International Journal of Evaluation and Research in Education (IJERE)
Vol. 13, No. 2, April 2024, pp. 1088~1099
ISSN: 2252-8822, DOI: 10.11591/ijere.v13i2.25721

Rasch model assessment of algebraic word problem among year 8 Malaysian students

Ku Soh Ting[1], Othman Talib[1], Ahmad Fauzi Mohd Ayub[2], Maslina Zolkepli[3], Chen Chuei Yee[4], Teh Chin Hoong[5]
[1]Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia, Serdang, Malaysia
[2]Department of Foundations of Education, Faculty of Educational Studies, Universiti Putra Malaysia Serdang, Malaysia
[3]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia
[4]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang, Malaysia
[5]ASASIpintar Program, Pusat GENIUS@Pintar Negara, Universiti Kebangsaan Malaysia, Bangi Malaysia

Article Info

Article history:

Received Dec 23, 2022
Revised Sep 2, 2023
Accepted Sep 26, 2023

Keywords:

Algebra
Arithmetic
Critical thinking
Linguistic
Person-item distribution map algebra

ABSTRACT

Word problems continue to be a challenge for students today. All students must meet the prerequisites for problem solving and reasoning skills, which are important components of the critical thinking component of 21st century skills. This study is being conducted to assess students' strategies for solving word problems with numbers, consecutives, and ages. The Rasch model is used to analyze the item difficulty level of word problems and students' strategies for solving ten-word problems at various levels of item difficulty in a similar trait. Then, Pearson correlation analysis is used to investigate the item difficulty level in relation to linguistic, algebraic, and arithmetic factors of word problems before evaluating students' performance on solving these word problems using various strategies. Rasch model found these algebraic word-problem questions are slightly harder for year 8 Malaysian students in relative to an international standard. Meanwhile, the item difficulty of word problems is driven by linguistic and algebra factors where students can score accurately if the word problems contained explicit information. However, the students encountered difficulties while losing their solution strategy when the questions contained implicit data that demanded critical thinking ability.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ku Soh Ting
Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia
43400 UPM Serdang, Selangor, Malaysia
Email: ting_blue2003@yahoo.com

1. INTRODUCTION

In 2013, the Ministry of Education Malaysia introduced the Malaysian Education Blueprint 2013-2025, which established visions and aspirations for education systems as well as aspirations for Malaysian students to have 21st century skills in order to compete in the modern labor market. One of the Malaysian education system's system aspirations is to have a quality education system which capable of placing Malaysia in the top third of countries in international assessments such as the program for international student assessment (PISA) and the trends in international mathematics and science study (TIMSS) in 15 years. Furthermore, Malaysian students are aspired to be globally competitive, with knowledge, critical thinking skills, leadership abilities, and bilingual proficiency [1].

According to recent PISA 2018 reports, Malaysian students scored lower than the average countries of the Organization for Economic Cooperation and Development (OECD) in reading, mathematics, and science, despite the fact that Malaysia achieved relatively higher mean mathematics and science performance than PISA 2012 [2]. Malaysian students' underperformance in PISA 2018 is attributed to curriculum, teaching quality, teacher welfare, and parental involvement, all of which have a significant impact on student performance [3], [4]. According to previous studies [4], [5], students' thinking and answering processes are becoming increasingly diverse, and students are unable to perform in international assessments that required high order thinking skills (HOTS) responses. Similar findings in the TIMSS were reported by Tajudin and Chinnappan [6], indicating that most high school students in Malaysia continue to perform below expectations, particularly in cognitively demanding tasks.

Furthermore, Perera and Asadullah [7] discovered that the incorporation of creative and critical thinking skills in the curriculum is limited in Malaysia, causing Malaysian students to perform poorly in the PISA assessment when compared to Singapore and Korea. The critical thinking skill is an important element of the 21st century learning skills and this skill can be developed through problem solving which requires the ability to solve problem effectively by using knowledge, facts, and data. For example, in mathematic syllabus, there is mathematical word problem solving questions which educates the students to use their critical thinking skills for the solutions.

According to previous studies [8]–[11], mathematical word problem solving is considered a hurdle for the students due to the learners' negative attitude, mindsets and interests in mathematics. On the other hand, another studies found word problems to be boring, difficult, and lacking in enjoyment while learning [10], [12], after collecting this information through questionnaires and interviews with the participants. Besides that, there are numerous factors, including linguistics comprehension and numerical complexity that contributed to the difficulty of the word problem [13]–[16].

To assess students' ability to solve word problems, a test instrument will be developed in which item tests were adapted from respective textbooks or reference books that correspond to their Mathematics syllabus, with the assistance of content experts determining the item difficulty based on their predefined perspectives [11], [17]–[19]. Furthermore, the performance of students on algebraic word problem-solving skills was determined and analyzed based on mean scores of the test instrument [11], [17], [18] which were thought to have limitations in the information that could be extracted from these studies. This does not adequately explain or represent the algebraic word problem-solving abilities of Malaysian students.

As a result, this study evaluates a group of year 8 students in order to gain insight into the ability of algebraic word problem solving among Malaysian students. Rasch model analysis was used to categorize the item difficulty level of ten-word problems containing numbers, consecutives, and ages, which is then investigated from the perspectives of linguistic, algebraic, and arithmetic factors by using Pearson correlation analysis. Rasch analysis was also used to evaluate the students' ability to solve word problems of varying difficulty in a similar characteristic. Because students cannot be solved without going through critical stages of the complicated problem-solving process as summarized by Depaepe *et al*. [20], the item tests in this study were designed as more difficult typed-word problems. According to Ibrahim *et al*. [19], the difficulty level of word problems as reflected in students' word problem solving performance was investigated in this study, which included the steps of how students interpret transfer situations, the socially situated nature of classroom transfer processes, and how learning transfer is carried out. This is followed by additional research into the relationship between arithmetic and algebraic knowledge -based solution strategies used by Malaysian students when solving word problems of varying difficulty.

