Gender Difference of Rural Grade Five Students' Performance in Solving Word Problems Involving Measurement Formulae and Higher-Order Thinking Skills

Menaga Suseelan[1], *Chew Cheng Meng[2], Chin Huan[3], Wun Thiam Yew[4], Lim Hooi Lian[5], Ahmad Zamri Khairani[6], Parmjit Singh[7]
[1,2,3,4,5,6]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia
[7]Faculty of Education, Universiti Teknologi MARA, Malaysia
cmchew@usm.my *corresponding author

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: This study assessed the gender difference in the performance of Grade Five elementary students from the rural region in solving word problems involving measurement formulae and higher-order thinking skills. The sample of this study comprised 109 students from three types of elementary schools, namely National School (NS), Chinese National-Type School (CNTS) and Tamil National-Type School (TNTS) located in the rural region of Penang, Malaysia. The findings indicate that there is no significant difference between the mean scores of the male and female students in the rural region of Penang, but both genders were found to perform poorly in solving measurement formulae word problems involving higher-order thinking skills. This implies the needs to shift the focus of mathematics teaching from procedural fluency to mastery of mathematical reasoning as well as problem solving skills to promote rural students' higher-order thinking skills.

Keywords: Gender differences, problem-solving skills, measurement formula, higher-order thinking skills, elementary school.

1. Introduction

Measurement formula falls under the principal real-world applications of mathematics which connects two key realms of mathematics, namely geometry and real numbers (Clements & Battista, 2001). After learning the basic measurement formulae such as perimeter, area and volume formulae, word problems involving measurement formulae will be introduced to students at the end of the topic. According to Verschaffel, Greer, and Decorte (2000), word problems involve text descriptions of problem situations in which one or more questions are needed to be answered using the application of mathematical operations to numerical information available in the problem given. Introducing word problems to the students after they have learnt the measurement formulae would help them bridge the concepts with the real word applications (Verschaffel, Van Dooren, Greer, & Mukhopadhyay, 2010), and ultimately develop their problem-solving competency (Schoenfeld, 1985). Nonetheless, Reiss and Torner (2007) argued that the word problems included in the textbook at the end of each topic might be too routine and can only be considered as an exercise for them rather than a problem. Thus, Jader, Lithner, and Sidenvall (2020) recommended that students should be engaged in solving non-routine mathematical word problems which involve higher-order thinking skills (HOTS) such as analysing, evaluating, and creating (Malaysian Ministry of Education [MOE], 2017) in order to enhance their problem-solving skills.

Despite the importance of word problems involving measurement formulae and HOTS, students worldwide are found to have difficulties in solving these problems (Machaba, 2016). In the Malaysian context, Grade Four students are reported to have weak conceptual understanding and adaptive reasoning in measurement formulae (Chew et al., 2016). In addition, the Malaysian students might also be weak in solving word problems involving measurement formulae and HOTS because these two interconnected strands enable students to relate the concepts and the situations stated in the word problems (Kilpatrick, 2001). Due to the socio-economic disadvantages of the schools, students from the rural region are reported to have lower mathematical proficiency as compared to students from the urban region (MOE, 2012). To address the equity gap between urban and rural students' mathematical proficiency and problem-solving skills, there is an urgent need to examine rural Malaysian students' performance in solving word problems involving measurement formulae and HOTS.

As advocated by the Organisation for Economic Co-operation and Development [OECD] (2009), a deeper understanding of students' problem-solving skills would be gained by studying the gender difference. Nonetheless, the findings of gender difference in problem-solving skills are inconclusive. For instance, Leder

(2019) reported that a small gender difference in complex problem solving were found to be in favour of boys. On the contrary, Pang and Seah (2020) as well as Thien (2016) found that girls were seen to perform better in solving the context-based word problems in the Programme for International Student Assessment (PISA) conducted in 2015. The inconclusive findings hinder the planning of remediation to address students' needs as well as to foster educational improvement and reduce gender disparity.

Thus, this paper sought to investigate the gender difference of rural Grade Five students in solving word problems involving measurement formulae and HOTS at both the assessment and item levels. The research questions which would be addressed in this study are as follows:

(1) What is rural Grade Five students' performance in solving word problems involving measurement formulae and HOTS in terms of gender?

(2) Is there any significant difference between the performance of rural Grade Five boys and girls in solving word problems involving measurement formulae and HOTS?

To answer the second research question, the null and alternative hypotheses were formulated as follows:

H_o: There is no significant difference between the performance of rural Grade Five boys and girls in solving word problems involving measurement formulae and HOTS.

H_a: There is a significant difference between the performance of rural Grade Five boys and girls in solving word problems involving measurement formulae and HOTS.

2. Literature Review

2.1. Gender Difference in Mathematics Performance

In line with the global education agenda to promote gender equity (Meinck & Brese, 2019; UNESCO, 2014), investigation of gender difference continues to serve as an important research area in mathematics education (Leder, 2019). In general, the Grade Four boys and girls performed equally well in the international large-scale assessment, namely the Trends in International Mathematics and Science Study (TIMSS) for the year of 2015 (Mullis, Martin, Foy, & Hooper, 2016). Yet, gender difference in mathematics achievement was found within some of the countries that participated in the study. Out of the 49 participating countries, eighteen were reported to have gender difference in mathematics performance which favoured boys, while eight were reported to have gender difference which favoured girls in TIMSS 2015 (Mullis et al., 2016). Nonetheless, the low-performing group (i.e., the bottom 20 %) consisted of almost equal percentage of boys and girls for each cycle of TIMSS from 1995 to 2015 (Meinck & Brese, 2019).

