Journal of Namibian Studies, 34 S2(2023): 171–187 ISSN: 2197-5523 (online)

Rasch Modelling in the Mathematical Reflective Thinking Scale for 21st Century Filipino Senior High School Learners

Dynah D. Soriano, Ph.D.[1]

[1] Don Honorio Ventura State University, Bacolor, Philippines

Abstract

Introduction: In the absence of a standardized tool that measures the Mathematical reflective thinking of Filipino Senior High School (SHS) learners in meeting the requirements of 21st century learning, this study proposes a scale which covers the entire problem-solving process, self-reflection, mastery, feedback, and realization. These factors indicate areas on how reflection is activated based on students' actions. This study is helpful in measuring and controlling the degree of possessing the reflective thinking skill and in the subject of planning research studies for development. This Mathematical Reflective Thinking Scale (MRTC) assists the teachers in determining on which step of the problem solving do their students experience difficulty on, what type of reflective habits they have and evaluating within the framework of scale dimensions, whether or not students possess any self-reflection, mastery, reasoning, questioning, evaluation, realization, and feedback habits and in designing instructional strategies.

Objectives: The study analyses the psychometric properties of the MRTC for Filipino SHS students employing the Rasch Model to provide a baseline reference in prescribing standardized tools in meeting 21st century learning.

Methods: The study employed descriptive-survey using the partial-credit Rasch Model analysis to evaluate the reliability and validity of MRTS for Filipino SHS. The participants of the study were 768 students coming from 47 SHS from the three major islands of the Philippines which are Luzon, Visayas and Mindanao, and they were selected using stratified-random sampling.

Results: Using a partial-credit Rasch modelling (Winsteps), the findings revealed that the new instrument was determined to be largely psychometrically sound, although two problematic items were also noted as reasons for further refinement. The study also evaluated the significance of several parameters as determined by the MRTS. items in contributing to students' reflective thinking in dealing with Mathematics problem-solving.

Conclusions: Finally, the proposed standardized tool is expected to contribute to the work of various stakeholders in Mathematics education in achieving quality basic education in the Philippines through the development of critical thinking and problem solving as reflective thinking skills to learners which facilitate reflective learning.

Keywords: reflective thinking, problem-solving, Senior High School, partial-credit, Rasch modelling, Philippines

Introduction

Educators agree that it is important for both teachers and learners to reflect on what students must learn and eventually achieve. In a constructivist approach, the purpose of teaching is a perspective on learning which emphasizes on how students actively develop knowledge out of their experiences (Mcleod, 2019). Accordingly, the role of reflective thinking is significant in the enhancement of higher order thinking that leads to excellent problem-solving skills. In addition, Kızılkaya and Aşkar (2009) emphasized that investigating, assessing, and measuring reflective thinking is a crucial skill.

Reflective thinking is a talent that can help with thinking development, along with problem-solving abilities. Reflective thinking can inspire someone to tackle an issue by following several problem-solving paths in addition to being a skill that matches up with many higher thinking abilities like critical thinking and problem solving (Can, 2015). In this regard, it is clear that problem solving skills and reflective thinking are closely associated with each other. Though there is a great deal of research conducted on reflective thinking skills, this research mostly focuses on the Mathematical reflective thinking skills of participants which serves as a basis for proposing a standard scale in complying with the requirements of 21st century learning.

In 2009, Kizilkaya and Aşkar developed an instrument on the Reflective Thinking Ability Scale towards Problem Solving whilst De Leon and Prudente (2018) ventured on developing a new reflective thinking scale that focuses on areas that determine the development of students' reflective thinking, particularly for the SHS students. However, the aforementioned questionnaires for Filipino SHS students' Reflective Thinking Scale and Reflective Thinking Ability Scale Towards Problem Solving is not yet verified to test the internal consistency of the constructs using a more complex test of validity and reliability like Rasch Modelling analysis.

Over the years, researchers have used several methods to investigate reflection. These are based on various theories. Finding out the validity and reliability of assessment instruments in exploring the internal consistency and factor structure for measuring reflection. Some of the recommendations included research on the development of teaching materials to develop high school students' reflective thinking skills regarding the development of a reflective thinking test instrument in Mathematics for high school students (Hendriana et al., 2019; Muntazhimah, 2019; and Nindiasari et al., 2016). The need to measure mathematical reflective thinking of learners is not well established in the literature and there is no standard test that will gauge learners' reflective thinking in Mathematics that will serve as their groundwork for 21st century education.

