International Journal of Evaluation and Research in Education (IJERE)
Vol. 10, No. 3, September 2021, pp. 956~965
ISSN: 2252-8822, DOI: 10.11591/ijere.v10i3.21439

Development and validation of year five geometrical measurement skills instrument

Siti Nur Annisa Mohd Nasser, Lim Hooi Lian
School of Educational Studies, Universiti Sains Malaysia, Malaysia

Article Info

Article history:

Received Jan 6, 2021
Revised Jun 18, 2021
Accepted Jul 13, 2021

Keywords:

Geometrical measurement skills
Numerical skills
Rasch analysis
SOLO Taxonomy
Spatial skills

ABSTRACT

Geometrical measurement is one of the most difficult fields in primary school mathematics and regularly found to be an area of weaknesses. The factors affecting the low understanding is due to the lack of valid and reliable assessment instruments. Thus, this study aimed to develop and validate a Geometrical Measurement Skill Instrument (GMSI) to assess Year Five National school pupils' geometrical measurement skills in geometrical measurement. GMSI was developed by applying the Structure of Observed Learning Outcomes (SOLO) Taxonomy and was constructed in the super item format which consisted of 24 items altogether. This study applied a survey approach to assess 132 Year Five pupils' geometrical measurement skills. The content validity was examined using the content validity index (CVI) analysis. For the construct validity, data were obtained and analyzed using the Rasch analysis. From the CVI analysis, the results showed that GMSI could be used for the pilot study. Results for the construct validity indicated that GMSI fulfilled the psychometric properties and is valid and reliable. Hence, the results could help teachers and pupils to diagnose the strength and weaknesses of geometrical measurement and help them to plan systematic remedy to improve teaching and learning.

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

Corresponding Author:

Siti Nur Annisa Mohd Nasser
School of Educational Studies
Universiti Sains Malaysia
11800 Gelugor, Penang, Malaysia
Email: sitinurannisa@student.usm.my

1. INTRODUCTION

Geometrical measurement has a specific and special role in almost every mathematics curriculum. Geometrical measurement can be defined as perimeter, area, and volume measurements in one, two and three dimensions, respectively [1], [2]. Besides, understanding of geometrical measurement in both conceptual areas and skills is very important, since it forms the basis for the study of other subjects such as physics, chemistry, biology, geology and geography, as well as art and architecture [1]. This is why geometrical measurement has been introduced since pre-school, developed and extended across primary and secondary schools [3]. However, previous studies have shown that geometrical measurement is one of the most difficult fields in mathematics especially in the primary level [4]. Not only that, studies have also shown that pupils have poor understanding of geometrical measurement due to the typical method of learning geometrical measurement, i.e. by memorising and applying step-by-step procedures without addressing the underlying concept [5]-[7]. This can be proved by a study conducted [8] with pupils in Indonesia, in which pupils claimed that reliance on complicated formulas makes the subject seem difficult and boring because they are unable to develop and explore the underlying concepts themselves. Besides, the instructional approach in major international assessment such as Trends in International Mathematics and Science Study (TIMSS) and National Assessment for Educational Progress (NAEP) has always been formula-centered without giving adequate attention to pupils' understanding of the underlying ideas of geometrical measurement. The portrayed issues happen not only globally, but also locally. In Malaysia, geometrical measurement is not only being assessed in the public examination such as Primary School Achievement Test (UPSR), School- Based Form Three Assessment (PT3) and Malaysia Certificate of Education (SPM), but also in the international examination such as the Trends in International Mathematics and Science Study (TIMSS) [9]. However, from the UPSR Answer Quality issued by Malaysian Examination Syndicate, pupils still have problems in understanding the underlying concept of geometrical measurement [10]-[14]. More worryingly, not only primary school pupils, but also secondary school students faced problems in understanding geometrical measurement. This can be proved by the analysis of TIMSS 2015 which stated that the ranked of Malaysia is still below the average score and is categorised as a low international benchmark [9]. Furthermore, according to the study, Malaysia was ranked low in geometrical measurement as compared to South Korea, Japan, Australia, the United States, Singapore, and Thailand.

In view of its important stance in mathematics, science and our lives, pupils should completely understand not only how to solve geometrical measurement problems numerically based on formula, but also the underlying concept of geometrical measurement. [15]. For example, according to Clements [16], pupils still have difficulties in understanding how the geometrical measurement formula works although they could use and calculate the formula correctly. One of the key reasons for this problem is the lack of assessment of the skills associated with conceptual understanding of geometrical measurement [4], [5], [17]. In order to grasp the conceptual basis of geometrical measurement, pupils need to have strong geometrical measurement skills, as geometrical measurement skills combine both spatial and numerical skills [18], [19]. Several past studies have shown that spatial and numerical skills facilitate pupils' learning in a variety of topics, especially geometrical measurement [20]. Besides, previous studies [21]–[23] stated that there were many problems in geometrical measurement due to a lack of link between spatial and numerical skills where pupils did not understand the underlying structure behind geometrical measurement formulas, e.g. area=length x width. However, mathematics researchers and educators have suggested that the inflexibility of pupils in managing geometrical measurement problems may result from the curriculum and the assessment of school mathematics [24]. Besides, there is inadequate number of assessment instrument developed to assess pupils' geometrical measurement skills based on spatial skills and also numerical skills [4]. Therefore, it is important to develop an assessment tool covering a wide range of geometrical measurement skills including the specific types of constructs i.e., spatial and numerical skills to better inform the teaching and learning in geometrical measurement. Due to the importance of geometrical measurement skills in assessing the performance of geometrical measurement, this study aims to develop a new assessment instrument i.e., Geometrical Measurement Skill Instrument that is theoretically grounded to determine the level of geometrical measurement skills of the pupils in geometrical measurement.

