

# Rasch Analysis Model: Reliability and Validity of Superitem Test Instrument

Mohd Faizul Ridzuan<sup>1</sup>, Lim Hooi Lian<sup>2</sup>, Farah Atiqah Ahmad Fozee<sup>3</sup>, Siti Nur Anissa Mohd Nasser<sup>4</sup>

<sup>1</sup>School of Educational Studies, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia,

<sup>2</sup>School of Educational Studies, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia,

<sup>3</sup>Unit of General Studies, Kolej Komuniti Bagan Serai, No. 29, Jalan Syed Thaupy 2, Pusat Bandar Baru 34300 Bagan Serai, Perak, Malaysia, <sup>4</sup>School of Educational Studies, Universiti Sains Malaysia 11800 Pulau Pinang, Malaysia

Email: mohdfaizul\_ridzuan@yahoo.com.my, hllim@usm.my, farah@kkbse.edu.my, sitinurannisa177@gmail.com

**To Link this Article:** <http://dx.doi.org/10.6007/IJARPED/v9-i4/8166>

DOI:10.6007/IJARPED/v9-i4/8166

**Published Online:** 29 December 2020

## Abstract

This study is conducted to measure the reliability and validity of superitem test based on Rasch Analysis Model. Superitem test instrument is built based from SOLO Model four hierarchical levels namely unistructural, multistructural, relational and extended abstract. There are four content domains of linear equations tested in superitem test which are linear pattern (pictorial), direct variation, concept of function and arithmetic sequence. Every correct response made by the students involved are references to any stages in SOLO Model which also shows their algebraic solving ability. This research is conducted among community college in Perak state that involved 390 respondents. The results demonstrated the individual reliability is high at 0.90 with separation index of 2.98 whereas the item reliability is also high at 0.99 with separation index of 9.41. Besides that, superitem test has fulfilled the validity aspect construction in Rasch Analysis Model in terms of item fit, polarity item and unidimensionality. In conclusion, superitem test instrument is valid and able to measure algebraic solving ability in community college students.

**Keywords:** Realibility, SOLO Model, Superitem Test, Validity.

## Introduction

Algebra is a process represents abstract reasoning which is the key to success in mathematics (Tan, 2015). However, not to some students (Kaput, 2008). There are several studies conducted have shown Malaysian students troubled in grasping algebraic concept (Ting, Rohani, Kamariah, & Dalia, 2017). Studies conducted by Nor Hasnida Che Ghazali and Effandi Zakaria (2011) onto 132 form two students showed 54.5% students got marks between 2.0 – 4.2 from overall total of 8 marks. These students are categorized as having low comprehension concept.

In addition, types of algebraic questions use in class nowadays do not meet the current needs which teachers should emphasize the use of real-life applications of algebra (Stacey & Macgregor, 1999). Some algebraic problem solving questions use in teaching and learning (T&L) at community college are as in Figure 1.

*Sehelai kain yang berbentuk segi empat tepat mempunyai panjang  $2xy$  meter dan lebar  $4y$  meter. Jika luas sebanyak  $3pq$  meter persegi dipotong daripada kain itu, kirakan baki kain itu dalam meter persegi.*

**Fig. 1** An Example of Community College Algebraic Solving Question

According to Stacey and Macgregor (1999) types of algebraic solving questions in Figure 1 is easy because it can be solved by routine procedure. They discussed that students need to be exposed with complex algebraic problem solving questions which requires them to apply the abstract concept of algebra to solve problems. This is to ensure the students would be able to appreciate the usage of algebra as a current problem solving method.

### Theoretical Background

Assessment is a process in teaching and learning that allow students and teachers to evaluate their teaching and learning process thus making improvement where needed (Shepard, 2005). A good assessment will enhance the student's understanding (Tan, 2015). According to Nor Hasnida Che Ghazali and Effandi Zakaria (2011) school assessment nowadays are prone towards measuring the student's achievement through examination. Hence, there are limited information to improve the teaching and learning process. A more specific measurement is needed to measure the students' weakness and strength (Lim & Noraini Idris, 2006).