Several studies have been conducted to compare the mathematics performance of boys and girls by region. Based on PISA 2009 database, Forgasz and Hill (2013) found that the Grade Eight boys performed significantly better than the girls in the western countries, such as Australia, United Kingdom, United States, and Canada. Similarly, Fryer Jr. and Levitt (2010) also found that the Grade Six girls in United States started to lose the stand in mathematics subject even though the mean difference did not exist between the boys and girls upon entry to school. In contrast to the studies conducted in the western countries, girls from non-OECD countries were found to outperform boys in mathematics in TIMSS 2011 (Reilly, Neumann, & Andrews, 2017). A similar trend was found in the middle east context. According to Ghasemi and Burley (2019), Saudi Arabi, Oman, Jordan, Bahrain, and Kuwait are among the countries which recorded the highest gender difference in mathematics achievement in favour of girls for TIMSS 2015. However, the findings in the Asian context are inconclusive. The Korean Grade Four boys significantly outperformed the girls in TIMSS 2015 (Pang & Seah, 2020), while the Indonesian Grade Four girls significantly outperformed the boys in TIMSS 2015 (Mullis et al., 2016). Meanwhile, gender difference in mathematics achievement for TIMSS 2015 was not significant in Singapore and Japan (Mullis et al., 2016).

Rather than the overall mathematical proficiency, several studies have been conducted to determine gender difference in achievement for each mathematics content domain. In TIMSS 2007, the Grade Eight girls scored significantly higher in Algebra, while the Grade Eight boys scored significantly higher in the other mathematics content domains such as Geometry, Data, and Number (Louis & Mistele, 2012). A similar finding was reported in PISA 2012 across OECD countries. As demonstrated by Leder and Forgasz (2018), the Grade Eight boys from OECD countries scored higher than the girls in the mathematics content domain of space and shape with a mean difference of nine points. This contradicts with the study conducted by Louis and Mistele (2012) where the Grade Eight girls in the OECD countries are found to have better performance than the boys in the mathematics content domain of uncertainty and data with a mean difference of nine points (Leder & Forgasz, 2018). Meanwhile, Weldeana (2015) found that there was no significant difference between the performance of boys and girls from the high-ability group on the stereotype-free tasks related to geometry. As problem solving is the heart of mathematics learning, several studies have been conducted to investigate gender difference in problem solving skills. Nonetheless, the findings of the studies consistently reported that boys performed better in the more difficult, unfamiliar, life-related mathematical problems, while girls performed better in the familiar, less difficult, and not life-related mathematical problems (Innabi & Dodeen, 2018; Leder, 2019).

In the Malaysian context, the past related studies focused on the overall mathematical proficiency, mathematical performance for each content domain as well as problem-solving skills. The gender difference in favour of Grade Eight girls was reported in TIMSS 2007 and 2011 (Saw, 2015) as well as PISA 2012 (Thien, 2016). Specifically, the Grade Eight girls significantly outperformed the boys in three out of four mathematics content domains, namely (1) change and relationship, (2) quantity, and (3) uncertainty and data in PISA 2012 (Thien, 2016). Even though the Grade Eight boys scored higher in geometry in PISA 2012, the mean differences were not significant (Thien, 2016). In terms of mathematical processes, gender difference that favoured girls was reported in all mathematical processes being assessed in PISA 2012, that is, employing, formulating and interpreting (Thien, 2016). In elementary school context, Abedalaziz (2011) and Kashefi et al. (2017) found that that there was no significant difference between the performance of boys and girls in solving mathematical problems involving higher-order thinking skills. Besides, the problem-solving strategies used, and the errors made were the same for both genders (Abedalaziz, 2011).

In general, the past-related studies mainly focused on gender difference in the overall mathematical proficiency of both Grade Four and Grade Eight students. In the Malaysian context, the gender differences have been investigated based on the overall mathematical proficiency, mathematical performance for each content domain as well as problem-solving skills. Whilst the past-related studies focused on students from both rural and urban regions, this study sought to investigate the gender difference in problem solving of rural students for the topic of measurement formulae which involves HOTS. With a more focused contain domain and a more in-depth analysis at the item-level, more comprehensive findings could be obtained.

2.2. Contributing to Difficulties in Mathematics Problem Solving

The special characteristic of word problems is that the question(s) needed to be answered are embedded in the situation which is presented in text descriptions (Verschaffel et al., 2000). Following this, read and understand the word problems would be the first step in the problem-solving process (Polya, 1945). In fact, read and understand the word problems require literacy skills (White, 2009) which allow students to interpret the question(s) in mathematical context (O'Connor & Norton, 2020). Poor literacy skills eventually hinder students from understanding the situation presented in the word problems and from determining the requirement of the word problems. Thus, the difficulties in solving word problems might be rooted in students' poor literary skills (Fatmanissa & Kusnandi, 2017; Pongsakdi et al., 2020).

Besides literacy skills, numeracy skills which enable students to understand the numerical information presented in word problems, is another key competency needed in problem solving (White, 2009). Apart from understanding the text descriptions, students must be able to decode the mathematical event stated in the word problems, followed by identifying the relationships among the numbers provided in the word problems as well as specifying the missing quantity that needs to be determined (Fuchs, Fuchs, Seethaler, & Barnes, 2020). With poor numeracy skills, students might not be able to identify the correct mathematical operation and form correct mathematical sentences with relevant numbers (O'Connor & Norton, 2020). Thus, poor numeracy skills might also contribute to the difficulties in solving word problems.