Several preliminary researches on the development of geometrical measurement skills instrument have been conducted. Vasilyeva, *et al.* [4] conducted a study to develop geometrical measurement skills instrument to evaluate primary pupils performance in geometrical measurement. However, the instrument was constructed in the multiple-choice format. Using the multiple-choice format, pupils are required to respond by selecting only one of the justifications given in the choices, limiting their ability to provide useful insight into their understanding of geometrical measurement. Therefore, in this study, geometrical measurement skills instrument is developed in the super item format i.e., the open-ended format based on Structure of Observed Learning Outcomes (SOLO) Taxonomy that focused on different levels of difficulties. Moreover, the literature did not explicitly document that existing geometrical measurement skills applied SOLO Taxonomy as the framework of the instrument development.

2. LITERATURE REVIEW

Even though the researchers in the field of measurement have concluded that geometrical measurement skills depend on both spatial and numerical skills, the relationship between these cognitive skills and geometrical measurement skills has not been explicitly assessed. In fact, hardly any empirical research has been conducted to explore the factors influencing geometrical measurement skills [25]. Besides, Crites, *et al.* [5] mentioned in their study that the routine assessment and instruction approach has always been formula-centered, which may potentially hinder the development of geometrical measurement concepts and skills. Moreover, Hwang, *et al.* [26] stated in their study that there was a lack of studies done to determine pupils' level of geometrical measurement skills and their weaknesses in geometrical measurement, and this issue caused problems for teachers to determine the level and weaknesses of the pupils.

It is therefore important to design an assessment tool i.e., the newly developed assessment instrument that can provide benefits not only to pupils but also teachers and researchers. By developing the assessment instrument, pupils' level, strength and weaknesses of geometrical measurement skills can be identified and this can help guide pupils in their learning process. Besides, the instrument might provide teachers with valuable perspectives and guideline to identify pupils' geometrical measurement skills and error in geometrical measurement. Furthermore, from the assessment instrument, the researchers could gain a larger picture of geometrical measurement skills and recognize constructs that generate the most difficulties for pupils.

Thus, this study aims to develop Geometrical Measurement Skill Instrument (GMSI) to assess Year Five pupils' geometrical measurement skills in geometrical measurement. The content domain of the instrument in this study is geometrical measurement consisting of perimeter, area and volume. Furthermore, in order to assess the geometrical measurement skills in this study, the integration of spatial and numerical skills was used as the constructs. Based on the literature, the subskills of spatial skills that are related to geometrical measurement assessment are spatial visualization and spatial structuring skills [15], [16], [27]-[29]. Thus, in this study, the definition for spatial skills for geometrical measurement is the integration of spatial visualisation and spatial structuring skill to mentally structure one, two and three-dimensional space from a single unit of measurement to a composite unit and to conceptualize the formula for perimeter, area and volume. Whereas, the definition for numerical skills for geometrical measurement is the ability to integrate counting, computing and applying formula skill to solve formula-based problems involving perimeter, area and volume using numerical calculations.

In this study, the model called SOLO Taxonomy which was developed by Biggs and Collis [30] was used as the framework to develop the assessment instrument. This is because SOLO Taxonomy plays an important role that values the balance of surface and deep levels of items [31]. The SOLO Taxonomy encompasses five key hierarchical levels that represent the quality of learning for a specific task i.e., pre-structural, unistructural, multistructural, relational and extended abstract. However, in this study only four levels of SOLO Taxonomy (unistructural, multistructural, relational and extended abstract) are used because the first level which is the prestructural level represents a response that is irrelevant or misses the point. Based on SOLO Taxonomy, there are two major categories, where each containing two increasingly complex stages: surface and deep (Surface=unistructural and multistructural; Deep=relational and extended abstract).

In the unistructural level, pupils need to use only one relevant information in the stem by applying spatial or numerical skills to solve perimeter, area and volume questions. While, in the multistructural level pupils need use several relevant information in the stem by applying spatial or numerical skills to solve perimeter, area and volume questions. Next, in the relational level pupils should be able to integrate all relevant information in the stem to generalize or create a clear structure or meaning of perimeter, area and volume problems by applying spatial or numerical skills. Lastly, in the extended abstract level pupils should have the ability to apply spatial and numerical skills to generalize the perimeter, area and volume problems into more abstract situation that leads to a new topic or area. Figure 1 summarizes the process of the development of GMSI.

Thus, the objective of the present study is to develop an alternative assessment instrument to assess Year Five National School pupils' geometrical measurement skills in geometrical measurement and validate the assessment instrument in terms of content validity and construct validity. Besides, this study aims to identify pupils' level of geometrical measurement skills based on the level of SOLO Taxonomy level of understanding i.e., unistructural, multistructural, relational and extended abstract. The last objective of this study is to identify the pupils' level of spatial skills and numerical skills in geometrical measurement.