Therefore, a new approach has been used in this study which presented superitem test and adapted it with different stages in SOLO (Structure of the Observed Learning Outcome) Model. SOLO Model is use to evaluate the students' understanding (Pegg & Tall, 2005). There are five stages in this model which are prestructural, unistructural, multistructural, relational and extended abstract. The higher the stage, the more understanding required to solve a problem (Chan, Tsui, Chan, & Hong, 2002; Biber & Incikabi, 2016).

In this study, superitem test has been used to evaluate the ability of algebraic solving in community college students according to four stages of SOLO Model which are unistructural, multistructural, relational and extended abstract. At unistructural and multistructural stage, students are able to use two or more information to provide responses. At relational and extended abstract stage, students are able to link and integrate the information in different situations. Each correct response given by students demonstrated their algebraic solving ability based on the SOLO Model stages.

### Superitem Test Development

Superitem test is built according to instrument development model proposed by Miller, Loveler and McIntire (2013). Superitem test involves four content domains of linear equations which are linear pattern (pictorial), direct variation, concept of function and arithmetic sequence in which two linear equations problems are developed for each content domain. Each problem developed in this superitem test is based on mathematics course

syllabus of certificate level in community college (Jabatan Pengajian Kolej Komuniti, 2017). Superitem test problems development is based on test specification table to ensure the questions tested are within the scope and relevant with the content.

#### a) Test Specification Table

Test specification table contains content domain and SOLO Model stages. Content domain involves four main linear equations domains namely linear pattern (pictorial), direct variation, concept of function and arithmetic sequence. Based on Kubiszyn and Borich (2003) the algebraic solving ability for each domain can be measured through pattern analysing, making pattern generalization and applying linear equations. Each question in superitem test is divided into four stages which are unistructural, multistructural, relational and extended abstract from incompetence level to expertise level based on the response structure characteristics of SOLO Model (Biggs & Collis, 1982). The purpose of this division is to facilitate the process assessment of community college students' algebraic solving ability.

From Table 1, there are two questions for each content domain of linear equations which are two questions of linear pattern (pictorial), direct variation, concept of function and arithmetic sequence. Thus, eight questions were assessed in this superitem test. Each question is divided into four sections which are analysing pattern, making algebra pattern generalization, making linear equation pattern generalization and applying linear equations. Findings from this superitem test are adapted to SOLO Model responses structure stages which are unistructural, multistructural, relational and extended abstract. Hence, the algebraic solving ability of a student is identified based on their response structure.

Table 1.

*Test Specification Table*

Content Domain	Number	SOLO Model Stages				Total
		Unistructural	Multistructural	Relational	Extended abstract	
Linear Pattern (Pictorial)	1	1	1	1	1	4
	2	1	1	1	1	4
Direct Variation	3	1	1	1	1	4
	4	1	1	1	1	4
Concept of function	5	1	1	1	1	4
	6	1	1	1	1	4
Arithmetic Sequence	7	1	1	1	1	4
	8	1	1	1	1	4
Overall Total		8	8	8	8	32

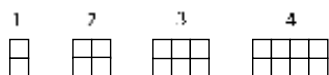
Table 2 below shows the questions related to linear equations for linear pattern (pictorial) domain use in this superitem test. At unistructural stage, students are required to use their algebra understanding to investigate the linear pattern of the problem and provide an exact unknown whereas at multistructural stage, students need to provide two or more exact unknowns. Next, at relational stage, students are required to make a generalization of linear pattern problem with developing linear equations. Lastly, at extended abstract stage, students are required to link the linear pattern concept and apply it in a more abstract

problem. Every exact response given by students portray their cognitive ability to solve problem that involved linear pattern (pictorial) based on the SOLO Model stages.

Table 2.