Thus, a variety of word problems with varying degrees of difficulty are evaluated as item-level difficulty of word problems in order to investigate the characteristics of word problems from the perspectives of linguistic, algebra, and arithmetic factors, and to assess students' capabilities to solve word problems through their arithmetic and algebra knowledge. Previous research on these relationships relied heavily on basic arithmetic word problems [21]–[25]. Therefore, procedural knowledge with arithmetic numbers is applied in this study as to represent the algebra skill in solving word problems. Furthermore, representation and fluency in interpreting the mapping of symbols and variables to word problems are required. The Rasch model allows for the determination of item characteristics such as item difficulty level for word problems based on the probability of a correct response at a given level of participant capability. The goal of this research is to answer the following research questions: Is there any relationship between the linguistic, algebraic, and arithmetic properties of the word problem with the item characteristics? Can the arithmetic and algebra skills of students be used to predict their ability to solve word problems?

2. RESEARCH METHOD

2.1. Participants and overall design

A total of 236 students who aged in between 13 and 14 years old from an international school in Malaysia were participated in this work. Before enrolling in this international school, these students completed their primary education at various Primary Chinese schools located throughout Malaysia's Klang Valley. There were 127 male students (53.8%) and 109 female students among the eighth-grade students from eight classes (46.2%) where the participants were consisted of 233 Chinese students (98.73%), an Indian student (0.42%), a Malay student (0.42%) and a Eurasian student (0.42%). Due to the sensitive ethical issue, all the name of the students will be converted into number regardless of their gender during data processing.

2.2. Instrument

The ten-word problems are created by adapting questions from international school mathematics textbooks and presented on paper and pencil. While the complexity of algebraic word problems varies, they are classified as algebraic because they usually require the use of letter-symbolic algebra to solve the word problems, including number problems, consecutive problems, and age problems. Thus, one unknown will be provided in the first category of word problems, where the unknown values can be derived from the known values by performing the arithmetic operations specified in the problem. Meanwhile, for the second and third categories of word problems, both necessitate the combination of multiple unknowns, which cannot be easily solved by performing the arithmetic operations as described in the problem directly from the known value. The difference between the third and second categories was that the third category typed questions required non-routine thinking to deduce the solution from a more difficult word problem. The word problems test was distributed to the students during the mathematics lesson in the classroom. Each group's students were evaluated individually for 45 minutes.

2.3. Measure

Simple letter-symbolic algebra and arithmetic methods are used to solve these word problems. Students were required to answer all questions and demonstrated all their working solutions. The accuracy score was determined by the number of word problems answered correctly. Meanwhile, arithmetic intrusions are used as a strategy use metric, referring to the percentage of attempted questions answered using non-algebraic methods. Each question's solution was coded exclusively as “algebraic”, “arithmetic” or “no strategy” based on the overall solution. Classification of the question's solution is not based on the initial problem representation but referred to the steps taken to arrive at the answer. For example, if a solution started with a model diagram followed by guess- and-check procedures, it is categorized as “arithmetic”. If the diagram is followed by an algebraic expression and the unknown solved via equivalent equations, the solution is considered as “algebraic”. In addition, if a solution started with some attempt at an algebraic expression (e.g., let the weight of the dog be X) but contained no equation or other symbolic notation, then it is labeled as “arithmetic”. The arithmetic and mix solutions were belonged to the non-algebraic methods. Meanwhile, “no strategy” is referred to the “solutions” with no discernible strategies. The results of the test are then utilized as data in this study. This set of algebraic word problems will be used to examine the students' algebraic and arithmetic knowledge.

2.4. Analysis

The analyses used in this study are divided into two phases. The first phase will be investigating the item characteristics and employing Rasch model to categorize these word problems based on their difficulty levels by using the Winsteps version 4.8.2.0. This is followed by determining the correlation of the item difficulty level with the linguistic, algebra and arithmetic factors. In the latter phase, students' preference for solving algebraic word problems either by adopting algebra or arithmetic methods, as well as students who failed to solve the word problems, are analyzed using the descriptive statistics as generated by IBM statistical package for the social sciences (SPSS) version 25.

3. RESULTS AND DISCUSSION

3.1. Item difficulty level of word problem

There was a total of 10 questions in this algebra and arithmetic based word problems' test with some of the questions have an additional one or two sub-questions which required students to show their calculations. The purpose of this word problem's test was to access students' understanding on solving algebraic and arithmetic based word problems. Students could use either algebra or arithmetic methods to solve these word problems. Software WINSTEPS version 4.8.2.0 is employed to perform the Rasch measurement on the dichotomous responses where students who can answer the word problems correctly will be denoted as (1). However, students who was failed to solve the math problems will be assigned as (0). In the Rasch measurement, Cronbach alpha value for the item and person reliability; and variable map regarding the student's ability on solving the word problems at various level of item difficulty in a similar scale is analyzed.

The person-item distribution map (PIDM) represented the distribution of item difficulty and students' ability to answer the algebra and arithmetic questions correctly on a same logit scale. The ability of students to respond correctly in the test is listed on the left side of the PIDM while the item difficulty is itemized on the right side of the map. Higher logits on the left map indicate students with higher ability to answer the test correctly while the most difficult item is represented by highest logits on the right map and vice versa Figure 1.