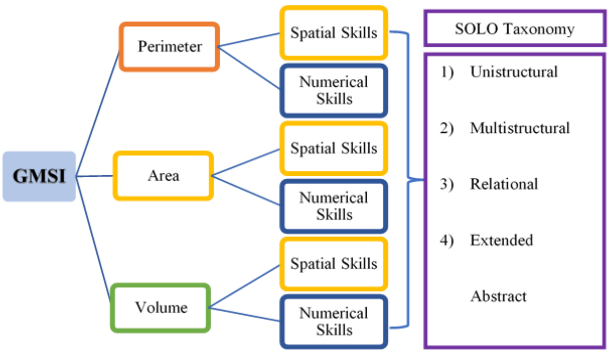

Figure 1. Process of the development of GMSI

3. RESEARCH METHOD

3.1. Research design

This study aimed to develop assessment to assess year five pupils' geometrical measurement skills in geometrical measurement. The research design implemented in this study is the cross-sectional survey design. A cross-sectional survey design allows data to be gathered from selected individuals at one time. Besides, a cross-sectional design provide valuable insight into current behaviour, attitudes and beliefs in a population [32]. In addition, the survey design is chosen because it can generate a large amount of data at a relatively low cost in a short time. Therefore, researchers can set a limited timeframe for a project that can help to plan and deliver results [33].

3.2. Instrumentation

In this study, the instrument development process introduced by Miller and Lovler [34] was used to develop the instrument. The selection of this model is due to the effectiveness of its application in previous research, where all the instruments developed using the steps proposed in this model obtained high validity and reliability [35]–[37]. This instrument development process consists of ten steps namely defining the instrument universe, audience, and purpose, developing an instrument plan, composing the instrument items, writing the administration instructions, conducting the pilot test, conducting the item analysis, revising the instrument, validate the instrument, developing norms and compiling the test manual. However, since the main purpose of this analysis is to develop and validate the instrument, the ninth and tenth steps i.e., developing norms and compiling the test manual in the test development process will not be included in this study [34].

The instrument in this study consisted of six super items where there were four items in each super item, giving a total of 24 items altogether. All the super items were open-ended questions. Two super items were developed for each content domain (perimeter, area and volume) involving spatial skills and numerical skills.

3.3. Scoring procedure

The scoring procedure for the instrument was based on the principles of Rasch modelling and the principles of measurement, in which scores are given in an orderly manner, such that each increase in score reflects an improvement in the ability of the pupils [38]. Thus, in this study, the ordered value of 0, 1, 2, 3 are given to the items as: 0=totally wrong, 1=partially correct, 2=almost completely correct, and 3=completely correct.

3.4. Participants

In Malaysia, geometrical measurement i.e., perimeter, area and volume are being taught to Year Four pupils (10 years old). However, this study involves Year Five pupils (11 years old) because of pupils' previous knowledge and the syllabus of the instrument involved the Year Five syllabus i.e., the perimeter, area and volume of the composite figure, considering that the composite figure items were mostly evaluated in UPSR. Thus, the participants of this study consist of 132 Year Five pupils from four National Schools in the Northeast district, Penang.

3.5. Research procedure

The first important procedure after composing a newly developed instrument is to obtain expert panel reviews to identify the content validity. Thus, in this study, six experts are appointed to evaluate the instrument consisting of three professional experts and three field experts. Experts in this study include a lecturer from the Institute of Teacher Education with psychometric background, School Improvement Specialist Coaches in Teaching and Mathematics, as well as teachers of Mathematics with more than ten years of experience. The content validation is conducted by face-to-face approach. After the analysis of the content validity, the test instrument is administered to the participants (pilot test) to determine the construct validity. The data collected from the sample are then analysed by the quantitative item analysis.

3.6. Data analysis

In this study, the content validity is represented by the Content Validity Index (CVI). The CVI is described to be the best practice to quantify the content validity of an instrument based on several recent studies [39]–[43]. The CVI analysis was done based on the six steps suggested by Yusoff [43]: 1) Preparing content validation form; 2) Selecting a review panel of experts; 3) Conducting content validation; 4) Reviewing domain and items; 5) Providing score on each item; and 6) Calculating CVI.

For the construct validity, the data collected from 132 Year Five pupils in the pilot test were analyzed by the application of partial credit Rasch model using the WINSTEPS Version 3.72.3. Construct validity ensures that the developed instrument measures correctly what it intended to measure. In Rasch analysis, the instrument is considered valid and reliable if the data fulfill the psychometric properties as: 1) Item dimensionality; 2) Item dependency; 3) Item fit; and 4) Item polarity. Thus, all the above-mentioned analysis is conducted and reported in order to determine the validity and reliability of the instrument.

4. RESULTS AND DISCUSSION

4.1. Content validity

From the content validity analysis, the CVI value for item (I-CVI) and the CVI value for scale (S-CVI) were calculated. From the calculation, the I-CVI value is 1 and the S-CVI/Ave is 1. The findings concluded that the instrument of this analysis had extremely good content validity because the I-CVI value met the criterion in which the minimum value for I-CVI was 0.78 for 6 to 10 experts [40]. In addition, the S-CVI/Ave value also fulfils the satisfactory level at which, according to Polit and Beck [44], the appropriate value for S-CVI/Ave must be more than 0.90. From the results, it can be concluded that all 24 items in the research instrument have been developed with good operationalization and conceptualization which can be used for the pilot study.