*Questions of Linear Pattern (Pictorial)*

*Rajah di bawah menunjukkan pola segiempat sama.*



Stage	Question	Explanation
Unistructural	<i>Berapakah jumlah segiempat sama pada langkah yang kelima?</i>	<ul style="list-style-type: none"> <li>Students are able to provide one exact unknown with information provided.</li> </ul>
Multistructural	<i>Berapakah jumlah segiempat sama pada langkah yang kelapan, langkah yang kesepuluh dan langkah m?</i>	<ul style="list-style-type: none"> <li>Students are able to explain linear pattern based on the problem given.</li> <li>However, students are not able to explain the connection in the linear pattern given.</li> </ul>
Relational	<i>Tuliskan persamaan bagi mencari jumlah segiempat sama sekiranya diberikan nilai langkah yang tertentu. Anggapkan T mewakili jumlah segiempat sama dan m sebagai nilai langkah yang tertentu.</i>	<ul style="list-style-type: none"> <li>Students are able to make a generalization for linear pattern in the problem given.</li> <li>Students are able to form linear equations based on the information provided.</li> </ul>
Extended Abstract	<i>Siti Fatimah bercadang untuk membentuk sebuah kubus dengan menggunakan segiempat sama yang terdapat pada langkah yang kedua. Berapakah jumlah segiempat sama yang diperlukan oleh beliau untuk membentuk kubus tersebut? Jelaskan.</i>	<ul style="list-style-type: none"> <li>Students are able to make connection on linear pattern concept and applying it in other situational problems.</li> <li>Students are able to get a new solving method based on a new situational problem.</li> </ul>

**Scoring Procedure**

Rating Scale Analyses is used to scoring question in superitem test. Through Rating Scale Analyses, students are given the same choice to answer a question (Wright, 1998). Arrangement values of 1, 2, 3 and 4 are used for scoring question in superitem test where

value 1 is referred to unistructural stage, 2 referred to multistructural stage, 3 referred to relational stage and 4 referred to extended abstract stage as shown in Table 3. Value 1 is the lowest level of responses while value 4 is the highest level of responses for each question in superitem test.

Table 3.

*Rating Scale Analyses of Superitem test*

1	2	3	4
Unistructural	Multistructural	Relational	Extended Abstract

**Rasch Analysis Model**

Rasch Analysis Model is able to measure latent traits (Abdul Aziz, Masodi, & Zaharim, 2013). According to Bond and Fox (2015) the probability of an individual to answer a correct response of item depends on the individual ability and the item difficulty level. The higher individual ability has the highest probability to answer all items whereas less difficult items to have a high probability to be getting correct response from all respondents when the differences between item difficulty and individual ability are monitored (Wright & Stone, 1979; Bond & Fox, 2015).

From Rasch Analysis Model, the reliability of one instrument can be measured through individual reliability and item and individual separation index and item. Meanwhile, the validity of an instrument can be measured through individual fit and item, polarity item and unidimensionality. Thus, Rasch Analysis Model is used in this study to determine the reliability and validity of superitem test instrument.

**Research Objectives**

Objectives of this study are as followed:

1. To identify the reliability of superitem test instrument using Rasch Analysis Model.
2. To determine the validity of superitem instrument using Rasch Analysis Model.

**Research Questions**

Based on the research objectives, there are two research questions:

1. What is the reliability of superitem test instrument in terms of individual reliability and item and individual separation index and item?
2. What is the validity of superitem test instrument in terms of individual fit and item, polarity item and unidimensionality?

**Methodology**

The survey method is used by using superitem test to collect data of the community college students' ability on algebraic solution. This research is conducted in Perak community college which involved 390 respondents. These respondents are students who are taking mathematics subject (SSM, 1022). Five experts are chosen to determine the validity of this superitem test content. They are lecturers from universities and community colleges that have vast knowledge in psychometrics and mathematics. Besides that, they also have more than five years teaching experience in psychometrics and mathematics.

### Data Analysis

Data collected from superitem test are inserted into WINSTEPS 3.71.0.1 software to be analysed. Rating scale is used in this study where value 1, 2, 3 and 4 are referred to the SOLO Model stages which are unistructural, multistructural, relational and extended abstract. Based on Bond and Fox (2015); Linacre (2005) statistics is used to measure the Rasch Analysis Model reliability are individual reliability and item and individual separation index and item as shown in Table 4.