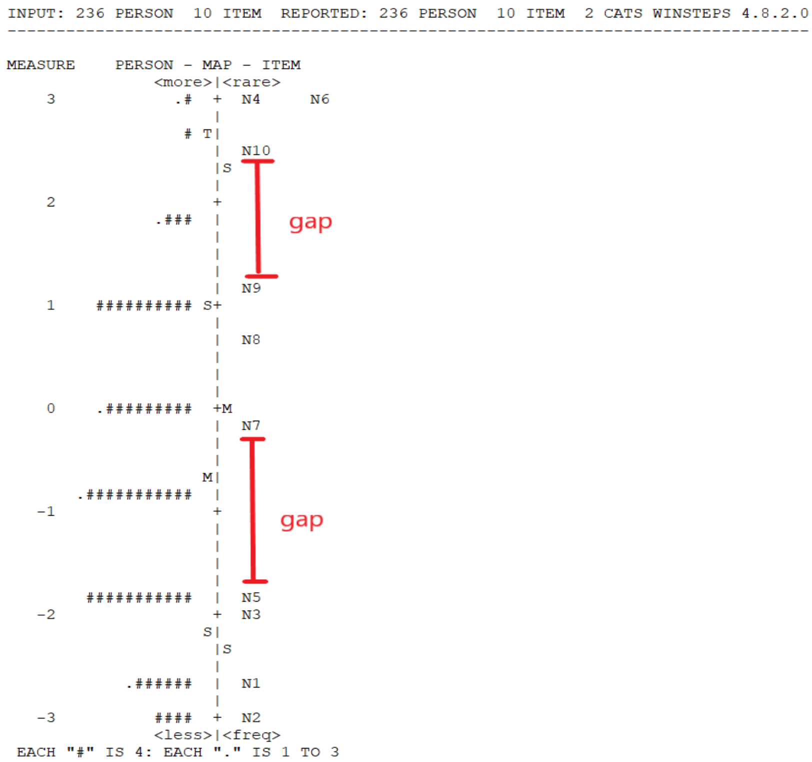

Figure 1. The person-item distribution map

Based on Figure 1, the mean of student's ability (M) is –0.60 logit which is observed lower than the averaged of the set items (0 logit), explaining that the item difficulty in the test is slightly harder for a year 8 Malaysian student, causing the performance of the students to be lower than the expected performance. Thus, some mediocre questions are required to fill in the gap between the N5 and N7 items as there was a big gap between these item difficulties to obtain more precise information about the students' understanding on algebra and arithmetic based word problems. Moreover, a difficult question is also required in between N9 and N10 items since the distribution of these item difficulties displayed larger than 0.5 logit [26]. As shown in PIDM (Figure 1), N4 (logit=3.10) and N6 (logit 3.01) items are most difficult word problem questions for these year 8 students. About 6 students (2.5%) managed to score this test with an excellent score. Meanwhile, N2 item (logit= –3.66) is easiest word problem question and most of the students were able to answer this question except 5 students (2.1%).

Figure 2 shows the comparison of person reliability between (Non-Extreme) person who summarized persons with non-extreme scores that excluding zero and perfect scores; and the (Extreme and Non-Extreme) person where it referred to person with all estimable scores that including zero and perfect scores. As a result, person reliability for 230 students under the (Non-Extreme) category is 0.56. Addition of one student who scored perfectly or in the top of categories and five students who has minimum extreme scores contributed to the increment of person reliability to 0.62 which can be observed under the (Extreme and Non-Extreme) group as predicted by Rasch model. In comparison to Cronbach alpha value of 0.64, the person reliability obtained from Rasch analysis (0.62) is slightly lower, indicating that the raw scores reproduce fair reliably as 236 students completing the word problem test. However, the low value of 1.27 on student's separation indicated that this test is not sensitive enough to distinguish between high and low achievement by students on algebra and arithmetic based word problem where more item difficulties are needed to fill in the gap between N5 to N7 as well as N9 to N10 item difficulty as shown in Figure 1 [26].

SUMMARY OF 230 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	4.3	10.0	-.60	.97	.98	-.14	1.01	.13
SEM	.1	.0	.11	.00	.04	.08	.10	.05
P.SD	1.8	.0	1.64	.06	.65	1.18	1.57	.80
S.SD	1.8	.0	1.64	.06	.65	1.18	1.57	.81
MAX.	9.0	10.0	3.77	1.19	2.90	2.61	9.90	3.46
MIN.	1.0	10.0	-3.83	.93	.28	-1.60	.18	-.78

REAL RMSE 1.08 TRUE SD 1.23 SEPARATION 1.14 PERSON RELIABILITY .56
MODEL RMSE .97 TRUE SD 1.32 SEPARATION 1.37 PERSON RELIABILITY .65
S.E. OF PERSON MEAN = .11

MAXIMUM EXTREME SCORE: 1 PERSON .4%
MINIMUM EXTREME SCORE: 5 PERSON 2.1%

SUMMARY OF 236 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	4.3	10.0	-.67	.99				
SEM	.1	.0	.12	.01				
P.SD	1.9	.0	1.80	.16				
S.SD	1.9	.0	1.80	.16				
MAX.	10.0	10.0	5.21	1.93				
MIN.	.0	10.0	-5.32	.93				

REAL RMSE 1.11 TRUE SD 1.41 SEPARATION 1.27 PERSON RELIABILITY .62
MODEL RMSE 1.00 TRUE SD 1.49 SEPARATION 1.48 PERSON RELIABILITY .69
S.E. OF PERSON MEAN = .12

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .64 SEM = 1.14
STANDARDIZED (50 ITEM) RELIABILITY = .92

Figure 2. Summaries of persons

Meanwhile, the reliability of item difficulty for 10 items in the word problem test as measured from Rasch analysis is 0.99 which is high with large item separation value of 10.27 as shown in Figure 3. This means that this test has wide item difficulty variance and is large enough to precisely pinpoint the different level of word problems' item difficulty hierarchy on the latent variable [26]. It is also reproducible to measure the understanding of the students on algebra and arithmetic based word problems. This agrees with the suggestion from previous study [27], where the ordering of item difficulty is replicable with different sets of students if the item reliability measured is high [27].

SUMMARY OF 10 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	100.3	236.0	.00	.21	.99	-.22	1.04	.06
SEM	23.6	.0	.78	.02	.04	.49	.13	.39
P.SD	70.9	.0	2.33	.05	.13	1.46	.38	1.18
S.SD	74.8	.0	2.46	.05	.14	1.54	.40	1.24
MAX.	208.0	236.0	3.10	.30	1.15	1.36	1.76	2.15
MIN.	16.0	236.0	-3.66	.16	.75	-3.17	.45	-2.46

REAL RMSE .23 TRUE SD 2.32 SEPARATION 10.27 ITEM RELIABILITY .99
MODEL RMSE .22 TRUE SD 2.32 SEPARATION 10.54 ITEM RELIABILITY .99
S.E. OF ITEM MEAN = .78

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.99
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 3. Summaries of items

The point measure correlation (PTMeasure Corr.), infit and outfit mean square (MNSQ) are used to measure the item validity. According to the item statistics in correlation order (Table 1), all the items displayed positive value of PTMeasure Corr. (x) and within the range of 0.4<x<0.8, except the N4 item with PTMeasure Corr. value of 0.34 which is found outside the acceptable range. Positive value of PTMeasure Corr. showed that no item should be removed from this word problem test. According to Bond and Fox [28], an item would be dropped or refined if the PTMeasure Corr. appeared in negative value since it showed that this item is unable to measure the construct or the item exhibits discriminant validity [29]. In addition, the high PTMeasure Corr. indicated that these items could be used to distinguish between the ability of the students on solving algebra and arithmetic based word problems.