4.2. Construct validity

Data collected from the test results were keyed in the IBM SPSS Version 26 software analyzed by applying the partial credit Rasch model using the WINSTEPS Version 3.72.3 software. The results of the analysis are presented as:

4.2.1. Item dimensionality

One of the most important aspect in Rasch analysis is to ensure that all the items are related to the same latent variable and point in the same direction [38]. To accomplish this, all the items in the instrument i.e., the GMSI have to satisfy the unidimensionality requirements to make sure that GMSI measure the same latent variable i.e., geometrical measurement skill. Thus, the Principal Component Analysis (PCA) of the Rasch residuals was performed and the results are as shown in Table 1. From the analysis, the variance explained by measures is 62.4% and closely match to the expected which is 63.5%. The largest secondary dimension, i.e., the 1st contrast in the residual explained 3.5% of the variance. Based on the general rules for PCA suggested by Linacre [45], GMSI is considered to be unidimensional considering that the value of explained by measures is larger than 60% and the value of unexplained variance in the 1st contrast is 3.5% and the value is in the acceptable range i.e., less than 15% as suggested by Fisher [46], Conrad, *et al.* [47]. Besides, the eigenvalue for unexplained variance in the 1st contrast is 2.3 which is less than 5 and this confirmed that there is no secondary dimension for GMSI [48].

Table 1. Standardized residual variance (in eigenvalue units)

		Empirical		Modelled
Total raw variance in observations	63.8	100.0%		100.0%
Raw variance explained by measures	39.8	62.4%		63.5%
Raw variance explained by persons	16.4	25.7%		26.2%
Raw Variance explained by items	23.4	36.7%		37.4%
Raw unexplained variance (total)	24.0	37.6%	100.0%	36.5%
Unexplained variance in 1st contrast	2.3	3.5%	9.4%	

4.2.2. Item dependency

Item dependency is one of the most important requirements of Rasch analysis in which all the items in GMSI must exhibit local independence. To satisfy the Rasch analysis requirement for item dependency, the results of the largest standardized residual correlation were analyzed as presented in Table 2. The results show that items that are more than 0.7 are highly correlated and are considered as locally dependent or redundant. A negative correlation explains the opposite of local dependence [49]. Based on the results, the correlation of all the item pairs is not more than 0.7 and it can be concluded that the items in GMSI are not locally dependent.

Table 2. Largest standardized residual correlations used to identify dependent item

Correlation	Entry number item	Entry number item
0.57	4 SSPEA	8 SSAU
0.43	1 SSPU	5 SSAEA
0.38	17 NSAU	21 NSVU
0.32	10 SSVM	11 SSAEA
0.32	18 NSAM	22 SSVR
0.29	3 SSPR	7 NSVM
0.29	2 SSPM	15 SSVR
-0.35	1 SSPU	12 SSVEA
-0.30	5 SSAU	12 NSVM
-0.29	10 SSVM	17 NSVU

4.2.3. Item fit

The fit analysis of the items in Rasch Model is to ensure the quality of a measurement instrument [50]. It helps to determine the misfit items so that researchers could refine the test instrument. In order to identify the fit items, the value of mean-square fit statistic (MNSQ) must ranges between 0.6 to 1.4 [51] and the value of the normalized and standardized infit and outfit (ZSTD) must ranges between -2.0 and 2.0 [38]. Table 3 shows the results of the fit analysis. From the results, all of the items are in the acceptable range except for items SSPEA, NSPEA, SSVEA and SSAEA. However, the misfit items are retained because high mean-squares are much greater threat to validity compared to low mean-squares [52]. Besides, according to Linacre [52], if the value of mean-squares are less than 0.5, the items are less productive for measurement, but not degrading. Besides, all the misfit items fulfill the requirements of the ZSTD values and also have the positive value of PT-Measure Correlation.

Table 3. Item statistic

Item	Measure	Infit		Outfit		PT-MEASURE
		MNSQ	ZSTD	MNSQ	ZSTD	CORR
SSPR	1.67	1.39	1.4	0.75	0.1	0.30
NSPR	-0.09	1.29	2.3	1.38	1.0	0.53
SSPU	-2.98	0.98	0.0	1.35	0.9	0.35
NSPU	-1.03	1.06	0.7	1.30	2.0	0.41
SSVM	-0.42	1.22	2.6	1.24	1.7	0.30
SSAR	0.97	1.20	1.2	0.99	0.3	0.44
NSAR	0.71	1.19	1.4	0.91	-0.1	0.49
SSPM	-2.10	0.97	-0.2	1.18	0.8	0.43
NSPM	-1.61	1.15	1.0	1.12	0.4	0.56
SSPEA*	2.42	1.07	0.3	0.22*	-0.4	0.25
SSVU	-1.41	1.05	0.5	1.01	0.1	0.43
SSAM	-0.43	1.02	0.3	1.01	0.1	0.45
SSAU	-2.27	0.93	-0.5	0.99	0.1	0.46
NSAEA	1.89	0.94	-0.1	0.85	-0.1	0.40
NSAM	-1.28	0.93	-0.4	0.74	-0.5	0.65
NSVEA	1.72	0.93	-0.2	0.91	0.0	0.44
NSVR	1.30	0.92	-0.3	0.42	-0.7	0.48
NSVM	-1.07	0.91	-0.7	0.69	-0.9	0.67
NSAU	-1.32	0.90	-1.0	0.82	-1.1	0.54
NSPEA*	2.37	0.89	-0.3	0.35*	-1.4	0.43
NSVU	-2.59	0.87	-0.8	0.67	-1.1	0.51
SSVEA*	2.41	0.85	-0.2	0.24*	-0.5	0.30
SSAEA*	2.39	0.77	-0.5	0.18*	-0.8	0.33
SSVR	0.75	0.75	-1.7	0.49	-0.7	0.60
Mean	0.00	1.01	0.2	0.83	0.0	
S. D	1.73	0.16	1.0	0.36	0.8	