Reliability of one instrument is referred to the consistency of measurement values acquired (Kamis, Ab. Rahim Bakar, Hamzah, & Asmiran, 2012). Individual realibility coefficient is referred to the consistency of respondent pattern when given other instrument set which measure the same construct while item reliability coefficient is referred to degree of consistency item arrangement on logit scale if the items are answered by different respondent group which acquire the same ability.

Table 4.

*Realibility of Rasch Analysis Model*

Statistic	Index	Statement
Individual Reliability and Item	< 0.6	Unaccepted
	0.6 – 0.8	Less accepted
	> 0.8	Strongly accepted
Individual Separation Index and Item	> 2	Good and accepted

However, validity in Rasch Analysis Model can be measured through item fit statistic, polarity item and unidimensionality (Bond & Fox, 2015). Statistics of item fit have shown how the data acquired are fit with Rasch Analysis Model (Ariffin, 2008). Mean-square value (MNSQ) and mean standardization (ZSTD) is used to determine the collected data compatibility with the model. Infit and Outfit MNSQ range that can be accepted for likert scale is between 0.6 to 1.4 (Bond & Fox, 2015) whereas cut off value for Infit ZSTD SD is |2.00| (Bode & Wright, 1999). However, ZSTD value can be neglected if MNSQ value is accepted (Linacre, 2005).

Polarity item is ascribed to early detection of construct validity in order for all items to contribute positively to measure latent traits (Bond & Fox, 2015). Point-measure Correlation value (PTMEA CORR.) is used to measure polarity item. If the value for PTMEA CORR. for all item are positive, hence, they function parrallely in measuring the construct. However, if the value of PTMEA CORR. acquired are negative, hence, the item need to be reviewed whether it should be fixed or omitted. The PTMEA CORR. value obtained for all items are preferable to be positive and more than 0.3 (Nunnally & Bernstein, 1994; Bond & Fox, 2015). A good range value of PTMEA COOR. for each instrument is between 0.3 to 0.6 (Bond & Fox, 2015).

Unidimensionality is a measurement of one latent trait or dimension at a time (Bond & Fox, 2015). Principal Component Analysis of Residuals (PCAR) is used to measure the research instrument dimension. Rasch Analysis Model needs at least a 40% raw variance explained by measures and is acceptable if it reaches up to 60% (Azrilah Abdul Aziz et al., 2013). Likewise, unexplained variance in 1<sup>st</sup> contrast should be in range of 5% to 15% (Fisher, 2007). Eigenvalue that is more than 5 will be a threat to unidimensionality probability (Linacre, 2005).

### Findings

Table 5 shows the reliability of individual and item and separation index of individual and item for superitem test instrument. Based on Rasch Analysis Model, the individual reliability value obtained is 0.90 and reliability item value is 0.99. The individual separation index value is 2.98 and item separation index is 9.41. According to Bond and Fox (2015) and Linacre (2005) research findings shows superitem test instrument have a strong acceptance of individual and item reliability and also acceptable individual and item separation index.

Table 5.

#### *Reliability of Superitem Test Instrument*

Statistic	Value	Statement
Individual Reliability	0.90	Strongly accepted
Item Reliability	0.99	Strongly accepted
Individual Separation Index	2.98	Good and accepted
Item Separation Index	9.41	Good and accepted

On the other hand, Table 6 shows the item fit and polarity item for superitem test instrument. Rasch Analysis Model have demonstrated Infit and Outfit MNSQ value for each item respectively between 0.7 to 1.4 and fulfilled the range of 0.6 to 1.4 (Bond & Fox, 2015). Value of PTMEA CORR. for all item are positive in range of 0.76 to 0.86 and have surpassed the minimum requirement of 0.3 (Nunnally & Bernstein, 1994).

Table 6.