It is interesting to note that, reflective thinking is all about critical thinking and together with problem-solving skills, it is considered as one of the requirements of the twenty-first century education among educators and learners. It is of paramount importance that educators and learners should be equipped with vast knowledge and skills as they engage in the contemporary needs of the times. These requirements should be manifested and realized from the teachers' preparation down to the students' learning assessment.

Undoubtedly, the 21st century demands a lot from teachers and students. However, these demands do not hinder the educators' undying passion for their chosen profession. These demands make teachers more heroic among other professions as they think of ways on how they can serve as bridges to help their students cross and reach the summit of success.

At present, assessment of learning should be reviewed and restructured to be relevant to the students' 21st century skills' requirements. Teachers play a vital role in educating students for them to utilize their 21st century skills in the teaching-learning process. Students may apply these abilities in their future practice as well as to improve their knowledge based on the 21st century needs. Mathematics students must be provided with the opportunity to develop their skills by engaging with authentic learning tasks, solving complex real-word problems, integrating technology into their instruction and assessments, while fostering cooperative learning in the classroom.

From this point, Filipino teachers may evaluate their students' level of reflective thinking by using a more sophisticated test. This may guide teachers in planning their lessons and in preparing appropriate assessment tools for students' performance. Since the weight of learning in a remote environment is cast on the shoulders of the students, it is necessary for them to assess their own level of reflective thinking to enhance their critical thinking and problem-solving skills.

From several previous studies mentioned above, research references on evaluating MRTS for SHS students are limited. Also, most research carried out employed classical test theory instead of a more sophisticated test such as the Rasch Model in establishing the validity and reliability to come up with a standard tool.

The utilization of Rasch modelling analysis in this study gauges the usefulness of survey instruments in testing the validity and reliability of the psychometric properties that identifies the Mathematical reflective thinking of the learners. The Rasch model is a powerful tool for the analysis and refinement of survey and test instruments especially with regards to increasing reliability and validity (Boone, 2016).

To measure student performance that is based on Outcomes-Based Education (OBE) by measuring students' Reflective Thinking in Mathematics, the researcher of this study hopes to produce an output that would help both Filipino instructors and students in the SHS. The SHS students' Reflective Thinking Scale and Reflective Thinking Ability Scale towards Problem Solving is combined and evaluated using the Rasch modelling analysis to gauge the Mathematical Reflective Thinking Scale intended for Filipino SHS students. It is a thinking process involving activities of reflecting ideas, problems, or information received or the process of interpretation that starts from one experience to the next by making a deeper relationship understanding and connecting other experiences or ideas.

In the absence of a standardized tool that measures the Mathematical reflective thinking of Filipino SHS learners in meeting the requirements of 21st century learning, this study proposes a scale which covers the entire problem-solving process, self-reflection, mastery, feedback, and realization. These factors indicate areas on how reflection is activated based on students’ actions. This study is helpful in measuring and controlling the degree of possessing the reflective thinking skill and in the subject of planning research studies for development. This MRTS assists the teachers in determining on which step of the problem solving do their students experience difficulty on, what type of reflective habits they have and evaluating within the framework of scale dimensions, whether or not students possess any self-reflection, mastery, reasoning, questioning, evaluation, realization, and feedback habits and in designing instructional strategies.

Objectives

The purpose of the study is to analyse the psychometric properties and goodness of fit of the proposed Mathematical Reflective thinking scale for Filipino SHS students in establishing its validity and reliability as a standardized tool to measure the mathematics competence of the 21st century learner.

Methods

This descriptive-survey study covers an analysis on the psychometric properties of the Mathematical Reflective thinking scale for Filipino SHS students employing the Rasch Model to evaluate the reliability and validity of the Reflective Thinking Scale for Filipino SHS students developed by De Leon and Prudente (2018) and Reflective Thinking Ability Scale towards Problem Solving developed by Kızılkaya and Aşkar (2009).

The utilization of the Rasch analysis model provides a baseline reference in prescribing standardized tools to gauge the Mathematical Reflective thinking of Filipino SHS students in order to assess if students are equipped with 21st century skills. The Rasch methodologies were utilized to record, assess, and offer a method for the survey's optimization. Also, to provide a context when explaining survey results.