An item would be considered misfit when these three parameters (PTMeasure Corr., MNSQ and ZSTD) are not within the ranges. The outfit MNSQ is more sensitive to the unexpected behavior by students on items far from the student's measure level where it is represented by the y-value within the range of productive of measurement of $0.5<y<1.5$ [26]. However, Bond and Fox [28] suggested that the infit and outfit MNSQ values should lied in between $0.6<y<1.4$, in order to determine the suitability of the item constructed. The latter item is referred confusing if more than 1.4 logit while an item with logit lower than 0.6 is believed to be too simple as expected by the students. Based on Table 1, the N1, N4 and N6 items did not have outfit MNSQ within the range. Meanwhile, only N1 and N8 items are having outfit ZSTD (z) not in the fit range of $-2<z<2$. As a result, none of these items (N1, N4, N6, and N8) have the values out of fit for all three parameters, thus these items are still considered fit in range and do not require further reviewed or omitted.

Table 1. Item statistics in correlation order

Items	Difficulty	Infit		Outfit		PTMeasure Corr.
		MNSQ	ZSTD	MNSQ	ZSTD	
N4	3.10	1.07	0.37	1.54	0.99	0.34
N6	3.01	0.88	–0.50	0.45	–1.04	0.44
N10	2.43	1.05	0.33	0.96	0.06	0.40
N9	1.22	1.12	1.23	0.99	0.06	0.46
N8	0.63	0.75	–3.17	0.56	–2.46	0.65
N7	–0.18	0.85	–2.02	0.96	–0.21	0.62
N5	–1.86	0.88	–1.47	0.99	0.03	0.58
N3	–2.04	1.11	1.20	1.27	1.05	0.48
N1	–2.65	1.15	1.36	1.76	2.15	0.42
N2	–3.66	1.07	0.46	0.91	–0.07	0.42

3.2. Item difficulty level of word problem characteristics concerning situation model and numerical model

The item difficulty was estimated using Rasch modelling. Next, the relationship between the item difficulty level with the respective linguistic, algebra and arithmetic factors of word problems were studied through Pearson correlation, and their information are summarized in Table 2. The item difficulty level estimates are centered at 0 logit [30]. Therefore, the negative logit indicates relatively easy items, while the positive logit representing the difficult items. The order of items was arranged based on their difficulty level, from the most difficult item (N4) to the easiest item (N2) which determined from the Rasch modelling.

Table 2. Item difficulty and other properties

Items	Difficulty level	Logits	Linguistic		Algebra		Arithmetic	
			Linguistic factor (No of words)	Implicit Info	Number of variables	Determine variables relationship within problem	Mathematical operation required	Solving steps required
N4	Very	3.1	32	Yes	2	Implied	d, s, m	3
N6	difficult	3.01	18	Yes	2	Implied	m, a, s	3
N10	Difficult	2.43	36	Yes	5	Implied	a, s, m	4
N9		1.22	24	Yes	4	Implied	a, m	4
N8	Mediocre	0.63	20	No	3	Explicitly	a	5
N7		-0.18	8	No	2	Explicitly	a	3
N5	Easy	-1.86	13	No	2	Explicitly	a, s	4
N3		-2.04	18	No	1	Explicitly	s, m	2
N1	Very easy	-2.65	12	No	1	Explicitly	s, d	1
N2		-3.66	12	No	1	Explicitly	a, m	2

In overall, the results showed that the association between word problems properties concerning linguistic and algebra with their respective item difficulty is simple and straightforward, but the arithmetic factor displayed the contradict results. Based on Table 3, there was a significant positive relationship between the number of words (linguistic factor) and the item difficulty ($r(8)=0.72$, $p=0.019$). Within these word problems, the need for realistic considerations did explain the difficulty, because the four-word problems requiring realistic consideration where N4, N6, N10 and N9 items were located at the ends of the difficulty dimension. Furthermore, the implicit information appeared to explain the difficulty of word problems where the four-word problems with the highest difficulty value had an implicit information. However, the rest of the word problems with lowest difficulty values (N1, N2, N3, N5, N7 and N8 items) contained explicit information which was clearly stated or explained and there is no room for confusion.

Other than that, there was a significant positive relationship between the number of variables (algebra factor) and the item difficulty (r(8)=0.631, p=0.05). Within these word problems, the need for realistic considerations did explain the difficulty, owing to the five-word problems requiring realistic consideration (N4, N6, N9 and N10 items) were found located at the ends of the difficulty dimension. Moreover, the implied variable relationship appeared to explain the difficulty of word problems where the four-word problems with the highest difficulty value had implied relationship within variables. Meanwhile, the rest of the word problems with lowest difficulty values (N1, N2, N3, N5, N7 and N8 items) had explicitly relationship within variables. Neither the number of mathematical operations (r(8)=0.078, p=0.83) nor the solving steps (r(8)=0.53, p=0.115) is correlated significantly with the item difficulty. As a result, the existence of the arithmetic factor of the number of mathematical operations and the solving steps did not distinguish between easy and difficult types of word problems.

