limited time and other workloads of the selected raters. Therefore, an assessment plan was set as in Table 1 to ensure each student's script can be at least marked by two raters and ensure there was an overlap in combinations of the involved raters. Each student's script was marked separately based on the type of scale (quality and quantity), thus only 5 452 ratings were achieved (29 criteria x 188 students). The limited assessment plan allowed each rater to assess all the five dimensions including 29 criteria with only one out of seven assessing work. Such plan could be executed well and used to analyse the effect of raters on students' performance (Lunz, 1997).

Table 1. Raters' Assessment Plan

Raters' Quantity Scale (Criteria 1-15)	Raters' Quality Scale (Criteria 16-29)							Total Students	Total Rating
	A	B	C	D	E	F	G		
A		/	/	/	/	/	/	38	822
B	/		/	/	/	/	/	25	725
C	/	/		/	/	/	/	34	776
D	/	/	/		/	/	/	33	775
E	/	/	/	/		/	/	18	746
F	/	/	/	/	/		/	16	828
G	/	/	/	/	/	/		24	780
Total	18	25	19	20	34	42	30	188	5452

The collected data were analysed and assessed in the final phase; evaluation phase. In this phase, the data were analysed based on the many-facet Rasch model to determine the functionality of the item, biasness of the raters, and functionality of the ranking scale, hence determine the students' mastery level of mathematical process.

Research Instrument

Students' mathematical process was measured based on their responses in answering the given mathematical problems and reflective writing. This was then scored according to the scoring rubric constructed which consists of 29 criteria adapted from previous theoretical and empirical literature. The criteria's descriptor or attribute to construct the scoring rubric were presented below:

Table 2: Sources of Reference for The Criteria used to Assess Mathematical Process

Dimension	Criteria's Descriptor/ Attribute	Source
Connection	Making connection with existing knowledge	Boud, Keogh, & Walker (1985)
	Making connection with context	Wen, Tajudin, & Masri, (2017)
	Making connection with the subject	
	Making connection with daily activities	Kolb (2012)
Representati on	Using representation	Carson (2007)
	Listing the importance of representation	Bruner (1960)
	Using various representation	Mcleod (2018)
	Using appropriate representation	Krawec (2010)
	Representing thought process and execution in solving problems	Bruner (1960)
	Interpreting representation	Scusa (2008)
Mathematica l	Writing numerical answer	
	Using Mathematical terms	Rothstein, Rothstein, &
Communicati on	Using Mathematical symbols	Lauber (2007)
	Explaining Mathematical idea	Wood (2001)
	Analysing Mathematical thinking and other students' strategies	Dollah, et al., (2016)
Reasoning	Using data to test idea	Noor, Zamri & Eu (2016)
	Using data for quantitative reasoning	
	Abstract reasoning	
	Explanation matches the problem's needs and support solution	(Jonassen, 2004)
	Identifying trend	Dumas, Alexander, & Grossnickle (2018)
	Using various reasoning techniques (inductive, deductive, abstract, and quantitative).	Ghazi (2014)
Problem Solving	Identifying problem	Jonassen (2004)
	Extracting information	Piaget (2008)
	Sequencing information	Piaget (2008)
	Strategizing	Costa (2004)
	Verifying strategy	Costa & Kallick (2008)
	Constructing solution	Schoenfeld (2013)
	Interpreting solution	Ponte (2007)
	Rechecking and reflecting	Jonassen (2004)

Findings and Discussion

Appropriateness of Model Data

The data is said to be appropriate for the model if only 5% of the standard residual value is above +2 or below -2, and approximately 1% above +3 or below -3 (Linacre, 2018). In this research, from 5,452 data gathered from the students' mathematical process scores, 234 (4.29%) had the standard residual value above +2 and below -2, while only 15 (0.20%) recorded the value above +3 and below -3. Therefore, the recorded values indicated that the analysis done could validate the model.

Reliability

The first analysis focused on the reliability. The reliability constructed from the Rasch analysis is the separation reliability which indicates how the elements in each facet could be separated so each facet can be well defined (Myford & Wolfe, 2003). The reliability value is the ratio of true-score variance to test-score variance. This Rasch reliability value is in between 0 to 1. High value indicates that there is a high separation level between the elements in a facet (Wright & Masters, 1982).