4.2.4. Item polarity

Another important aspect in identifying the unidimensionality of the instrument is the item polarity. Item polarity can be determined by looking at the results of PT-MEASURE correlation [38]. The value of PT-MEASURE correlation must be positive to indicate that all the items are moving in the same direction. Whereas, negative PT-MEASURE correlation means that the items are problematic and not consistent with the construct [38]. Based on the results, the value of PT-MEASURE correlation for all items are positive as shown in Table 3, and it can be concluded that GMSI moves in the same direction to measure the construct.

4.2.5. Person and item reliability and separation indices

Table 4 shows the person and item reliability and separation indices. From the analysis, the person reliability which is similar to the value of Cronbach alpha (α) is 0.83 and according to Bond and Fox [38], GMSI is considered to have a good and acceptable level of consistency because it has the value greater than 0.7. Furthermore, the value of person separation index is 2.2 and this value is acceptable because a good separation index value is to be greater than 2.0 [53]. Apart from that, the result of the item reliability of 0.98 demonstrate that GMSI has high item reliability because the value is greater than 0.8 [38]. Besides, the value of 7.89 for the item separation is acceptable because the value is greater than 2.0 [53]. The result shows that GMSI has the ability to classify the items in various difficulties.

Table 4. Person-item reliability and separation indices

	Separation	Reliability
Person	2.20	0.83
Item	7.89	0.98

4.3. Discussion

GMSI was developed for the purpose to assess Year Five pupils' geometrical measurement skills in geometrical measurement. Apart from determining the pupils' measurement skills performance, this instrument helps pupils to grasp the conceptual understanding in geometrical measurement by integrating spatial skills and numerical skills in the instrument. This is due to several past studies had stated that spatial skills and numerical skills do predict geometrical measurement skills [5], [25], [54]. In addition, by integrating spatial skills and numerical skills, pupils and teachers could recognize which constructs lead to geometrical measurement limitations, and this could help them improve teaching and learning.

To ensure that the items developed measure the constructs of geometrical measurement skills in geometrical measurement, content validity and construct validity were executed. The CVI analysis was used for examining the content validity and the results of the CVI analysis is found to be valid in terms of its content [44]. Besides, in order to determine the construct validity, the partial credit Rasch model was applied. There are five assumptions that need to be accomplished to confirm the construct validity which are unidimensionality, local independence, item fit, item polarity and person-item reliability and separation indices. Based on the findings, all the items in GMSI have a good unidimensionality [45] and were locally independent [38]. Besides, all the items were consider fit due to all the items have high reliability and fulfilled the requirements for the ZSTD values and also have the positive PT-MEASURE correlation [52]. Furthermore, from the findings, GMSI have a good and acceptable person reliability and high item reliability [38]. Moreover, according to Linacre [53], GMSI acquired a good person separation index that could classify person's ability and also a good separation index for item that has the ability to classify items in various difficulties. Therefore, it can be concluded that GMSI had fulfill all the assumptions of the Rasch model i.e., the psychometric properties and has the ability to measure geometrical measurement skills in geometrical measurement among Year Five pupils.

Besides, by looking at the Person- Item Distribution Map (PIDM) in Rasch Analysis as shown in Figure 2, the ability of the pupils and item difficulty can be viewed clearer to better inform the correlation between the person and the item on the same logit scale. According to the PIDM, pupils find it more challenging to solve spatial skills items than numerical skills items since three of the spatial skills items, SSAEA, SSPEA, and SSVEA, are located at the highest value of the logit scale. This explains why majority of the pupils acquire lack of conceptual understanding of geometrical measurement as mentioned by Clements and Sarama [55] where spatial skills plays an imperative role for pupils to grasp the underlying concepts of geometrical measurement. Furthermore, as the items of GMSI were developed according to the level of SOLO Taxonomy, it can be seen that, the difficulty of the items were aligned with the SOLO Taxonomy level of understanding i.e., unistructural, multistructural, relational and extendented Abstract except for items SSVU, NSPU and NSAU. Based on PIDM, majority of the pupils were at the multistructural level. This finding indicates that the pupil's ability in geometrical measurement skills is still at the surface level of understanding, as according to Hattie and Brown [31], the two SOLO Taxonomy levels of understanding i.e., unistructural and multistructural are categorized as surface level of understanding.