Moreover, lack of conceptual understanding as well as arithmetic skills will also contribute to the difficulties in mathematics problem solving (Scheibling-Seve, Pasquinelli, & Sander, 2020). In contrast with the basic arithmetic tasks, solving word problems requires both conceptual understanding and procedural knowledge (Scheibling-Seve et al., 2020). Conceptual understanding enables students to know the underlying reasoning for certain arithmetic operations and procedures (Crooks & Alibali, 2014). Lack of conceptual understanding will lead to the failure of students to apply the mathematical concept(s), identify the correct mathematical operation(s) and form correct mathematical sentence(s). Meanwhile, poor procedural knowledge will hinder students from executing the answer using correct arithmetic algorithm (Daroczy, Wolska, Meurers, & Nuerk, 2015). Thus, students' difficulties in solving word problems could be due to poor conceptual understanding as well as non-mastery of procedural knowledge.

In addition, weak metacognitive skills will also contribute to the difficulties in solving word problems especially non-routine word problems (Abdullah, Rahman, & Hamzah, 2017). According to Jader et al. (2020) and Kloosterman (1992), non-routine word problems cannot be solved using known methods or formulae at first sight. In this sense, students could not solve the problems by identifying keywords (Salemeh & Etchells, 2016) and applying memorized methods (Palraj, DeWitt, & Alias, 2017). Indeed, students need to interpret the word problems and outline the key information, followed by choosing appropriate heuristic strategy, carry out the plan, evaluate the strategy used and revise the plan if necessary (Abdullah et al., 2017). In this regard, students will have to acquire metacognitive skills such as planning, monitoring, and evaluating (Spada, Georgiou, & Wells, 2010) to solve the non-routine word problems. Lack of metacognitive skills leads to the failure of students in choosing appropriate heuristic strategy. Besides, they may not be able to evaluate the plan, make necessary revision or consider other alternative solutions to solve the problems (Abdullah et al., 2017). Thus, students' difficulties in solving word problems might also be caused by weak metacognitive skills.

3. Methodology

This study employed a cross-sectional survey research design which provides a succinct plan in the data collection and data analysis to address the research problems. Specifically, the use of a cross-sectional survey would provide a snapshot of the rural Grade Five students' performance in solving word problems involving measurement formulae and higher-order thinking skills at a single point of time (Gay, Mills, & Airasian, 2012).

3.1. Population and Sampling

The targeted population of the study were Grade Five students in the rural region of Malaysia. Due to the logistic constraints, the sampling frame was reduced to a state, rather than the whole country. Thus, the accessible sampling of this study was Grade Five students in the rural region of Penang state. The sample of the study was selected using non-proportional stratified cluster sampling to ensure the representation of relevant strata within the sample (Gay et al., 2012). The population of this study was first being stratified into three strata based on school type, namely National School (NS), Chinese National-Type School (CNTS) and Tamil National-Type School (TNTS). Although the same mathematics curriculum was used in the three types of school, the medium of instruction for mathematics lessons is different. After stratifying the population into the three strata, the sample was randomly selected by cluster which was defined as a school in each stratum. In other words, all Grade Five students in the selected schools would participate in the study. As shown in Table 1, a total of 36, 35 and 38 Grade Five students from NS, CNTS and TNTS were selected as the sample of this study.

Table 1. Demographic Information of the Sample of Study

	Gender		Total	
School Type	**F**	**M**	**n**	**%**
NS	22	14	36	33.03
CNTS	21	14	35	32.11
TNTS	16	22	38	34.86
Total	59	50	109	100.00

3.2. Instrument of the Study

In this study, Grade Five students' performance in solving word problems involving measurement formulae and HOTS was measured using a problem-solving test. The problem-solving test consisted of eight open-ended word problems adapted from the Grade Five mathematics workbooks. The contents covered by these eight items include the perimeter of squares, perimeter of rectangles, perimeter of triangles, perimeter of regular polygons, area of squares, area of rectangles, area of triangles, volume of cubes, and volume of cuboids. The problem-solving test consisted of four items involving a single concept, and four items involving multiple concepts. These eight items were categorised as HOTS because they fall in the top three levels of the Revised Bloom's Taxonomy (MOE, 2017), namely analysing, evaluating, and creating (Anderson et al., 2001). The students would have to analyse the requirement of the word problems, evaluate whether the measurement formulae can be directly applied to solve the word problems, and synthesise the strategy to solve the word problems. The content covered by each item and the cognitive domain involved in the items are as summarised in Table 2. The marks allocated for the word problems ranged from 4 to 6 depending on the difficulty of the word problems as well as the number of steps involved.

The instrument was developed in English and translated into Malay, Mandarin and Tamil language to comply with the medium of instruction used in NS, CNTS and TNTS, respectively. The Malay, Mandarin and Tamil versions of the problem-solving test were validated by two subject matter experts from NS, CNTS and TNTS respectively, to ensure the item relevance and the content coverage of the problem-solving test. The results of the validation are tabulated in Table 3. The S-CVI for content coverage was 1.00 for all the three types of schools, while the S-CVI for the item relevance was .94, 1.00 and 1.00 for NS, CNTS and TNTS, respectively. With the values surpassing the minimum threshold of .80, the problem-solving test had a high content coverage, and all the items were relevant to the learning standards (Polit & Beck, 2006) related to measurement formulae in Grade Five Mathematics Curriculum. After the validation process, Cronbach's Alpha was used to determine the reliability of each version of the problem-solving test which consisted of polytomous items. With the Cronbach's Alpha value of .83, .81 and .78 respectively, the Malay, Mandarin and Tamil versions of the problem-solving test were reliable (Pallant, 2016).