#### *The fitness and polarity of items for Superitem Test Instrument*

Item Number	Total Score	Measure	S.E	MNSQ		PTMEA	
				Infit	Outfit	Corr.	Exp.
1	1155	-1.81	0.11	1.07	1.08	0.82	0.79
2	956	0.46	0.11	1.23	1.19	0.80	0.79
3	1070	-0.84	0.11	0.76	0.79	0.76	0.79
4	1056	-0.68	0.11	0.72	0.73	0.77	0.79
5	925	0.82	0.11	0.92	0.91	0.80	0.79
6	839	1.85	0.11	0.81	0.86	0.76	0.78
7	956	0.46	0.11	0.96	0.95	0.86	0.79
8	1019	-0.26	0.11	1.41	1.41	0.84	0.79

Table 7 illustrates the results of Principal Component Analysis of Residuals (PCAR) for superitem test instrument. Based on Rasch Analysis Model, raw variance explained by measures is 64.9% which have surpassed the 40% minimum requirement of unidimensionality in Rasch Analysis Model (Azrillah Abdul Aziz et al., 2013). Meanwhile, unexplained variance in 1<sup>st</sup> contrast is 11.3% which is still in range of 5% to 15% (Fisher, 2007). Eigenvalue of 2.6 shows the non-existent of second dimension (Linacre, 2005).



Table 7.

*The results of Principal Component Analysis of Residuals (PCAR)*

		Empirical	Modeled
Total raw variance in observations	22.8	100.0%	100.0%
Raw variance explained by measures	14.8	64.9%	64.4%
Raw variance explained by persons	11.2	49.1%	48.7%
Raw variance explained by items	3.6	15.8%	15.7%
Raw unexplained variance (total)	8.0	35.1%	100.0%
Unexplained variance in 1st contrast	2.6	11.3%	32.2%

**Discussion**

The study findings from Rasch Analysis Model demonstrated that superitem test has a high value of reliability on individual and item respectively with 0.90 and 0.99. According to Bond and Fox (2015) reliability value that is more than 0.8 could be strongly accepted while the value that is less than 0.6 is not accepted. Meanwhile Fisher (2007) stated that reliability value that is more than 0.94 is excellent. Therefore, superitem test is a useful and reliable instrument to measure the algebraic solving ability of community college students.

Likewise, individual separation index and item value are 2.98 and 9.41 respectively. As stated by Linacre (2005) separation index value that is more than 2 is preferred and could be accepted. Research findings have shown that the community college students' ability could be divided into three strata: low, medium and high. Furthermore, item separation index showed item in superitem test could be categorized to 9 strata and enable it to measure the algebraic solving ability for each level of community college students.

Research findings also have highlighted all item have fulfilled the Infit and Outfit MNSQ range set by Bond and Fox (2015) which is between 0.6 to 1.4. This suggests that all items are consistent with item measurement. Besides that, Point-measure Correlation (PTMEA CORR.) value have shown all items are positive with range of 0.77 to 0.84. Therefore, PTMEA CORR. value for all item have surpassed the 0.3 minimum requirement (Nunnally & Bernstein, 1994). The positive PTMEA CORR. values illustrates that the items in superitem test are able to differentiate the ability of community college students.

Meanwhile, unidimensionality provides unexplained variance in 1<sup>st</sup> contrast is 11.3% which is still in range of 5% to 15% (Fisher, 2007). Moreover, raw variance explained by measure is 64.9% which have surpassed the 40% minimum requirement needs for unidimensionality in Rasch Analysis Model (Azrilah Abdul Aziz et al., 2013). The biggest factor taken from residual only have three strongest items and far from five items that are structured to create the second construct (Linacre, 2005). Therefore, items in superitem test instrument have fulfilled the unidimensionality aspects in which only algebraic solving ability in community college students are measured.

**Conclusion**

Superitem test instrument use in this study is built based on four levels of SOLO Model namely unistructural, multistructural, relational and extended abstract. There are four content domains of linear equations tested in superitem test which are linear pattern (pictorial), direct variation, concept of function and arithmetic sequence. In the early phase, students are tested with easy questions and at the end of the phase, they are tested with more challenging questions (Biggs & Collis, 1982; Lim & Idris, 2006). Hence, this superitem



test instrument which is built through the hierarchical level able to differentiate the students' cognitive ability level.