The Alpha Cronbach value in this research for students was 0.94, raters (0.99), and mathematical process criteria (0.99). Table 3 below shows the reliability values for the students, raters, and the mathematical process criteria, that was between 0.94 to 0.99 with the separation index between 3.94 and 11.00. This value complements the Rasch measurement model. Therefore, the value of the respondents' reliability was ≥ 0.8 and the separation index value was ≥ 2 , and they were well accepted (Bond & Fox, 2015; Linacre, 2018).

Table 3. Reliability and Separation Index for the ProM3 Instrument's Facets

Facet	Reliability value	Separation Index
Students	0.94	3.94
Raters	0.98	11.00
Mathematical Process Criteria	0.99	8.74

Validity

The second analysis is pertaining the validity of each facet based on the Infit MNSQ and Outfit MNSQ values in the range of 0.6 until 1.4. Infit (information-weighted mean square residuals) is more sensitive towards the assessment that is expected to happen on students as compared to outfit (outfit sensitive fit statistics) (Linacre, 2019). The range of Infit MNSQ statistics in the previous studies as recorded were 0.4 until 1.2 (Wright & Linacre, 1989), 0.5 until 2.0 (Myford & Mislevy, 1995) and 0.75 until 1.3 (McNamara, 1996).

The Infit MNSQ and Outfit MNSQ statistics with the range of 0.6 until 1.4 was used for the rating scale in Rasch analysis, thus be used in this research (Linacre, 2019). As for the Outfit MNSQ statistics, it is similar to Infit MNSQ, but more sensitive towards the outliers. The range accepted to identify the non-fitting Outfit MNSQ is similar to the Infit MNSQ; between 0.6 until 1.4. Both Infit MNSQ and Outfit MNSQ statistics were checked to identify the data that did not fit with MRFM. The validity of each facet will be discussed below.

(i) Students Facet

Table 4 shows the students' measurement report, including students' measurement values, RMSE values, standard error, separation index, and chi-square values. The students' measurement values were in between -3.45 logit (standard error = 0.38, students 138), which was the lowest measurement, and 4.71 logit (standard error = 0.62, students 141), which was the highest measurement. The RMSE value for the students' facet was 0.31. The separation index was 3.94 which indicated that the students were able to be separated according to their abilities. This was further verified by the significant value of chi-square; $\chi^2 = 2666.1$, $p < 0.01$, $df=187$. The significant separation index and chi-square values also indicated that the measurement recorded for the students could be used to differentiate them according to their abilities (Linacre, 2019).

The validity refers to the Infit MNSQ and Outfit MNSQ statistics which was to identify the data that fit the measurement model. The findings revealed that 161 (85.64%) from the original population of students taking part in this research recorded the Infit MNSQ value

between 0.60 – 1.40 and Outfit MNSQ (0.60 – 1.40), which are in the acceptable range. This means that the data of students' facet fulfils the expected measurement model, thus has validity.

Table 4. Students' Ability

Students	Measurement Value	Standard Error (SE)	Infit MNSQ	Outfit MNSQ
141	4.71	0.62	1.16	1.22
10	1.39	0.31	1.40	1.35
46	1.53	0.28	1.23	1.40
71	-0.42	0.26	0.60	0.60
138	-3.48	0.38	1.39	1.44

Note. Chi-square value: 2666.1; df: 187; significant (probability): 0.00 Value RMSE: 0.31; Separation Index: 3.94

(ii) Rater Facet

Table 5 shows the statistical measurement for raters' facet which consists of severity, standard error, and the values of Infit and Outfit MNSQ. Raters' severity refers to their tendency either to score their students leniently or strictly (Eckes, 2015). In this research, the raters' severity was in the range of -0.59 logit (SE=006) for Rater B, who was the most lenient, until 0.74 logit (SE=0.06) for Rater D, who was the strictest. The stated standard error indicates the accuracy of each strictness measurement (Linacre, 2018).

The difference in strictness based on the chi-square test shows that there was a significant difference ($\chi^2 = 386.9$, $df = 6$, $p < 0.01$) between the raters. This indicates that the raters possess different severity in scoring. However, from the 1170 opportunities of agreement between the raters, 88 agreements were achieved that is 51.8% as compared to 73.5 (43.2%) expected agreements.

The values of Infit and Outfit MNSQ for each rater scoring the students' progress depicts their consistency in assessing. The Infit MNSQ values for each rater lies in the acceptable range, which is between 0.87 and 1.05, and 0.89 until 1.11 for the Outfit MNSQ. The readings are in the acceptable range which means that the raters assessed each student consistently, thus their assessments' scores have validity (Eckes, 2015).