Table 3. Pearson correlation between item difficulty with linguistic, algebra and arithmetic

		Linguistic factor	Item difficulty (logits)	Number of variables	Solving steps required	Number possible operation required
Linguistic factor	Pearson correlation	1	.720*	.697*	.406	-.105
	Sig. (2-tailed)		.019	.025	.244	.774
	N	20	10	10	10	10
Item difficulty (logits)	Pearson correlation	.729*	1	.631	.530	.078
	Sig. (2-tailed)	.019		.050	.115	.830
	N	10	10	10	10	10
Number of variables	Pearson correlation	.697*	.631	1	.742*	.191
	Sig. (2-tailed)	.025	.050		014	.596
	N	10	10	10	10	10
Solving steps required	Pearson correlation	.406	.530	.742*	1	-.013
	Sig. (2-tailed)	.244	.115	.014		.972
	N	10	10	10	10	10
Number possible operation required	Pearson correlation	-.105	.078	.191	-.013	1
	Sig. (2-tailed)	.774	.830	.596	.972	
	N	10	10	10	10	10

*. Correlation is significant at the 0.05 level (2-tailed)

3.3. Word problem solving strategy relates to algebra and arithmetic

Following that, the students' strategies for solving word problems are investigated, with solutions chosen by students for each question is classified as algebraic, arithmetic, or no strategy as summarized in Table 4. There are six questions about number-based word problems where each with a different level of difficulty. The N4 and N6 items being the most difficult while the N1, N2, N3, and N5 items being the easiest. Except for N5 item where there were 64.8% of students employed arithmetic strategy, majority of students chose algebra strategy to solve these number-based word problems especially for N1, N2 and N3 items. However, it is found that N4 and N6 items had the lowest percentage of correct responses (6.8% and 7.2%, respectively), although an average group of students (about 42% to 72%) selected algebra strategy to solve these items. In contrast, students achieved more than 71% correct responses for the easier N1, N2, and N3 items where more than 94% of students adopted algebra method to solve these items.

Students are noticed to be struggled with most difficult N4 item, since it required procedural knowledge in algebra to solve the equation of ($\frac{n}{2}$ -25=3n). Students are believed struggling to remove the value of 2 from $\frac{n}{2}$ fraction where students know to double the value of 25 but did not double the value of 3n. Moreover, our findings also found that N6 item has low correct response of 7.2%. The phrase "16 less than 9 times the number" in the N6 item is believed to be misinterpreted by students as "16-9n" rather than "9n-16." Due to this common error, 42.4% of students solved this number-based word problem using algebra strategy. As for N5 item, it is observed that 64.8% of students preferred to solve this item using arithmetic method. Therefore, students can solve the number-based word problems utilizing arithmetic and logical thinking, as evidenced by these results.

For N10, N9, N8 and N7 items, findings shown that students are having difficulty with age- and consecutive-based word problems which relied heavily on arithmetic strategies. Although students selected arithmetic approach to solve these items, but the percentages of accuracy are remained low. The correct response percentages for consecutive-based word problems, N7 (41.9%) and N8 (29.7%) items are found two times higher than that of respective age-based word problems, N9 (22%) and N10 (10.6%) items. The difficulty in solving these age- and consecutive-based word problems is owing to lack of comprehension of the text and an inability to define the variables. Due to the variables are not explicitly stated, students are unable to connect them to construct the equation.

Table 4. The performance of students on solving different types of word problems using various strategies

Word problem		Performance		Strategy in solution		
Items	Types	wrong (%)	correct (%)	No strategy (%)	Algebra (%)	Arithmetic (%)
N4	Numbers	93.2	6.8	11	72	16.9
N6	Numbers	92.8	7.2	38.6	42.4	19.1
N10	Age problem	89.4	10.6	30.1	19.9	50
N9	Age problem	78	22	16.5	21.2	62.3
N8	Consecutives	70.3	29.7	33.9	9.3	56.8
N7	Consecutives	58.1	41.9	17.8	8.5	73.7
N5	Numbers	31.4	68.6	6.4	28.8	64.8
N3	Numbers	28.8	71.2	1.3	98.7	-
N1	Numbers	21.2	78.8	4.2	94.5	1.3
N2	Numbers	11.9	88.1	-	99.6	0.4

For consecutive-based word problems (N7 and N8 items), students are unable to connect the variables of even consecutive numbers in equations, for example three even consecutive numbers, should return them in the form of x-2, x, and x+2. However, students are failed to do so in algebra, with majority of students relying on guessing and checking, to solve consecutive-based word problems successfully. For N9 and N10 items, it is demonstrated that students employed the arithmetic approach of backward working and algebra with two variables to solve these word problems with the success rate of less than 22%. However, most students are struggling to grasp the concept of word problems, converting, and connecting the variables into equations.

3.4. Discussion

Previous studies found that performance on solving word problems is influenced by linguistic factors [15], [21], [23], [25], algebraic reasoning and arithmetic ability [15], [21]–[25]. However, these investigations concentrated exclusively on performance based on linguistic, algebraic, and arithmetic word problems with basic semantic frameworks, without regard to difficulty of the word problems. Therefore, the difficulty of the word problems is measured by using the Rasch modelling in this study and the item difficulty level in relation to linguistic, algebraic, and arithmetic factors of word problems is investigated.

Moreover, strategies that are adopted by students for solving word problems either using arithmetic or algebraic reasoning are evaluated. The evaluation covered a range of challenges ranging from word problems to arithmetic and algebraic reasoning that required the eighth-grade students to demonstrate their problem-solving abilities. The purpose of this study is to investigate the Malaysian year 8 students' performance in solving word problems that require them to construct a proper situation model and cannot be solved only via the use of superficial coping methods such as the keywords approach. The Rasch model is used for evaluating the ability of the students on solving the items at various difficulty level in a similar scale where the characteristics of word problem items is characterized in terms of linguistic, algebraic, and arithmetic aspects.

This study is also conducted to determine the performance of students on solving word problem whether they will retain their early algebraic reasoning abilities when struggling with various difficulty levels of word problems. Moreover, the relationship between selected linguistic factors (the length of the word problem question and implicit information), algebraic factors (the number of variables and relationships within the problem), and arithmetic factors (mathematical operations and solving steps required) and the difficulty level of developing situation models is investigated and analyzed to see if it can adequately explain the difficulty level of these demanding word problems. In general, the findings suggested that superficial linguistic features appropriately characterized the difficulty of word problems which requiring implicit information inference. Similar correlations were discovered for algebra; the results revealed a link between algebra skills and performance on both easy and difficult items.