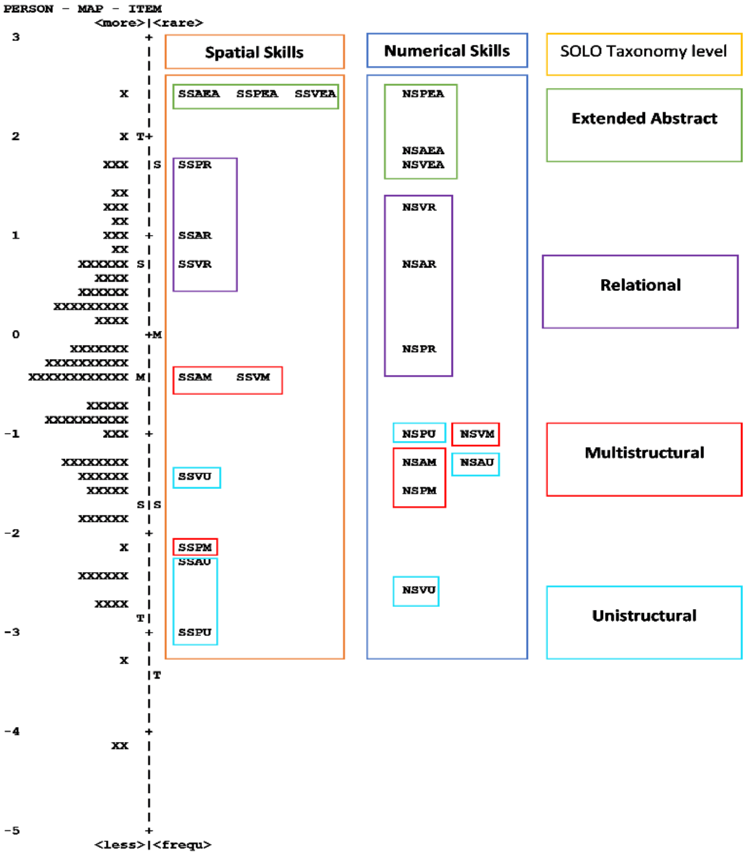

Figure 2. Person-item distribution map

5. CONCLUSION

This study aimed to examine content validity and construct validity of a newly developed instrument i.e., the GMSI. From the result of CVI analysis and Rasch analysis, it can be concluded that GMSI has an extremely good content validity and also fulfilled the psychometric properties for the construct validity. Therefore, GMSI is found to be a valid and reliable instrument to assess pupils' geometrical measurement skills in geometrical measurement. They are several implications of this this study. First, the results of this study provide a new insight, that is, a new assessment approach to pupils, teachers and also the education system in teaching and learning geometrical measurement. Besides, the results of this study help pupils and teachers to diagnose the strength and weaknesses in geometrical measurement and this could help teachers to plan accurate and systematic remedy to improve teaching and learning in geometrical measurement. Furthermore, the valid and reliable instrument could be used as a template and guidance for teachers and pupils in various geometric topics. Perhaps, GMSI could be a reference to other researchers who would like to further their research in other fields of study.

However, this study has some limitations. First, although GMSI has been determined to be valid and reliable, future study could be carried out to improve the test, such as adding more items and conducting the instrument online with feedback given to pupils. Besides that, this study only included primary pupils, and the syllabus was limited to that of a primary school. As a result, there is a need for research to extend the established instrument for secondary students that covers the secondary syllabus. Since the aim of this study was to develop and validate a new assessment instrument, the GMSI, future research on gender differences in geometrical measurement skills, spatial skills, and numerical skills could be conducted. Furthermore, future research on error analysis should be performed in order to identify the errors and misconceptions of the pupils in geometrical measurement

ACKNOWLEDGEMENTS

The authors would like to express their utmost gratitude to Education Sponsorship Division, Ministry of Education Malaysia and also School of Educational Studies, University Sains Malaysia for making this research possible.