Besides that, the analysis based on Rasch Analysis Model indicates that superitem test is a reliable and valid instrument to be used to measure the community college students' algebraic solving ability. However, the reliability and validity of superitem test instrument needs to be further analysed if there is an interest from other researcher to use superitem test instrument towards different targeted groups.

This study could also help the lecturers at Community College to improve the teaching and learning (T&L) process aspects mainly in the assessment, diagnosis and intervention. Developing a high-quality test is strongly emphasized to produce meaningful results. Therefore, lecturers can use this superitem test instrument to gather responses or feedback from students regarding algebraic expressions topic and linear equation furthermore drafting the intervention plan suitable for the students.

### Acknowledgements

This study was funded by the Ministry of Higher Education Malaysia through the Federal Training Prize. I would like to express many thanks to the Department of Polytechnic Education and Community College for allowing this study to be conducted at community college in Perak state.

### References

- Abdul Aziz, A., Masodi, M. S., & Zaharim, A. (2013). *Asas model pengukuran Rasch: pembentukan skala dan struktur pengukuran*. Universiti Kebangsaan Malaysia.
- Biber, A. C., & Incikabi, L. (2016). Problems Posed by Prospective Elementary Mathematics Teachers in the Concept of Functions: An Analysis Based on SOLO Taxonomy. *Mersin University Journal of the Faculty of Education*, 12(3), 796–809.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. Academic Press.
- Bode, R. K., & Wright, B. D. (1999). Rasch measurement in higher education. In J. C. Smart & W. G. Tierney (Eds.), *Higher Education: Handbook of Theory and Research* (pp. 287–316). Springer Netherlands.
- Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) Taxonomy on Student's Learning Outcomes: An empirical study. *Assessment & Evaluation in Higher Education*, 27(6), 511–527.
- Fisher, J. W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Jabatan Pengajian Kolej Komuniti. (2017). *Matematik (SSM 1022)*. Putrajaya, Malaysia.
- Kaput, J. (2008). What is algebra? What is algebraic reasoning? In J. Kaput, D. Carraher, & M. Blanton (Eds.), *Algebra in the Early Grade* (pp. 5–17). Lawrence Erlbaum/Taylor & Francis Group & National Council of Teachers of Mathematics.
- Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice* (7th ed.). John Wiley & Sons Inc.
- Lim, H. L., & Idris, N. (2006). Assessing Algebraic Solving Ability of Form Four Students. *International Electronic Journal of Mathematics Education*, 1(1), 55–76.
- Linacre, J. M. (2005). *A user's guide to WINSTEPS: Rasch Model Computer Programs*. MESA

Press.

- Miller, L. A., Loveler, R. L., & McIntire, S. A. (2013). *Foundations of psychological testing: a practical approach* (4th ed.). SAGE Publications Inc.
- Ghazali, N. H. C., & Zakaria, E. (2011). Students' Procedural and Conceptual Understanding of Mathematics. *Australian Journal of Basic and Applied Sciences*, 5(7), 684–691.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychological theory*. MacGraw-Hill.
- Pegg, J., & Tall, D. (2005). The fundamental cycle of concept construction underlying various theoretical frameworks. *Analyses ZDM*, 37(6), 468–475.
- Shepard, L. A. (2005). Formative assessment: Caveat emptor. *ETS Invitational Conference the Future of Assessment: Shaping Teaching and Learning*. New York.
- Stacey, K., & Macgregor, M. (1999). Implications for mathematics education policy of research on algebra learning. *Australian Journal of Education*, 43(1), 58–71.
- Tan, Y. H. (2015). *The cognitive diagnostic assessment of the learning of algebraic expressions for form two students (Unpublished doctoral dissertations)*. Universiti Sains Malaysia.
- Wright, B. D. (1998). Rating scale model (RSM) or partial credit model (PCM). *Rasch Measurement Transactions*, 12(3), 641–642.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.