Both age-based word problems (N9 and N10 items) were difficult. However, the arithmetic factor, on the other hand, did not predict the difficulty of these word problem, whereas the linguistic and algebraic factors did. These findings are unsurprising, given the form and context of these challenging word problems that required critical thinking, as well as the fact that there are only a few strategies for solving them, such as algebra and arithmetic methods. As a result, constructing an equation from the word problem question is the most difficult aspect of determining the difficulty of word problems in algebra. For instance, the variables required in the equations were not explicitly stated, and students were required to infer them from the text content of the word problem which are abstract where their meaning often elusive, leading the students to mistranslations or misinterpretations [31]. Individual items seem to have unique characteristics within its deeper structure that contribute to the difficulty level of the item. For example, the key factor that affecting the difficulty of N4 item is the requirement for text comprehension ability, while the difficulty level of N6 item appears to be the requirement for procedural knowledge in algebra to find the final solution to this word problem.

Majority of students were able to answer these word problems (N1, N2, N3 and N5 items) which involving one variable through constructing equations and applying their algebraic procedural skills. One possible explanation for why some of the number-based word problems (N4 and N6) are more difficult than others is that the underlying structure and implicit information which are different. The N7 and N8 items are parts of the difficult items where the linguistic aspect of the term 'consecutive' creates an impediment for students on resolving this word problem. Apart from that, the failure was caused by the presence of multiple variables in this word problem. To solve these consecutive-based word problems, conceptual understanding of consecutive and algebra is necessary. The findings significantly contribute to the research on mathematics development in various ways. First, eighth-grade students are capable of correctly solving word problems including at least some equations with a single variable. It demonstrates that children can reason algebraically and emphasizes the connection between arithmetic and algebraic reasoning, as stated by [32]. Additionally, research implies that with effective instruction [31] in the real-world context issue [33], these students may properly solve even more of these words problem, laying the way for better algebra proficiency.

As a result, the difficulty level of word problems is determined not only by their linguistic, algebraic, and arithmetic characteristics, but also by their requirement for non-routine thinking [34], different combinations of cognitive skills [35] and deeper comprehension [36], [37], which can contribute for the differences in difficulty levels. However, the evidence for this finding is insufficient since the word problems employed in this study are not designed in a systematic manner to compare these characteristics. Future research should focus on word problems with systematic variation in the surface and deep structure aspects of the word problems. Moreover, when examining the ability of individual differences in solving word problems during text comprehension, algebra, and arithmetic skills, it is found that other general non-routine thinking abilities (for example, critical thinking and logical thinking) are omitted. As for this study, word problem solving performance and word problem difficulty are examined solely through the achievement of word problem tests. The future research should examine the solution processes for word problems that answered by students in greater depth via stimulated recall interviews, in order to have better understand of their difficulties in handling various demanding word problems.

3.5. Limitations of the study

A significant limitation of this study is the gaps that existed between categorization of word problems by arithmetic and algebra characteristics. The test is not sensitive enough to differentiate students' high and low achievement in algebra and arithmetic, and additional item difficulties are required to close the gap between N5 and N7 items, as well as in between N9 and N10 items. Moreover, lack of systematic design in word problem causes it unable to create distinct theory-based subcategories. Therefore, several well-studied arithmetic and algebraic features, such as categorizing students according to their arithmetic and algebraic abilities to examine their relationship with word problem performance should be conducted.

3.6. Educational implication

Difficulty with word problems can arise even when other aspects of mathematical cognition are intact [38], [39], as they require non-routine thinking and knowledge of the real-life. The difficulty with word problems is partly due to the fact that the cognitive resources that required to solve them are different and more numerous than those required for arithmetic knowledge and proficiency [40]. The findings of this study found that challenging non-routine word problems also required advanced text comprehension and analysis abilities. Thus, practice with more challenging word problems is beneficial not only for mathematics learning but also for developing problem solving and critical thinking skills. The implementation of different levels of word problems provides valuable content for mathematics education, and it is important to reach all students throughout mathematics lessons through efficient word-problem classroom instruction.

4. CONCLUSION

In this study, Rasch analysis was able to quantitatively categorize the 10 word-problems containing numbers, consecutives, and ages in the test instrument to various levels of item difficulty, which was found to be more convincing than determining item difficulty of word-problems based on experts' predefined perspectives, which varied from one to another. According to the logits as measured by the Rasch measurement, the item difficulty level of these word problems can be classified into five levels. Although the Cronbach alpha value (α=0.64) determined for this test instrument is fairly reliable where two additional item difficulties were suggested to fill in the gap in between these ten word-problems, but the reliability of item reliability determined for this test instrument is found high with large separation, indicating that the item difficulty levels of these word-problems can be ascertained and used to measure the word-problem solving ability among the students. Meanwhile, the PIDM, which measured students' ability on word-problem solving, discovered that students were performing lower than expected, owing to the test instrument's standard being slightly harder for their current level. Further investigation revealed that the item difficulty as determined by these ten-word problems was related to linguistic and algebra factors such as the number of words and variables requiring realistic considerations for explaining the difficulty. However, arithmetic factors such as the number of mathematical operations and the number of solving steps have no effect on the difficulty level of word problems. More than half of the students solved word problems with consecutives and ages using arithmetic strategies, while the majority of students solved numbers-based word problems easily using algebra methods. However, the students encountered difficulties while losing their solution strategy when the questions contained implicit data that demanded critical thinking ability.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude for the financial aids and supports provided by Universiti Putra Malaysia (UPM) through the research grant number GP-IPS/2020/9683300.