Table 2. Cognitive Domain Involved and Content Covered in the Problem-Solving Test

Items	Constructs measured	Cognitive Domain	Mark Allocation
Q1	• Perimeter of rectangle	• Analysing • Evaluating • Creating	4
Q2	• Perimeter of polygon	• Analysing • Evaluating • Creating	4
Q3	• Area of rectangle • Area of triangle	• Analysing • Evaluating • Creating	6
Q4	• Perimeter of rectangle • Area of rectangle	• Analysing • Evaluating • Creating	6
Q5	• Perimeter of square • Area of square	• Analysing • Evaluating • Creating	5
Q6	• Perimeter of square • Perimeter of triangle • Area of triangle	• Analysing • Evaluating • Creating	6
Q7	• Volume of cube	• Analysing • Evaluating • Creating	6
Q8	• Volume of cuboid	• Analysing • Evaluating • Creating	4

Table 3. Validity and Reliability of Problem-Solving Test

Version	Content Coverage (S-CVI)	Item Relevance (S-CVI)	Reliability (Cronbach Alpha)
Malay	1.00	.94	.83
Mandarin	1.00	1.00	.81
Tamil	1.00	1.00	.78

3.3. Research Procedure

The data collection of the study began by administering the problem-solving test which comprised eight open-ended word problems involving measurement formulae and HOTS to the participants. The problem-solving test took 1.5 hours. Clear instructions were given to the students so that they would show all their workings on the answer scripts. After administering the test, the answer scripts were collected and marked based on the marking scheme and the scores were categorised into grades and proficiency level proposed by the Malaysian Examination Syndicate (2016) as shown in Table 4. Then, the independent-samples t-test was performed to determine the gender difference in the participants' test scores. The data were analysed by using the Statistical Package for the Social Sciences (SPSS) version 26.

Table 4. Guideline for Categorising Scores into Proficiency Level

Percent Score	Grade	Proficiency Level
80 – 100	A	Excellent proficiency level was achieved
65 – 79	B	Good proficiency level was achieved
50 – 64	C	Satisfactory proficiency level was achieved
40 – 49	D	Minimum proficiency level was achieved
0 – 39	E	Minimum proficiency level was not achieved

Source: Malaysia Examination Syndicate (2016)

4. Findings

During the data analysis, the participants' scores were categorised into grades and the corresponding proficiency levels based on the guideline provided by the Malaysian Examination Syndicate (2016). The profile of all the 109 participants' performance in solving the word problems involving measurement formulae and HOTS is reported in Table 5. As shown in Table 5, none of the rural Grade Five students obtained grade A in the NS, CNTS and TNTS. There were only two Grade Five boys but no Grade Five girls from the CNTS obtained grade B in the test. However, none of the Grade Five students in the NS and TNTS scored grade B in the problem-solving test.

Table 5. Grades of Students Categorised According to School Type and Gender

	NS				**CNTS**				**TNTS**				**Overall**			
	Gender		**Total**		**Gender**		**Total**		**Gender**		**Total**		**Gender**		**Total**	
Grade	**F**	**M**	***n***	**%**	**F**	**M**	***n***	**%**	**F**	**M**	***n***	**%**	**F**	**M**	***N***	**%**
A	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00
B	0	0	0	0.00	0	2	2	5.71	0	0	0	0.00	0	2	2	1.83
C	2	2	4	11.11	1	0	1	2.86	3	2	5	13.16	6	4	10	9.17
D	1	1	2	5.56	2	0	2	5.71	2	4	6	15.79	5	5	10	9.17
E	19	11	30	83.33	18	12	30	85.71	11	16	27	71.05	48	39	87	79.82
Total	22	14	36	100.00	21	14	35	100.00	16	22	38	100.00	59	50	109	100.00

By comparing the Grade Five students' grades across gender, the findings show a similar trend in all the three types of schools. Regardless of gender, majority of the Grade Five students (NS: 83.33%; CNTS: 85.71%; TNTS: 71.05 %) in the three types of rural schools scored grade E in the problem-solving test. A similar finding was reflected in the overall sample. A total of 79.82 percent of the Grade Five students scored grade E in the test. This indicates that more than three quarters of the Grade Five students in the three types of rural schools did not even achieve a minimum proficiency level in solving the word problems involving measurement formulae and HOTS.

After profiling the participants' grades, the independent-samples t-test was conducted to determine gender difference of the participants within and across the three types of schools in their performance in solving the word problems involving measurement formulae and HOTS. The results of the independent-samples t-test are as reported in Table 6.

Table 6. Result of Independent Sample t-Tests at Assessment Level

School Type	**Gender**	***n***	***M***	***SD***	***t***	***p***
NS	Male	14	22.86	15.31	-.74	.46
	Female	22	26.06	10.67		
CNTS	Male	14	16.83	20.47	-.63	.54
	Female	21	20.42	13.57		
TNTS	Male	22	20.51	16.60	-.72	.48
	Female	16	24.44	16.59		
Overall	Male	50	20.13	17.23	-1.18	.24
	Female	59	23.62	13.49		

The mean scores of Grade Five girls (NS: M = 26.06, SD = 10.67; CNTS: M = 20.42, SD = 13.57; TNTS: M = 24.44, SD = 16.59) were slightly higher than Grade Five boys (NS: M = 22.86, SD = 15.31; CNTS: M = 16.83, SD = 20.47; TNTS: M = 20.51, SD = 16.60 in the NS, CNTS and TNTS, respectively. However, the results of the independent-samples t-test indicate that there was no significant difference in the mean scores of Grade Five boys and girls [NS: t (34) = -.74, p = .46; CNTS: t (33) = -.63, p = .54; TNTS: t (34) = -.72, p = .48] in the three types of rural schools. With an eta squared of .02, the magnitude of the difference in the mean scores (mean difference = 3.90) was very small in NS (Cohen, 1988). Likewise, the magnitude of the difference in the mean scores (CNTS: mean difference = 3.59; TNTS: mean difference = 3.48) was also very small in CNTS and TNTS with an eta squared of .01 (Cohen, 1988).