In accordance with the aims of this study in standardizing an instrument for Filipino SHS learners, students from different major island groups were sampled. Seven hundred sixty-eight (768) students representing forty-seven (47) SHS from Luzon, Visayas, and Mindanao from twenty-three (23) private schools and twenty-four (24) public schools recognized by the Department of Education and offering at least four strands made up the study's respondents. The sample was selected via stratified random selection, and it was made up of students, ranging in age from 15 to 19, from various class levels and strands.

A survey questionnaire was given to participants through a variety of sources including the social media, Google forms, and referrals upon the grant of permission from concerned school administrators and provided informed consent forms to research participants. A 4-point Likert scale on Reflective thinking with 15 items developed by De Leon and Prudente (2018), and a 5-point Likert scale on the Reflective Thinking Ability towards Problem Solving with 14 items developed by Kızılkaya and Aşkar (2009) was used. The students were asked to rate each item based on their agreement measuring their self-reflection, mastery, feedback, realization, questioning, reasoning and evaluation on students' reflective thinking in Mathematics.

By comparing the fit of the rating scale on the aspects of unidimensionality, item and person separation, internal consistency value of Cronbach's alpha, rating scale effectiveness as well as item and person fit, and providing the Wright item-person map for determining whether the items are appropriate for the respondents, partial-credit Rasch modelling was performed using the WinStep software.

Results

The Mathematical Reflective Thinking Scale instrument validity and reliability analysis was conducted on the aspects of unidimensionality, item and person level of difficulty, as well as item and respondent fit, rating scale effectiveness, construct key map, and the Wright item-person map.

Table 1. Standardized Residual Variance

INPUT: 768 PERSONS 29 ITEMS MEASURED: 768 PERSONS 29 ITEMS 129 CATS 3.66.0

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --			Modeled
Total raw variance in observations	=	42.3	100.0%		100.0%
Raw variance explained by measures	=	13.3	31.4%		31.9%
Raw variance explained by persons	=	10.1	24.0%		24.4%
Raw Variance explained by items	=	3.1	7.4%		7.5%
Raw unexplained variance (total)	=	29.0	68.6%	100.0%	68.1%
Unexplned variance in 1st contrast	=	3.3	7.9%	11.5%	
Unexplned variance in 2nd contrast	=	2.3	5.5%	8.1%	
Unexplned variance in 3rd contrast	=	1.7	4.0%	5.9%	
Unexplned variance in 4th contrast	=	1.5	3.7%	5.3%	
Unexplned variance in 5th contrast	=	1.3	3.2%	4.6%	

Based on Table 1, the value for the raw variance explained by measurements is 31.4% explained by the Rasch dimension. Sumintono and Widhiarso (2015) define a number higher than 20% as "acceptable" and claim that this value provides proof of the Mathematical Reflective Thinking Scale's unidimensionality, meaning that the scale unquestionably measured the concept of statistical reasoning. The greatest secondary dimension, indicated by the first comparison under Winsteps, explains just 7.9% of the raw variance with an eigenvalue of 3.3, the strength of at most three items. As a result, there is a roughly 3.3:1 ratio between the variation explained by items (7.4%) and the variance explained by the second greatest dimension (7.9%). Other than that, the 1st to 5th contrast's unexplained variance is less than 8%, which is within the ideal range of less than 15%. It is also true that almost all survey datasets have several dimensions (few datasets are completely unidimensional), albeit to varied degrees. A conceivable secondary dimension would have at most three items (Royal & Gonzalez, 2016). The study comes to the conclusion that the unidimensionality assumption is reasonably satisfied for a unidimensional Rasch analysis given the data supporting a single, primary underlying construct being measured by the Rasch dimension (Linacre, 2018, pp. 557-558; Royal, Gilliland, & Kernick, 2014).

Based on the Rasch analysis in WINSTEPS, Table 2 displays the values for person reliability, item reliability, person separation, item separation, and Cronbach's alpha for the Mathematical Reflective Thinking Scale instrument. The Mathematics Reflective Thinking Scale discriminates the sample into sufficient levels, yielding a value for person reliability of 0.86 and a person separation value of 2.52. When the value of person reliability is higher than 0.80, according to Sumintono and Widhiarso (2015), and Bond and Fox (2007), the respondent is providing a "good" and consistent response. According to Linacre (2003), a good separation value of item difficulty is appropriate if the person separation value is higher than 2.00. The value of 2.52 is "good" for the person separation.