REFERENCES

[1] Malaysia Ministry of Education, *Malaysia education blueprint 2013-2025 (preschool to post-secondary education)*. Putrajaya, Malaysia, 2013.
[2] F. Avvisati, A. Echazarra, P. Givord, and M. Schwabe, "Malaysia-country note-PISA 2018 results." OECD, 2019.
[3] K. H. Kok, "PISA 2018 and Malaysia," *International Journal of Advanced Research in Education and Society*, vol. 2, no. 3, pp. 12–18, 2020.
[4] A. S. Md Yunus, A. F. M. Ayub, and T. H. Tan, "Geometric thinking of Malaysian elementary school students," *International Journal of Instruction*, vol. 12, no. 1, pp. 1095–1112, 2019.
[5] C. Chin, "Time for teachers and parents to step up," *The Star*, pp. 6–7, Feb. 2020.
[6] N. M. Tajudin and M. Chinnappan, "The link between higher order thinking skills, representation and concepts in enhancing TIMSS tasks," *International Journal of Instruction*, vol. 9, no. 2, pp. 199–214, Jul. 2016, doi: 10.12973/iji.2016.9214a.
[7] L. D. Perera and M. N. Asadullah, "Mind the gap: What explains Malaysia's underperformance in PISA?" *International Journal of Educational Development*, vol. 65, pp. 254–263, 2019, doi: 10.1016/j.ijedudev.2018.08.010.
[8] L. Bragg, "Students' conflicting attitudes towards games as a vehicle for learning mathematics: A methodological dilemma," *Mathematics Education Research Journal*, vol. 19, no. 1, pp. 29–44, Jun. 2007, doi: 10.1007/BF03217448.
[9] M. C. Utsumi and C. R. Mendes, "Researching the attitudes towards mathematics in basic education," *Educational Psychology*, vol. 20, no. 2, pp. 237–243, 2000, doi: 10.1080/713663712.
[10] G. P. Mingke and E. M. Alegre, "Difficulties encountered in mathematical word problem solving of the grade six learners," *International Journal of Scientific and Research Publications*, vol. 9, no. 6, pp. 336–345, 2019, doi: 10.29322/IJSRP.9.06.2019.p9053.
[11] E. Zakaria and N. Yusoff, "Attitudes and problem-solving skills in algebra among Malaysian matriculation college students," *European Journal of Social Sciences*, vol. 8, no. 2, pp. 232–245, 2009.
[12] A. Noyes and M. Adkins, "Reconsidering the rise in A-level mathematics participation," *Teaching Mathematics and its Applications*, vol. 35, no. 1, pp. 1–13, 2016, doi: 10.1093/teamat/hrv016.
[13] O. C. Barbu and C. R. Beal, "Effects of linguistic complexity and math difficulty on word problem solving by English learners," *International Journal of Education*, vol. 2, no. 2, pp. 1–19, Dec. 2010, doi: 10.5296/ije.v2i2.508.
[14] G. Daroczy, M. Wolska, W. D. Meurers, and H. C. Nuerk, "Word problems: A review of linguistic and numerical factors contributing to their difficulty," *Frontiers in Psychology*, vol. 6, no. APR, pp. 1–13, 2015, doi: 10.3389/fpsyg.2015.00348.
[15] A. J. H. Boonen, B. B. de Koning, J. Jolles, and M. van der Schoot, "Word problem solving in contemporary math education: A plea for reading comprehension skills training," *Frontiers in Psychology*, vol. 7, pp. 1–10, Feb. 2016, doi: 10.3389/fpsyg.2016.00191.
[16] P. Sepeng and A. Madzorera, "Sources of difficulty in comprehending and solving mathematical word problems," *International Journal of Educational Sciences*, vol. 6, no. 2, pp. 217–225, Mar. 2014, doi: 10.1080/09751122.2014.11890134.
[17] S. H. Teoh, P. Singh, T. H. Cheong, N. A. Nasir, N. S. M. Rasid, and N. Zainal, "An analysis of knowledge in STEM: Solving algebraic problems," *Asian Journal of University Education*, vol. 26, no. 2, pp. 131–140, 2020, doi: 10.24191/ajue.v16i2.10304.
[18] N. M. Siew, J. Geofrey, and B. N. Lee, "Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using dragonbox 12+App," *The Research Journal of Mathematics and Technology*, vol. 5, no. 1, pp. 66–79, 2016.
[19] N. N. Ibrahim, A. F. M. Ayub, A. S. M. Yunus, R. Mahmud, and K. A. Bakar, "Effects of higher order thinking module approach on pupils' performance at primary rural school," *Malaysian Journal of Mathematical Sciences*, vol. 13, no. 2, pp. 211–229, 2019.
[20] F. Depaepe, E. de Corte, and L. Verschaffel, "Students' non-realistic mathematical modeling as a drawback of teachers' beliefs about and approaches to word problem solving," in *From beliefs to dynamic affect systems in mathematics education. Advances in Mathematics Education*, Cham: Springer, 2015, pp. 137–156, doi: 10.1007/978-3-319-06808-4_7.
[21] A. J. H. Boonen, M. van der Schoot, F. van Wesel, M. H. de Vries, and J. Jolles, "What underlies successful word problem solving? A path analysis in sixth grade students," *Contemporary Educational Psychology*, vol. 38, pp. 271–279, 2013, doi: 10.1016/j.cedpsych.2013.05.001.
[22] L. S. Fuchs *et al.*, "The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems.," *Journal of Educational Psychology*, vol. 98, no. 1, pp. 29–43, Feb. 2006, doi: 10.1037/0022-0663.98.1.29.