Overall, the Grade Five girls in the sample had a slightly higher mean score than that of the Grade Five boys. The result of the independent-samples t-test indicate there was no significant difference in the mean scores of the Grade Five boys (M = 20.13, SD = 17.23) and the Grade Five girls (M = 23.62, SD = 13.49; t (107) = -1.18, p = .24 > .05) in the overall sample. With an eta squared of .01, the magnitude of the difference in the mean scores (mean difference = -3.48) was very small (Cohen, 1988).

Regardless within or across the three types of rural schools, the results of the independent-samples t-tests indicate that the gender difference among Grade Five students in the rural region was not significant, even though the girls slightly outperformed the boys. None of the null hypotheses was rejected. This implies that the performance of the Grade Five boys and girls in solving the word problems involving measurement formulae and HOTS was almost equivalent in each of the three types of rural schools and in the overall sample.

The data was further analysed at the item level. The independent-samples t-test was conducted to determine the gender difference in student performance for solving each of the word problems involving measurement formulae and HOTS. The results of the independent-sample t-test are reported in Table 7. For the three types of rural schools, the mean score of Grade Five girls was higher than the Grade Five boys for items Q2, Q3, and Q4, whereas the mean score of Grade Five boys was higher than that of Grade Five girls for item Q8. The comparisons of Grade Five student's performance on items Q1, Q5, Q6 and Q7 by gender were inconclusive across the three types of rural schools.

In general, there was no significant difference in the mean item scores between the Grade Five boys and Grade Five girls for each type of rural schools, except for item Q4 in the NS. With a p-value of .00 (<.05), the Grade Five girls (M = 3.36, SD = 1.43) scored significantly higher than the Grade Five boys (M = 1.43, SD = 1.65) on item Q4 with a mean difference of 1.93. With an eta squared of .03, the magnitude of the difference in the mean scores was very small (Cohen, 1988). A similar finding was found in the independent-samples t-test conducted using the overall sample of the study. There was a significant difference between the mean scores of Grade Five boys (M = 1.08, SD = 1.59) and Grade Five girls (M = 2.12, SD = 1.96; t (107) = -3.06, p = .00 < .05). Although the girls performed significantly better than the boys on item Q4 with a mean score difference of 1.04, the eta square of .01 indicated that the magnitude of the mean difference was very small (Cohen, 1988).

Table 7. Result of Independent Sample t-Tests at Assessment Level at Item-level

School Type	Item	Male		Female		Mean Difference	*t*	*p*
		M	***SD***	***M***	***SD***			
NS	Q1	1.93	1.49	1.95	1.46	-0.02	-0.05	.96
	Q2	0.93	1.38	1.14	1.46	-0.21	-0.43	.67
	Q3	2.14	1.79	2.55	1.53	-0.41	-0.72	.48
	Q4	1.43	1.65	3.36	1.43	-1.93	-3.73	.00
	Q5	2.36	2.21	1.77	1.66	0.59	0.91	.37
	Q6	0.14	0.36	0.32	0.48	-0.18	-1.17	.25
	Q7	0.71	1.33	0.18	0.59	0.53	1.65	.11
	Q8	0.64	0.93	0.45	1.06	0.19	.56	.59
CNTS	Q1	0.50	1.29	0.57	1.08	-0.07	-0.18	.86
	Q2	1.64	1.98	2.62	1.91	-0.98	-1.46	.15
	Q3	0.86	0.66	1.05	0.80	-0.19	-0.73	.47
	Q4	0.36	0.84	0.95	1.47	-0.60	-1.37	.18
	Q5	1.21	2.12	1.90	2.19	-0.69	-0.93	.36
	Q6	0.93	2.16	0.81	1.40	0.12	0.20	.84
	Q7	0.86	1.56	0.43	0.60	0.43	1.15	.26
	Q8	1.21	1.58	0.86	1.39	0.36	0.71	.49
TNTS	Q1	1.45	1.41	1.44	1.03	0.01	0.04	.97
	Q2	1.77	1.93	2.13	1.89	-0.36	-0.56	.58
	Q3	2.09	2.04	2.38	2.00	-0.28	-0.43	.67
	Q4	1.32	1.81	1.94	2.21	-0.62	-0.95	.35
	Q5	1.64	1.92	2.13	1.96	-0.49	-0.77	.45
	Q6	0.05	0.21	0.06	0.25	-0.01	-0.23	.82
	Q7	0.55	0.51	0.63	0.50	-0.08	-0.48	.64
	Q8	0.36	0.95	0.31	0.60	0.05	0.19	.85
Overall	Q1	1.32	1.48	1.32	1.34	0.00	-0.01	.99
	Q2	1.50	1.81	1.93	1.84	-0.43	-1.23	.22
	Q3	1.76	1.76	1.97	1.61	-0.21	-0.64	.52
	Q4	1.08	1.59	2.12	1.96	-1.04	-3.06	.00
	Q5	1.72	2.06	1.92	1.91	-0.20	-0.51	.61

	Q6	0.32	1.20	0.42	0.93	-0.10	-0.51	.61
	Q7	0.68	1.11	0.39	0.59	0.29	1.74	.09
	Q8	0.68	1.19	0.56	1.10	0.12	0.55	.58

5. Discussion

In this study, the scores of most students from the rural region of Penang fall under grade E for the problem-solving test involving measurement formulae and HOTS regardless of gender in the NS, CNTS and TNTS as well as in the overall sample. In other words, majority of the students did not even achieve a minimum proficiency level in solving the word problems involving measurement formulae and HOTS. This finding is parallel to the finding of the study conducted by Caponera and Losito (2016) in which the Malaysian students from low socio-economic status schools commonly found in rural region only scored 373 on average in TIMSS 2011 because most of them were unable to apply their mathematical knowledge to solve problems even in straight forward situations (Mullis, Martin, Foy, & Arora, 2012).