Table 2. The Value for Person Reliability, Item Reliability, Person Separation, Item Separation and Cronbach's Alpha Value of the Mathematical Reflective Thinking Scale Instrument

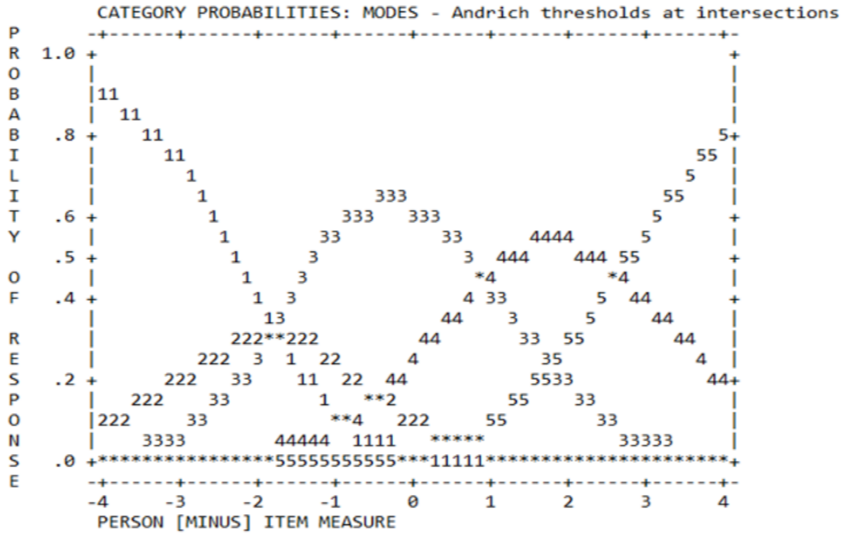
Statistics	Value	Interpretation
Cronbach's Alpha	0.85	Very high
Person Reliability	0.86	Good
Item Reliability	0.99	Excellent
Person Separation	2.52	Good
Item Separation	8.92	Good

According to Krishnan and Idris (2014), the person separation must be more than 1.00 to support measuring the students throughout the spread. The high item separation statistic suggests the student sample is large enough to confirm the item difficulty hierarchy, and the high person separation statistic suggests MRTS is sufficiently sensitive to distinguish between individual students with higher and lower levels of Reflective Thinking ability. Overall, these findings are consistent with the instrument's construct validity. In this study, the value for item reliability is 0.99 (i.e., the sample is large enough to precisely locate the items on the latent difficulty continuum) with an item separation value of 8.92. According to Bond and Fox (2007), a value less than 0.80 is less acceptable, but a value higher than 0.80 has a good value and is strongly acceptable. Regarding the item separation value, the value of 8.92 is considered high and meets Linacre's requirement (2003). According to Linacre's assertion from 2003, an item separation value greater than 2.00 is favorable. According to Krishnan and Idris (2014), the items have sufficient spread when the item separation value is greater than 1.00. High person reliability may be caused by wide ability variance. Contrarily, a high level of item reliability could be linked to a wide range in item difficulty and a sizable student population.

Also, the MRTS instrument's Cronbach's alpha score of 0.85 suggests that it has a high reliability of internal consistency (Sumintono & Widhiarso, 2015). Meanwhile, Bond and Fox (2007) claimed that the value of Cronbach's alpha (which is based on the Rasch analysis approach) that spans from 0.71 until 0.99 is acceptable as it is at the best level. This suggests that the MRTS instrument is ideal for doing genuine research.

The rating scale effectiveness, as shown in the category structure calibration in Table 3 and the response category probability curves in Figure 1, shows whether the response categories performed as intended and whether respondents were consistently and correctly able to comprehend and interpret the response categories.

Figure 1. Response category probability curves



Each category has a distinct peak, indicating that it is a meaningful endorsement choice for students at a particular mathematical reflective thinking ability level, according to the response category probability curves in Figure 1. To put it another way, students can sufficiently distinguish between distinct response options, adding to the validity.