Table 5. Raters' Measurement Facet

Rater	Severity Measurement	Standard Error (SE)	Infit MNSQ	Outfit MNSQ
D	0.74	0.06	1.02	1.02
G	0.26	0.06	0.95	0.98
C	0.24	0.06	0.98	0.99
E	-0.03	0.06	1.05	1.04
A	-0.06	0.06	1.01	1.01
F	-0.57	0.06	0.87	0.89
B	-0.59	0.06	1.05	1.11

Note. Chi-square value 386.9; df: 6; Significant (Probability): 0.00

Inter-Rater agreement opportunities: 170 Exact agreements: 88 = 51.80% Expected: 73.5 = 43.2%

(iii) Mathematical Process Criteria Facet

Based on Table 6, quantitative reasoning was the hardest criteria (1.45 logit, SE=0.12), while problem identification was the easiest criteria (-2.38 logit, SE=0.13). The chi-square test value was 2065.6, df was 28, $p < 0.01$, which indicates that there was a significant difference in difficulties of the assessed criteria (Linacre, 2018). This strengthens the claim that there is a significant variation of difficulties among the criteria. Figure 1 shows the positions of each criterion according to the difficulty level. The criteria on the most top level is the hardest, which is the quantitative reasoning, while the one at the lowest level is the easiest, which is problem identification.

The accuracy of the measurement can be seen based on the standard error (SE) recorded in Table 6. The standard error for all criteria were between 0.11 and 0.14. Then, the value of MNSQ was analysed to determine if the responses for each criterion fits the expected measurement of Rasch model. Table 6 shows the range of Infit MNSQ for all criteria (between 0.60 – 1.38) whereas the Outfit MNSQ lies between 0.63 – 1.36. The values are in the acceptable range of 0.6 – 1.4 (Linacre, 2019). Thus, it can be said that all the criteria have appropriate values and complements the Rasch measurement model, hence has validity.

Table 6. Measurement Facet of Mathematical Process Criteria

Criteria	Measurement	Standard Error (SE)	Infit MNSQ	Outfit MNSQ
Quantitative Reasoning	1.45	0.12	1.37	1.31
Relation with subject	0.20	0.12	1.38	1.36
Identifying trend	1.33	0.14	1.17	1.14
Abstract reasoning	0.74	0.11	0.60	0.63
Identifying problem	-2.38	0.13	0.63	0.76

Note. Chi-square value: 2065.6; df: 28; Significant (probability): 0.00

Facets program sequences the level of students' ability, raters' strictness, and criteria's difficulty from assessment scale to logit scale, and prepares a single reference to interpret the findings. In the Wright map, all the measurements for students, raters, criteria and scales' categories were vertically positioned on the same dimension, with logit as the measurement unit (Eckes, 2015).

Figure 1 shows the Wright map, which depicts the positioning of each facet involved. The first column represents the measurement scale in logit, followed by the second column of student' ability parameter, raters' severity in the third column, level of criteria's difficulty in column four, and the scale of rubric's score in the fifth column.

Each star in the first row represents two students, while each dot represents one student. The students' ability was arranged according to the positive orientation where the students with higher ability positioned at the top, while the lower ability students were placed at the bottom part of the map. Contrastingly, the mapping of raters' severity and criteria's difficulty were arranged with the strictest rater and hardest criteria being on top, whereas the most lenient rater and easiest criteria were positioned at the bottom.

As observed, the variation in the measurement of raters' severity was in the acceptable range with the logit distribution of 1.33, which is 16.2% of the observed logit distribution of students' ability (8.9 logits). This indicates that although there was a difference in the raters' severity level, the use of proM3 rubric as a scoring manual reduced clashes among raters in rating the students' abilities. This situation normally happens for the expert

raters and although there is a practice to prioritize mutual agreement with other raters, but they can be seen to be working independently (Linacre, 2018).