The relatively poor mathematical performance of rural students could be due to the socio-economic factor (Cheema & Kitsantas, 2013; Gilleece, Cosgrove, & Sofroniou, 2010) at both the student and school levels (Khairani, 2016; Thien, 2016). Since the study was conducted in rural region, most of the students might come from low socio-economic families. With the disadvantages in terms of socio-economic status, they had limited access to the educational resources which could promote their higher-order thinking skills (Noor Ibrahim, Mohd Ayub & Md Yunus, 2020). At the school level, Marwan, Sumintono, and Mislan (2012) reported that the teacher quality was another concern in rural schools, besides lacking educational resources and infrastructures. Due to the limited educational resources, the teachers in rural schools had less exposure to the teaching approach for promoting HOTS (Noor Ibrahim et al., 2020). Thus, some of them emphasized more on the memorisation of facts and concepts in mathematics classroom rather than teaching mathematical problem solving (Palraj et al., 2017). Consequently, students were only able to answer the mathematics questions involving cognitive domains such as remembering and understanding, but not the word problems which involved higher-order thinking skills.

The unsatisfactory performance of the rural Grade Five students in solving the word problems involving measurement formulae and HOTS could also be affected by the individual-level factors such as poor metacognitive skills (Abdullah et al., 2017; Verschaffel, Schukajlow, Star, & Van Dooren, 2020), poor basic arithmetic skills (Pongsakdi et al., 2020; Voyer, 2011), poor conceptual understanding, as well as lack of motivation (Bates & Wiest, 2004; Fadlelmula, Cakiroglu, & Sungur, 2015). In fact, poor conceptual understanding and metacognitive skills might prevent them from using their prior mathematical knowledge to plan strategies to solve the word problems involving HOTS (Abdullah et al., 2017; Verschaffel et al., 2020). As a result, they might fail to formulate the correct mathematical sentences to solve the word problems. Besides, poor arithmetic skills might hinder the students from performing the correct algorithms required to obtain the answers (Pongsakdi, et al., 2020, Voyer, 2011). Due to lack of motivation in reading the lengthy word problems, students might choose not to attempt the word problems at all (Nor, Ismail, & Yusof, 2016; Shamsuddin, 2016). These factors might further affect the rural Grade Five students' performance in solving the word problems involving measurement formulae and HOTS.

Although the students' performance in solving the word problems involving measurement formulae and HOTS was poor, there was no significant difference between the performance of the Grade Five girls and boys in the NS, CNTS and TNTS, and in the overall sample. In fact, the mean score of the girls was slightly higher than the mean score of the boys in each type of rural school and in the overall sample. The findings of the study contradicted with the findings of the study conducted by Forgasz and Hill (2013) as well as Innabi and Dodeen (2018) in the western context and middle east context, respectively, but they were consistent with the findings of the study conducted by Pang and Seah (2020) in the Asian context. In the Malaysian context, the findings were supported by Abedalaziz (2011), Kashefi et al. (2017) and Thien (2016). The findings of these studies highlighted that there was no significant difference among the performance of rural male and female elementary students (Abedalaziz, 2011; Kashefi et al., 2017) and secondary students (Thien, 2016) in solving word problems. According to Else-Quest, Hyde, and Linn (2010), the inconsistency of findings on gender difference in various contexts might be due to cultural variations in the opportunity structures for the female students. In Malaysia, the education system greatly emphasises gender equality (MOE, 2012). Both male and female Malaysian students are given equal educational opportunity. In the context of this study, the Grade Five students in the rural region of Penang are given equal exposure to the word problems involving measurement formulae and HOTS regardless of gender after they have learned the topic. Thus, their performance was found to be almost equal in this study.

To provide a more comprehensive understanding of gender differences of the rural Penang Grade Five students in solving the word problems involving measurement formulae and HOTS, the data were further analysed at the item level. In general, there was no significant difference in the mean item scores between the Grade Five boys and Grade Five girls except for item Q4. The findings of the overall sample highlighted that the Grade Five girls performed significantly better than the Grade Five boys in answering item Q4 which involved the perimeter and area of a rectangle. This item required a high cognitive demand in which the students had to analyse the requirements of the word problem, evaluate whether the measurement formula can be directly applied to solve the word problem, and synthesise the strategy to solve the word problem. In other words, solving item Q4 involved the mathematical processes such as interpreting, employing the mathematical knowledge, and formulating the correct mathematical sentences (Thien, 2016). It might be possible that with a better acquisition of skills in these mathematical processes as compared to their counterparts, the Grade Five girls in rural region of Penang performed remarkably better on item Q4.

6. Conclusion

In conclusion, this study provides empirical evidence regarding the rural Penang Grade Five students' performance and the gender differences in solving word problems involving measurement formulae and HOTS. Even though no gender difference was found in general, the findings of this study highlighted the poor performance of Grade Five boys and girls from the rural region of Penang in solving the word problems involving measurement formulae and HOTS. These findings reveal a dire need to inform teachers and educators to shift the focus of mathematics teaching from procedural skill fluency to mastery of reasoning skills as well as problem solving skills (Jader et al., 2020) to promote rural students' higher-order thinking skills. The findings of this study also inform the need of interventions which could promote rural students' ability in solving word problems involving measurement formulae and HOTS. Thus, future studies are suggested to develop such interventions to support teachers in enhancing rural students' performance in solving word problems involving measurement formulae and HOTS.