REFERENCES

[1] K. Fuson, D. Clements, and S. B. Kazez, *Geometry, Spatial Reasoning, and Measurement*. National Council of Teachers of Mathematics, 2010.
[2] E. M. Kim, J. Haberstroh, S. Peters, H. Howell, and L. Nabors Oláh, "A Learning Progression for Geometrical Measurement in One, Two, and Three Dimensions," *ETS Res. Rep. Ser.*, vol. 2017, no. 1, pp. 1-26, 2017, doi: 10.1002/ets2.12189.
[3] L. Outhred and D. McPhail, "A framework for teaching early measurement," In J. Bana, A. Chapman, Eds., *Mathematics education beyond 2000*, Perth: Mathematics Education Research Group of Australasia Incorporated, 2000, pp. 487-494.
[4] M. Vasilyeva, L. H. Ludlow, B. M. Casey, and C. S. Onge, "Examination of the psychometric properties of the measurement skills assessment," *Educ. Psychol. Meas.*, vol. 69, no. 1, pp. 106-130, 2009, doi: 10.1177/0013164408318774.
[5] T. Crites, B. J. Dougherty, H. Slovin, and K. Karp, "Geometric Measurement," in *Putting Essential Understanding into Practice: Geometry, 6-8*, National Council of Teachers of Mathematics, 2018, pp. 11–34.
[6] I. V. Mullis, M. O. Martin, E. J. Gonzalez, and S. J. Chrostowski, *TIMSS 2003 International Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2004.
[7] W. Wartono, J. R. Batlolona, and A. Putirulan, "Cognitive Conflict Strategy and Simulation Practicum to Overcome Student Misconception on Light Topics," *J. Educ. Learn.*, vol. 12, no. 4, pp. 747-756, 2018, doi: 10.11591/edulearn.v12i4.10433.
[8] T. Laurens, F. A. Batlolona, J. R. Batlolona, and M. Leasa, "How does realistic mathematics education (RME) improve students' mathematics cognitive achievement?" *Eurasia J. Math. Sci. Technol. Educ.*, vol. 14, no. 2, pp. 569-578, 2018, doi: 10.12973/ejmste/76959.
[9] S. Y. Chog, T. J. Wong, and A. Halim, "Pencapaian Matematik TIMSS 1999, 2003, 2007, 2011 dan 2015: Di Mana Kedudukan Malaysia Dalam Kalangan Negara Asia Tenggara?" *Malaysian J. High. Order Think. Ski. Educ.*, vol. 2, no. 23, pp. 54-108, 2018.
[10] Malaysian Examination Syndicate, "Kupasan Mutu Jawapan UPSR Matematik 015/025/035," Putrajaya, 2017.
[11] Malaysian Examination Syndicate, "Kupasan Mutu Jawapan UPSR Matematik 015/025/035," Putrajaya, 2013.
[12] Malaysian Examination Syndicate, "Kupasan Mutu Jawapan UPSR Matematik 015/025/035," Putrajaya, 2014.
[13] Malaysian Examination Syndicate, "Kupasan Mutu Jawapan UPSR Matematik 015/025/035," Putrajaya, 2012.
[14] Malaysian Examination Syndicate, "Kupasan Mutu Jawapan UPSR Matematik 015/025/035," Putrajaya, 2010.
[15] G. Tan Sisman and M. Aksu, "A study on sixth grade students' misconceptions and errors in spatial measurement: Length, area, and volume," *Int. J. Sci. Math. Educ.*, vol. 14, no. 7, pp. 1293-1319, 2016, doi: 10.1007/s10763-015-9642-5.
[16] D. Clements, *Learning and Teaching Measurement (2003 Yearbook)*. National Council of Teachers of Mathematics, Dec. 2003.
[17] J. Hannighofer, M. van den Heuvel-Panhuizen, S. Weirich, and A. Robitzsch, "Revealing German primary school students' achievement in measurement," *ZDM - Int. J. Math. Educ.*, vol. 43, no. 5, pp. 651-665, 2011, doi: 10.1007/s11858-011-0357-y.
[18] D. H. Clements and M. T. Battista, "Geometry and Spatial Reasoning," in *Handbook of research on mathematics teaching and learning*. New York: Macmillan, 1992, pp. 420-464.
[19] M. Vasilyeva, B. M. Casey, E. Dearing, and C. M. Ganley, "Measurement skills in low-income elementary school students: Exploring the nature of gender differences," *Cogn. Instr.*, vol. 27, no. 4, pp. 401-428, 2009, doi: 10.1080/07370000903221809.
[20] V. Tůmová and N. Vondrová, "Links between Success in Non-measurement and Calculation Tasks in Area and Volume Measurement and Pupils' Problems," *Sci. Educ.*, vol. 8, no. 2, pp. 100-129, 2017.
[21] M. T. Battista, "The Development of Geometric and Spatial Thinking," In F. K. Lester, Ed., *Second Handbook of Research on Mathematics Teaching and Learning*. Charlotte, NC: Information Age, 2007.
[22] H.-M. E. Huang and K. G. Witz, "Children's Conceptions of Area Measurement and Their Strategies for Solving Area Measurement Problems," *J. Curric. Teach.*, vol. 2, no. 1, pp. 10-26, 2012, doi: 10.5430/jct.v2n1p10.
[23] M. Battista, *Cognition-based assessment and teaching of geometric measurement*. *Building on Students' Reasoning*. Heinemann, 2012.
[24] H.-M. E. Huang and K. G. Witz, "Developing children's conceptual understanding of area measurement: A curriculum and teaching experiment," *Learn. Instr.*, vol. 21, no. 1, pp. 1-13, 2011.
[25] B. M. Casey, E. Dearing, M. Vasilyeva, C. M. Ganley, and M. Tine, "Spatial and numerical predictors of measurement performance: The moderating effects of community income and gender," *J. Educ. Psychol.*, vol. 103, no. 2, pp. 296-311, 2011, doi: 10.1037/a0022516.
[26] W. Y. Hwang, S. W. D. Purba, Y. F. Liu, Y. Y. Zhang, and N. S. Chen, "An investigation of the effects of measuring authentic contexts on geometry learning achievement," *IEEE Trans. Learn. Technol.*, vol. 12, no. 3, pp. 291-302, 2019, doi: 10.1109/TLT.2018.2853750.
[27] M. T. Battista, M. L. Winer, and L. M. Frazee, "How spatial reasoning and numerical reasoning are related in geometric measurement," *International Group for the Psychology of Mathematics Education*, 2017, pp. 355-362.
[28] N. S. Newcombe, J. L. Booth, and E. A. Gunderson, "Spatial Skills, Reasoning, and Mathematics," in *The Cambridge Handbook of Cognition and Education*. Cambridge University Press, 2019, pp. 100-123, doi: 10.1017/9781108235631.006.