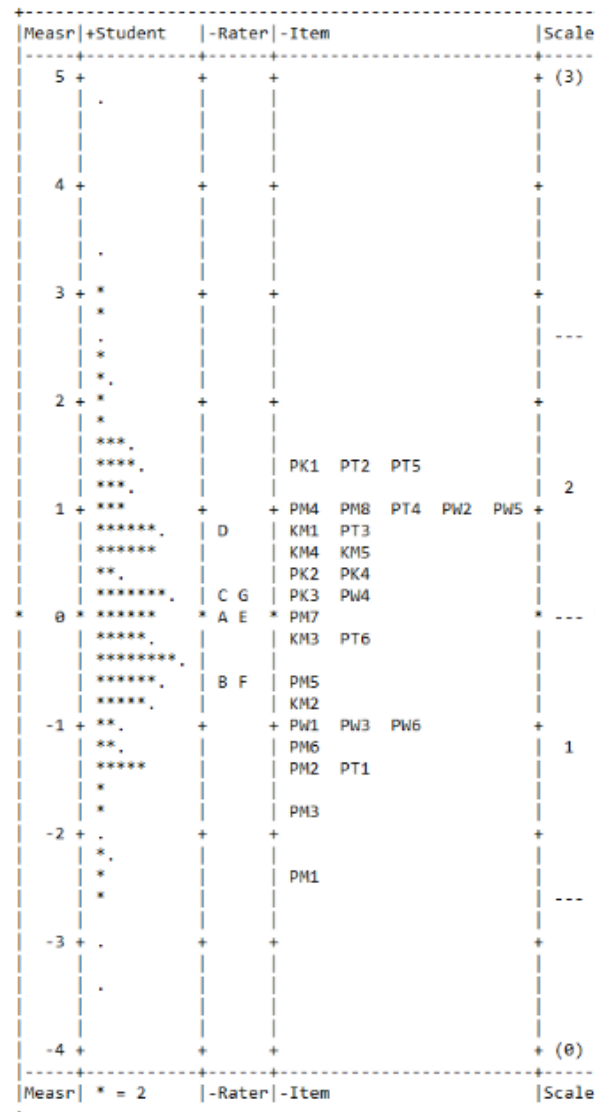


Figure 1. All Facet Vertical Ruler

Implication and Suggestion

This research shows that the analysis based on MRFM connects three variables including students' mathematical processing ability, the severity of raters, and the difficulty of criteria in ProM3 rubric. By collecting the information, the data can be researched holistically for each individual, which is difficult to be done in a traditional test analysis. Therefore, Rasch measurement can provide sufficient beneficial information to develop an assessment rubric to measure students' mathematical process. Besides, by using the Rasch measurement model, the expected reliability of students' mathematical process can be elevated by identifying the elements which affect the scores.

Rasch analysis allows to study the raters, criteria, and the combination of raters over criteria and specific students' group that affects the reliability value (Eckes, 2015). Thus, further analysis on raters effect such as their severity effect, halo effect and the bias effect should be carried out to gather more information pertaining the validity of the developed rubric.

Moreover, this research only focuses on one problem solving and reflective writing task, and involves only Form 1 students from boarding schools in northern and middle zones. The future research should focus on the validity and reliability of ProM3 Rubric in other tasks or assignments or different groups of students.

Conclusion

The new curriculum of Mathematics focuses on the students' mathematical process and their skills in solving problems to encourage them develop their understanding and concept independently. The process of problem solving gives an opportunity for the students their full potential to experience cognitive process of mathematics by making connection, reasoning, representation, mathematical communication, and problem solving. A well planned and structured activity of problem solving include situation and a valid rubric that allows students' mathematical process to be measured, and the findings of the measurement can be used to aid students' learning. MRFM analysis can be used to analyse the quality of items holistically not only pertaining the mathematical processing ability of students and the criteria's difficulty, but also the raters' severity to further verify the developed rubric. The findings of Facets analysis recorded high reliability index for students, raters and criteria, the hierarchical positioning of items according the difficulty level and the variation of raters' severity measurement in a minimum range. This indicates the ProM3 Rubric can be used to measure students' mathematical process, and it is valid and reliable.

References

- Adnan, M., & Jalil, N. S. (2016). Keupayaan Murid Cemerlang Akademik Tingkatan Empat dalam Menyelesaikan Masalah Algebra Bukan Rutin. *Jurnal Pendidikan Sains & Matematik Malaysia*, 6(1), 58–67.
- Akdemir, E. (2018). Investigating the Reflective Thinking Skills of Students for Problem Solving. *Online Submission*, 1(November), 774–780. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED595226&site=ehost-live>
- Arter, J., & McTighe, J. (2001). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance* (T. R. Guskey & R. J. Marzano, eds.). California: Corwin Press.
- Ayllon, M., Gomez, I., & Ballesta-Claver, J. (2016). Mathematical Thinking and Creativity through Mathematical Problem Posing and Solving. *Propositos y Representaciones*, 4(1), 169–218. Retrieved from <https://eric.ed.gov/?id=EJ1126306><http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1126306&site=ehost-live>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Third). New York: Routledge Taylor and Francis Group.
- Boud, D., Keogh, R., & Walker, D. (1985). Reflection: Turning Experience into Learning. In *Learning*. <https://doi.org/10.1145/1734263.1734380>
- Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. New York: Springer Science + Bussiness Media.
- Brookhart, S. M. (2010). General Principles for Assessing Higher-Order Thinking. In *How to Assess Higher-Order Thinking Skills in Your Classroom* (pp. 17–38). Retrieved from http://www.ascd.org/publications/books/109111/chapters/General_Principles_for_Assessing_Higher-Order_Thinking.aspx
- Bruner, J. (1960). *The Process of Education*. New York: Vintage.