Acknowledgement

This study was made possible with funding from the Fundamental Research Grant Scheme (FRGS) 203.PGURU.6711744 of the Malaysian Ministry of Higher Education (MOHE). The authors would like to thank all the teachers and Year Five pupils who voluntarily participated in this study.

References

[1]. Abdullah, A.H., Rahman, S. N. S., & Hamzah, M.H. (2017). Metacognitive skills of Malaysian students in non-routine mathematical problem solving. *Bolema*, *31*(57), 310–322.

[2]. Abedalaziz, N. (2011). Gender-related differences of Malaysian students in their solution processes of solving mathematical problems. OIDA *International Journal of Sustainable Development*, *2*, 11-36.

[3]. Anderson, L. W., Krathwohl, D. R. Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman

[4]. Bates, E.T., & Wiest, L. R. (2004). Impact of personalization of mathematical word problems on student performance. *The Mathematics Educator*, *14*(2), 17–26.

[5]. Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-scale Assessments in Education*, *4*(12), 1-22.

[6]. Cheema, J. R., & Kitsantas, A. (2013). Influences of disciplinary classroom climate on high school student self-efficacy. *International Journal of Science and Mathematics Education*, *12*, 1261–1279. doi: 10.1007/s10763-013-9454-4

[7]. Chew, C.M., Wun, T.Y., Lim, H.L., & Wong, S.E. (2016). A multi-strand test for assessing Year 4 pupils' proficiency in area formulae. *Open Journal of Social Sciences, 4*(2), 14-19. doi: 10.4236/jss.2016.42003.

[8]. Clements, D. H., & Battista, M. T. (2001). Length, perimeter, area and volume. In L. S. Grinstein & S. I. Lipsey (Eds). *Encyclopedia of mathematics education* (pp. 406- 410. New York, NY: Routledge

[9]. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Routledge Academic.

[10]. Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*(4), 344-377.

[11]. Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *6*(348), 1-13. doi: 10.3389/fpsyg.2015.00348

[12]. Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103-127.
[13]. Fadlelmula, F. K., Cakiroglu, E., & Sungur, S. (2015). Developing a structural model on the relationship among motivational beliefs, self-regulated learning strategies, and achievement in mathematics. *International Journal of Science and Mathematics Education*, *13*, 1355–1375. doi: 10.1007/s10763-013-9499-4
[14]. Fatmanissa, N., & Kusnandi, K. (2017). The linguistic challenges of mathematics word problems: A research and literature review. *Malaysian Journal of Learning and Instruction, Special issue on Graduate Students Research on Education*, 73-92.
[15]. Forgasz, H. J., & Hill, J. C. (2013). Factors implicated in high mathematics achievement. *International Journal of Science and Mathematics Education*, *11*, 481–499. doi: 10.1007/s10763-012-9348-x
[16]. Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210-240.
[17]. Fuchs, L., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (2020). Addressing the role of working memory in mathematical word-problem solving when designing intervention for struggling learners. *ZDM*, *52*(1), 87-96.
[18]. Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Upper Saddle River, NJ: Merrill.
[19]. Ghasemi, E., & Burley, H. (2019). Gender, affect, and math: A cross-national meta-analysis of Trends in International Mathematics and Science Study 2015 outcomes. *Large-scale Assessments in Education*, *7*(10), 1-25.
[20]. Gilleece, L., Cosgrove, J., & Sofroniou, N. (2010). Equity in mathematics and science outcomes: Characteristics associated with high and low achievement on PISA 2006 in Ireland. *International Journal of Science and Mathematics Education*, *8*, 475–496. doi: 10.1007/s10763-010-9199-2
[21]. Innabi, H., & Dodeen, H. (2018). Gender differences in mathematics achievement in Jordan: A differential item functioning analysis of the 2015 TIMSS. *School Science and Mathematics*, *118*(3), 127-137.
[22]. Jader, J., Lithner, J., & Sidenvall, J. (2020). Mathematical problem solving in textbooks from twelve countries. *International Journal of Mathematical Education in Science and Technology*, *51*(7), 1120-1136.
[23]. Kashefi, H., Yusof, Y. M., Ismail, Z., Men, O.L., Lee, T.J., & Joo, T.K., (November, 2017). *Gender and mathematics performance of primary students in higher order thinking skills*. Paper presented at 2017 7th World Engineering Education Forum (WEEF), Kuala Lumpur. doi: 10.1109/WEEF.2017.8467086
[24]. Khairani, A. Z. (August, 2016). *Assessing urban and rural teachers' competencies in STEM integrated education in Malaysia*. Paper presented at MATEC Web of Conferences (ENCON 2016). doi: 10.1051/MATECCONF/20178704004
[25]. Kilpatrick, J. (2001). Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics*, *47*(1), 101-116.
[26]. Kloosterman, P. (1992). Non-routine word problems: One part of a problem-solving program in the elementary school. *School Science and Mathematics*, *92*(1), 31-37.
[27]. Leder, G. C., & Forgasz, H. J. (2018). Measuring who counts: Gender and mathematics assessment. *ZDM*, *50*(4), 687-697.
[28]. Leder, G. C. (2019). Gender and mathematics education: An overview. In G. Kaiser, N. C. Presmeg (Eds.), *Compendium for early career researchers in mathematics education (*pp. 289-308). Cham: Springer.
[29]. Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education*, *10*, 1163–1190. doi: 10.1007/s10763-011-9325-9
[30]. Machaba, F. M. (2016). The concepts of area and perimeter: Insights and misconceptions of Grade 10 learners. *Pythagoras*, *37*(1), 1-11. doi: 10.4102/pythagoras. v37i1.304
[31]. Malaysian Examination Syndicate. (2016). *Gred sekolah rendah* [Primary school grade]. https://sapsnkra.moe.gov.my/dokumen/GRED.pdf
[32]. Malaysian Ministry of Education. [MOE] (2012). Preliminary report Malaysia education blueprint 2013-2025. Retrieved from http://www.moe.gov.my/userfiles/file/PPP/Preliminary-Blueprint-Eng.pdf.
[33]. Malaysian Ministry of Education. [MOE] (2017). *Kurikulum standard sekolah rendah: Standard kurikulum dan pentaksiran (semakan) Matematik Tahun 4* [Primary school standard curriculum: Year 4 Mathematics curriculum and assessment standard document (revised)]. Putrajaya: Malaysian Ministry of Education.