[23] P. M. Vilenius-Tuohimaa, K. Aunola, and J. E. Nurmi, "The association between mathematical word problems and reading comprehension," *Educational Psychology*, vol. 28, no. 4, pp. 409–426, 2008, doi: 10.1080/01443410701708228.
[24] L. S. Fuchs, J. K. Gilbert, D. Fuchs, P. M. Seethaler, and B. N. Martin, "Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension," *Scientific Studies of Reading*, vol. 22, no. 2, pp. 152–166, 2018, doi: 10.1080/10888438.2017.1398259n.
[25] H. L. Swanson, J. B. Cooney, and S. Brock, "The influence of working memory and classification ability on children's word problem solution," *Journal of Experimental Child Psychology*, vol. 55, no. 3, pp. 374–395, 1993, doi: 10.1006/jecp.1993.1021.
[26] J. M. Linacre, *A user's guide to winsteps, ministeps: Rasch model computer program manual 3.74.0*. Winsteps.com, 2012.
[27] A. Z. bin Khairani and N. bin A. Razak, "Advance in educational measurement: a Rasch model analysis of mathematics proficiency test," *International Journal of Social Science and Humanity*, vol. 2, no. 3, pp. 248–251, 2013, doi: 10.7763/IJSSH.2012.V2.104.
[28] T. G. Bond and C. M. Fox, *Applying the rasch model: Fundamental measurement in the human sciences*, 2nd Ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 2007.
[29] D. J. Krus and R. G. Ney, "Convergent and discriminant validity in item analysis," *Educational and Psychological Measurement*, vol. 38, pp. 135–137, 1978, doi: 10.1177/001316447803800118.
[30] W. J. Boone, "Rasch analysis for instrument development: why, when, and how?" *CBE—Life Sciences Education*, vol. 15, no. 4, Dec. 2016, doi: 10.1187/cbe.16-04-0148.
[31] M. J. Nathan, W. Kintsch, and E. Young, "A theory of algebra-word-problem comprehension and its implications for the design of learning environments," *Cognition and Instruction*, vol. 9, no. 4, pp. 329–389, Dec. 1992, doi: 10.1207/s1532690xci0904_2.
[32] L. S. Fuchs *et al.*, "Contributions of domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge," *Developmental Psychology*, vol. 48, no. 5, pp. 1315–1326, 2012, doi: 10.1037/a0027475.
[33] S. L. Ku, F. Razali, and A. F. M. Ayub, "The effectiveness of realistic mathematics education approach toward students learning: A systematic literature review of empirical evidence," *Journal of Critical Reviews*, vol. 7, no. 1, pp. 548–552, 2020.
[34] N. Pongsakdi, A. Kajamies, K. Veermans, K. Lertola, M. Vauras, and E. Lehtinen, "What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills," *ZDM*, vol. 52, no. 1, pp. 33–44, 2020, doi: 10.1007/s11858-019-01118-9.
[35] A. R. Strohmaier, F. Reinhold, S. Hofer, M. Berkowitz, B. Vogel-Heuser, and K. Reiss, "Different complex word problems require different combinations of cognitive skills," *Educational Studies in Mathematics*, vol. 109, pp. 89–114, 2022, doi: 10.1007/S10649-021-10079-4.
[36] S. Stephany, "The influence of reading comprehension on solving mathematical word problems: A situation model approach," in *Diversity Dimensions in Mathematics and Language Learning*, Berlin: Walter de Gruyter GmbH, 2021, pp. 370–395, doi: 10.1515/9783110661941-019.
[37] D. Hadianto, V. S. Damaianti, Y. Mulyati, and A. Sastromiharjo, "Does reading comprehension competence determine level of solving mathematical word problems competence?" *Journal of Physics: Conference Series*, vol. 1806, no. 2021, p. 012049, 2021, doi: 10.1088/1742-6596/1806/1/012049.
[38] D. D. Cummins, W. Kintsch, K. Reusser, and R. Weimer, "The role of understanding in solving word problems," *Cognitive Psychology*, vol. 20, pp. 405–438, 1988, doi: 10.1016/0010- 0285(88)90011-4.
[39] K. R. Koedinger and M. J. Nathan, "The real story behind story problems: effects of representations on quantitative reasoning," *Journal of the Learning Sciences*, vol. 13, no. 2, pp. 129–164, Apr. 2004, doi: 10.1207/s15327809jls1302_1.
[40] H. L. Swanson and M. Beebe-Frankenberger, "The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties," *Journal of Educational Psychology*, vol. 96, pp. 471–491, 2004, doi: 10.1037/0022-0663.96.3.471.

BIOGRAPHIES OF AUTHORS

Ku Soh Ting is a PhD candidate from the Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia (UPM). She earned her bachelor's degree in Materials Science from Universiti Kebangsaan Malaysia (UKM) in 2008, and she obtained her Master of Education from Asia e University (AeU) in 2015. Her research focuses on mathematics education, computational thinking, and problem-based learning. She can be contacted at email: ting_blue2003@yahoo.com.

Othman Talib is a Senior Lecturer at the Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia (UPM). He completed his first degree in Chemistry from Universiti Kebangsaan Malaysia (UKM) in 1986. A year later he was appointed as a Chemistry teacher at the Matriculation Centre, UPM. He completed his master in Pedagogy from UPM in 1999. He was appointed lecturer with the Faculty of Educational Studies in the university in February 2000 and went on to pursue his graduate studies in education at the University of Adelaide, Australia in 2003. He obtained his Doctor of Education degree in 2007. His research interests are on science education, chemistry education and apps. He can be contacted at email: otalib@upm.edu.my.

Ahmad Fauzi Mohd Ayub, Ph.D., is a Professor at the Faculty Educational Studies, University Putra Malaysia with a doctorate degree. His research interests are educational technology, multimedia, learning management system, technology-mediated tools in teaching, and learning. He can be contacted at email: afmy@upm.edu.my.

Maslina Zolkepli currently is a senior lecturer at the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She received her Bachelor of Computer Science in 2007, Master of Computer Science in 2010, both from Universiti Putra Malaysia. In 2015, she received her Doctor of Engineering Degree in the field of Computational Intelligence and Systems Science from Tokyo Institute of Technology, Japan. Her research interests include data science, fuzzy systems, complex networks, and computational intelligence. She can be contacted at email: masz@upm.edu.my.

Chen Chuei Yee is received the DPhil in Mathematics from University of Oxford, United Kingdom, in 2014. Prior to that, she obtained her Bachelor and Master degrees from Universiti Putra Malaysia in 2007 and 2009, respectively. She is currently a senior lecturer in the Department of Mathematics and Statistics, Universiti Putra Malaysia. Her research interests include calculus of variations, fuzzy variational problems, optimization and applications, optimal control problems, and mathematics education. She can be contacted at email: cychen@upm.edu.my.

Teh Chin Hoong is a senior lecturer at the ASASIpintar Program, Pusat GENIUS@Pintar Negara, Universiti Kebangsaan Malaysia (UKM). He received his Bachelor of Science (BSc) and Master of Science (MSc) in Materials Science at Universiti Kebangsaan Malaysia (UKM) in 2005 and 2009, respectively. Later, he studied organic chemistry and obtained his Doctor of Philosophy (PhD) in Chemistry from UKM in 2013. His research interests include project-based learning and chemistry education. He can be reached at email: chteh@ukm.edu.my.