- Carson, J. (2007). A Problem With Problem Solving : Teaching Thinking Without Teaching Knowledge. *The Mathematics Educator*, 17(2), 7–14.
- Costa, A. L. (2004). Habits of Mind. *Current Surgery*, 61(1), 101–103. <https://doi.org/10.1016/j.cursur.2003.09.009>
- Costa, A. L., & Kallick, B. (2008). *Learning and Leading with Habits of Mind: 16 Essential Characteristics for Success*. Alexandria: Association for Supervision and Curriculum Development.
- Dollah, M. U., Saad, N. S., Abdullah, M. F. N., & Yusof, Q. (2016). Penerapan Nilai Rasionalisme dan Objektivisme dalam Pengajaran Guru Matematik Sekolah Rendah. *Jurnal Pendidikan Sains & Matematik Malaysia*, 6(2), 85–108.
- Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2018). Relational Reasoning and Its Manifestations in the Educational Context : a Systematic Review of the Literature. *Educational Psychologist Review*, 25(3), 391–427.
- Eckes, T. (2015). *Introduction to Many-Facet Rasch Measurement : Analyzing and Evaluating Rater Mediated Assessment* (2nd ed.). <https://doi.org/10.3726/978-3-653-04844-5>
- Edelen, D., & Bush, S. B. (2020). A Cross-Cultural Approach for Communication in the Mathematics Classroom and Beyond. *Southeastern Regional Association of Teacher Educators*, 29(2). Retrieved from <https://eric.ed.gov/?id=EJ1268308>
- Fuson, K. C., Kalchman, M., & Bransford, J. D. (2005). Mathematical Understanding: An Introduction. In *How Students Learn: History, Mathematics, and Science in the Classroom* (pp. 215–256). <https://doi.org/10.17226/10126>
- Ghazi, S. R. (2014). *Formal Operational Stage of Piaget' s Cognitive Development Theory : An Implication in Learning Mathematics*. 17(2), 71–85.
- Jonassen, D. H. (2004). *Learning to Solve Problems ; An Instructional Design Guide*. San Francisco: Pfeiffer.
- Kadir, J., & Parman, M. S. (2013). Mathematical Communication Skills of Junior Secondary School Students in Coastal Area. *Jurnal Teknologi*, 2, 77–83.
- Klerlein, J., & Hervey, S. (2000). *Mathematics as a Complex Problem-Solving Activity* : Retrieved from <https://www.generationready.com/mathematics-as-a-complex-problem-solving-activity/>
- Kolb, A. Y., & Kolb, D. A. (2012). Experiential Learning Theory. In *Encyclopedia of the Sciences of Learning*. https://doi.org/10.1007/978-1-4419-1428-6_227
- Krawec, J. L. (2010). *Problem Representation and Mathematical Problem Solving of Students of Varying Math Ability*. University of Miami.
- Lantz, H. B. (2004). *Rubrics for Assessing Student Achievement in Science Grades K-12*. California: Corwin Press.
- Li, Y., & Schoenfeld, A. H. (2019). Problematizing Teaching and Learning Mathematics as Given in STEM Education. *International Journal of STEM Education*, 6(1). <https://doi.org/10.1186/s40594-019-0197-9>
- Linacre, J. M. (2018). *A User's Guide to FACETS Rasch Model Computer Programs*.
- Linacre, J. M. (2019). *Winsteps Help for Rasch Analysis 4.4.0*.
- Lunz, M. E. (1997). Performance Examinations: Technology for Analysis and Standard Setting. *Annual Meeting of the National Council of Measurement in Education*. <https://doi.org/10.1158/1535-7163.mct-16-0142>
- Malaysian Education Ministry. (2016). Standard Curriculum for Secondary School : Mathematics Form One. In *Malaysian Education Ministry*. Putrajaya.
- Mcleod, S. (2017). Kolb-Learning Styles. Retrieved from www.simplypsychology.org/learning-