[34]. Marwan, A., Sumintono, B., & Mislan, N. (2012). Revitalizing rural schools: A challenge for Malaysia. *Educational Issues, Research and Policies, 1*, 172-188.

[35]. Meinck, S., & Brese, F. (2019). Trends in gender gaps: Using 20 years of evidence from TIMSS. *Large-scale Assessments in Education*, *7*(1), 1-23.

[36]. Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMMS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

[37]. Mullis, I. V. S., Martin, M.O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics.* Retrieved from Boston College, TIMSS & PIRLS International Study CenterRetrieved from http://timssandpirls.bc.edu/timss2015/international-results/

[38]. Noor Ibrahim, N., Mohd Ayub, A. F., & Md Yunus, A. S. (2020). Impact of Higher Order Thinking Skills (HOTS) module based on the Cognitive Apprenticeship Model (CAM) on student's performance. *International Journal of Learning, Teaching and Educational Research*, *19*(7), 246-262.

[39]. Nor, N.A.K., Ismail, Z., & Yusof, Y. (2016). The relationship between emotional intelligence and mathematical competency among secondary school students. *Journal on Mathematics Education*, *7*(2), 91-100.

[40]. O'Connor, B. R., & Norton, S. (2020). Supporting indigenous primary students' success in problem-solving: Learning from Newman interviews. *Mathematics Education Research Journal*, 1-24.

[41]. OECD. (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school* Retrieved from https://www.oecd.org/pisa/pisap roduc ts/42843 625.pdf.

[42]. Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using SPSS.* Maidenhead: McGraw-Hill.

[43]. Palraj, S., DeWitt, D., & Alias, N. (2017). Teachers beliefs in problem solving in rural Malaysian secondary schools. *Malaysian Online Journal of Educational Technology*, *5*(4), 45-57.

[44]. Pang, J., & Seah, W. T. (2020). Excellent mathematical performance despite "negative" affect of students in Korea: The values perspective. *ECNU Review of Education*. doi: 10.1177/2096531120930726

[45]. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497.

[46]. Polya, G. (1945). How to solve it. New Jersey: Princeton University.

[47]. Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*, 33–44. doi: 10.1007/s11858-019-01118-9

[48]. Reilly, D., Neumann, D. L., & Andrews, G. (2017). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*. doi: 10.1007/s11165- 017-9630-6.

[49]. Reiss, K., & Torner, G. (2007). Problem solving in the mathematics classroom: The German perspective. *ZDM*, *39*(5-6), 431-441.

[50]. Salemeh, Z., & Etchells, M. J. (2016). A case study: Sources of difficulties in solving word problems in an international private school. *Electronic International Journal of Education, Arts, and Science*, *2*, 149-163.

[51]. Saw, G. (2015). Patterns and trends in achievement gaps in Malaysian secondary schools (1999–2011): Gender, ethnicity, and socioeconomic status. *Educational Research for Policy and Practice*, *15*(1), 41-54. doi: 10.1007/s10671-015-9175-2.

[52]. Scheibling-Seve, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics*, 1-19.

[53]. Schoenfeld, A. H. (1985). Making sense of "out loud" problem-solving protocols. *The Journal of Mathematical Behavior*, *4*(2), 171-191.

[54]. Shamsuddin. S (2016). *Mimi'S garden of reflections: Transpiring a teacher's journey to inspire.* Singapore: Partridge Publishing Singapore.

[55]. Spada, M. M., Georgiou, G. A., & Wells, A. (2010). The relationship among metacognitions, attentional control, and state anxiety. *Cognitive Behaviour Therapy*, *39*(1), 64-71.

[56]. Thien, L.M. (2016). Malaysian students' performance in mathematics literacy in PISA from gender and socioeconomic status perspective. *The Asia-Pacific Education Researcher*, *25*, 657-666.

[57]. UNESCO. (2014). UNESCO education strategy 2014–2021. France: UNESCO

[58]. Verschaffel, L., Greer, B., & DeCorte, E. (2000). *Making sense of word problems*. Netherlands: Swets and Zietlinger Publishers.

[59]. Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, 1-16.

[60]. Verschaffel, L., Van Dooren, W., Greer, B., & Mukhopadhyay, S. (2010). Reconceptualising word problems as exercises in mathematical modelling. *Journal für Mathematik-Didaktik*, *31*(1), 9-29.

[61]. Voyer, D. (2011). Performance in mathematical problem solving as a function of comprehension and arithmetic skills. *International Journal of Science and Mathematics Education*, *9*, 1073–1092. doi: 10.1007/s10763-010-9239-y

[62]. Weldeana, H. N. (2015). Gender positions and high school students' attainment in local geometry. *International Journal of Science and Mathematics Education*, *13*(6), 1331-1354.

[63]. White, A. L. (2009). A revaluation of Newman's error analysis. In D. Martin, T. Fitzpatrick, R. Hunting, D. Itter, C. Lenard, T. Mills, L. Milne (Eds), *Mathematics of prime importance, Proceedings of 46th Annual Conference of the Mathematics Association of Victoria* (pp. 249-257). Brunswick, Melbourne: M.A.V.