[29] R. Lehrer, L. Jaslow, and C. L. Curtis, “Developing an understanding of measurement in the elementary grades,” in D. H. Clements and G. W. Bright, Eds., *Learning and teaching measurement*. Reston, Va: National Council of Teachers of Mathematics, 2003, pp. 100-121.
[30] J. Biggs and K. F. Collis, *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. Academic Press, 1982.
[31] J. Hattie and G. T. L. Brown, “Cognitive processes in asTTle: The SOLO taxonomy,” asTTle Technical Report #43, University of Auckland/Ministry of Education, 2004, [Online]. Available: http://tinyurl.com/j2epdkq.
[32] L. R. Gay, G. E. Mills, and P. AirAsian, “Survey Research,” in *Educational Research Competencies for Analysis and Applications*, 9th ed. Pearson Education, Inc, 2012, pp. 175–191.
[33] K. Kelley, B. Clark, V. Brown, and J. Sitzia, “Good practice in the conduct and reporting of survey research,” *Int. J. Qual. Heal. Care*, vol. 15, no. 3, pp. 261-266, 2003.
[34] L. A. Miller and R. L. Lovler, *Foundations of psychological testing: A practical approach*. Carlifonia: Sage Publications, 2018.
[35] M. E. E. M. Matore, “Pembinaan Instrumen Kecerdasan Menghadapi Cabaran (Ikbar) Bagi Pelajar Politeknik Menggunakan Model Rasch,” PhD Thesis, Universiti Sains Malaysia, 2015.
[36] A. L. Adibah, “Pembanglinan Instrumen Penilaian Akhlak Pelajar di Institusi Pengajian Ttnggi Awam,” PhD Thesis, Universiti Teknologi Malaysia, 2013.
[37] H. Azman, “Instrumen Penilaian Pembimbing dalam Pelaksanaan Pembelajaran Berasaskan Kerja Pelajar di Industri,” PhD Thesis, Universiti Teknologi Malaysia, 2012.
[38] T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental Measurement in the Human Sciences*, 3rd ed. New York: Routledge, 2015.
[39] J. S. Grant and L. L. Davis, “Selection and use of content experts for instrument development,” *Res. Nurs. Health*, vol. 20, no. 3, pp. 269-274, 1997, doi: 10.1002/(sici)1098-240x(199706)20:3<269::aid-nur9>3.3.co;2-3.
[40] Mary R. Lynn, “Determination and quantificatiom of content validity,” *Journal of Experimental Psychology: General*, vol. 136, no. 1. pp. 382-386, 1986, doi: 10.1161/CIRCULATIONAHA.112.092098.
[41] M. F. M. Marzuki, N. A. Yaacob, and N. M. Yaacob, “Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps,” *J. Med. Internet Res.*, vol. 5, no. 2, pp. 1-7, 2018, doi: 10.2196/10308.
[42] M. M. Ozair, K. A. Baharuddin, S. A. Mohamed, W. Esa, and M. S. B. Yusoff, “Development and validation of the knowledge and clinical reasoning of acute asthma management in emergency department (K-CRAMED),” *Educ. Med. J.*, vol. 9, no. 2, pp. 1-17, 2017, doi: 10.21315/eimj2017.9.2.1.
[43] M. S. B. Yusoff, “ABC of content validation and content validity index calculation,” *Educ. Med. J.*, vol. 11, no. 2, pp. 49-54, 2019, doi: 10.21315/eimj2019.11.2.6.
[44] D. F. Polit and C. T. Beck, “The content validity index: Are you sure you know what’s being reported? Critique and recommendations,” *Res. Nurs. Health*, vol. 29, no. 5, pp. 489-497, 2006, doi: 10.1002/nur.20147.
[45] J. M. Linacre, “Data variance explained by Rasch measures,” Rasch Measurement Transactions, vol. 20, no. 1, p. 1045, 2006.
[46] W. P. Fisher, “Rating scale instrument quality criteria,” Rasch Measurement Transactions, vol. 21, no. 1, p. 1095, 2007.
[47] K. Conrad, K. Conrad, M. Dennis, B. Riley, and R. Funk, “Validation of the Behavioral Complexity Scale (BCS) to the Rasch Measurement Model,” GAIN Methods Report 1.2, 2011.
[48] J. M. Linacre, *Winsteps Rasch measurement computer program User’s Guide.* Beaverton, Oregon, 2020.
[49] J. M. Linacre, “Help for Winsteps Rasch Measurement and Rasch Analysis Software:Largest residual correlations for items,” 2012. [Online]. Available: https://www.winsteps.com/winman/table23_99.htm.
[50] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch analysis in the human sciences*. Netherlands: Springer, 2014.
[51] J. M. Linacre, “Reasonable mean-square fit values,” Rasch Measurement Transactions, vol. 8, no. 3, p. 370, 1994.
[52] J. M. Linacre, “Fit diagnosis: Infit outfit mean-square standardized,” Help for Winsteps Rasch Measurement and Rasch Analysis Software, 2014.
[53] J. M. Linacre, “Reliability and separation of measures,” 2020. [Online]. Available: https://www.winsteps.com/winman/reliability.htm.
[54] K. Cipora, P. A. Schroeder, and H.-C. Nuerk, “On the multitude of mathematics skills: Spatial-numerical associations and geometry skill?” in *Visualizing Mathematics. Research in Mathematics Education*. Springer, Cham, 2018, pp. 361–370.
[55] D. Clements and J. Sarama, *Learning and Teaching Early Math: The Learning Trajectories Approach*. New York: Routledge, 2014.