kolb.html

- McLeod, S. (2018). *Bruner's Three Modes of Representation*. 1–5.
- McNamara, T. F. (1996). *Measuring Second Language Performance*. London: Longman.
- Moskal, B. M. (2003). Recommendations for Developing Classroom Performance Assessments and Scoring Rubrics. *Practical Assessment, Research & Evaluation*, 8(14), 1–5.
- Myford, C. M., & Mislevy, R. J. (1995). *Monitoring and Improving a Portfolio Assessment System*. Princeton: Center for Performance Assessment Educational Testing Service.
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement : Part I. *Journal of Applied Measurement*, 4(4), 386–422.
- Noor, F. M., Zamri, S. N. A & Eu, L. K. (2016). Penaakulan Perkadaran Murid Tahun Lima dalam Topik Nisbah dan Kadaran. *Jurnal Pendidikan Malaysia*, 41(2), 99–105.
- OECD. (2018). *Pisa 2021 Mathematics Framework (Draft)*. Retrieved from <https://pisa2021-maths.oecd.org/ca/index.html>
- Piaget, J. (2008). Intellectual Evolution from Adolescence to Adulthood 1. *Human Development*, 14, 40–47. <https://doi.org/10.1159/000112531>
- Ponte, J. P. (2007). Investigations and Explorations in The Mathematics Classroom. *ZDM - International Journal on Mathematics Education*, 39(5–6), 419–430. <https://doi.org/10.1007/s11858-007-0054-z>
- Rasiman, Prasetyowati, D., & Kartinah. (2020). Development of Learning Videos for Junior High School Math Subject to enhance Mathematical Reasoning. *International Journal of Education and Practice*, 8(1), 18–25. <https://doi.org/10.18488/journal.61.2020.81.18.25>
- Rothstein, A., Rothstein, E., & Lauber, G. (2007). *Write for Mathematics* (2nd ed.). California, USA: Corwin Press.
- Schoenfeld, A. H. (2013). Reflections on Problem Solving Theory and Practice. *The Mathematics Enthusiast*, 10(1).
- Scusa, T. (2008). *Five Processes of Mathematical Thinking* (University of Nebraska-Lincoln). Retrieved from <http://digitalcommons.unl.edu/mathmidsummative/38>
- Stergar, C. (2005). *Performance Task, Checklist and Rubrics*. Illinois: Pearson Education.
- Sujadi, I., & Masamah, U. (2017). *Mathematical Reflective Thinking Processes of Senior High School Students*. 25, 115–126.
- Sundstrom, T. (2014). *Mathematical Reasoning : Writing and Proof* (1.1). California, USA: Creative Commons.
- Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Cognitive Components of a Mathematical Processing Network in 9-year-old Children. *Developmental Science*, 17(4), 506–524. <https://doi.org/10.1111/desc.12144>
- Usta, N. (2020). Evaluation of Preservice Teachers' Skills in Solving Non-Routine Mathematical Problems through Various Strategies. *Asian Journal of Education and Training*, 6(3), 362–383. <https://doi.org/10.20448/journal.522.2020.63.362.383>
- Wen, T. K., Tajudin, N. M., & Masri, R. (2017). Pembinaan Item dalam Bidang Perkaitan bagi Mata Pelajaran Matematik Tingkatan Empat Berdasarkan Model Taksonomi Pemprosesan Maklumat. *Jurnal Pendidikan Sains & Matematik Malaysia*, 7(2), 1–18.
- Wood, T. (2001). Teaching Differently : Creating Opportunities for Learning Mathematics. *Theory Into Practice*, 40(2), 110–117. <https://doi.org/10.1207/s15430421tip4002>
- Wright, B. D., & Linacre, J. M. (1989). Observations Are Always Ordinal; Measurements, However, Must Be Interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857–860.
- Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: MESA Press Chicago.

Yee, C. S., Tze, W. J., & Abdullah, A. H. (2017). Pencapaian Matematik TIMSS 1999 , 2003 , 2007 , 2011 dan 2015 : Di Mana Kedudukan Malaysia Dalam Kalangan Negara Asia Tenggara ? *Malaysian Journal of Higher Order Thinking Skills In Education*, (December).