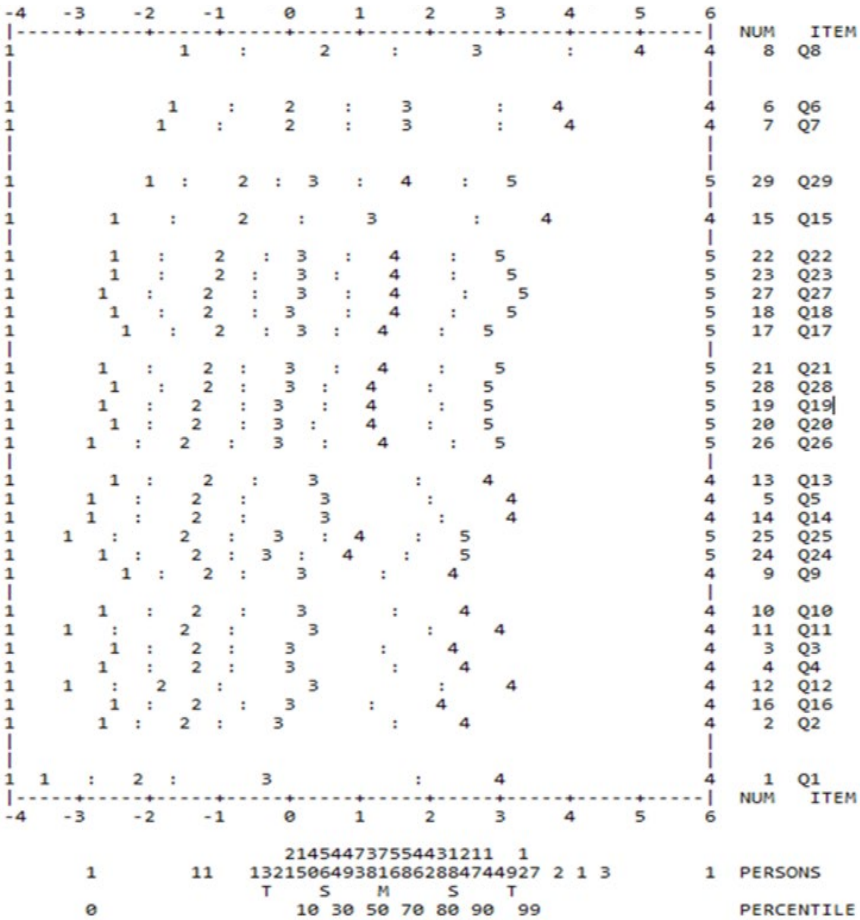
Table 3. Category Structure Calibration

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	75	4	-1.73	-2.14	1.49	1.59	NONE	(-3.26)	1
2	2	116	6	-.63	-.56	1.00	.96	-1.70	-1.84	2
3	3	710	38	.43	.44	.98	1.05	-1.85	-.31	3
4	4	688	37	1.20	1.33	1.12	2.84	.92	1.80	4
5	5	286	15	2.42	2.15	.70	.72	2.63	(3.85)	5

Next, the shape of the response category count distribution in Table 3 shows that even though students use all four of the options for reflective thinking and all five of the options for problem-solving ability, SHS students still tend to favor the options that fall on the side of agreement, particularly the "agree" and "sometimes" options. Notably, almost all infit and outfit MNSQ statistics fall within the advised range of 0.50 to 1.50 (Linacre, 2018, pp. 582-588), with only the outfit MNSQ for the categories of "strongly disagree" (for reflective thinking) and "never" (for problem-solving ability) being only 0.09 points higher than 1.50.

Figure 2. Construct key map



As anticipated, both the category measures and the Andrich threshold measures progress in steps. Finally, the construct key map for the five response types was investigated (see Figure 2). Items are ranked on the map from least endorsable (top) item Q8 (*When I do something enough times, I start to do it without thinking*) to most endorsable (bottom) item Q1) (*I think of a better way of doing things*). Obviously, throughout all 29 items, the categories are still listed in the following order: 1, 2, 3, 4, and 5. This consistency shows that none of the 29 components could possibly lead to confusion or unexpected results, proving the validity of the MRTS (Yang, Su, & Bradley, 2020). In conclusion, these findings show that the MRTS rating scale structure worked as anticipated and that research participants consistently and accurately comprehended the response options.

Item fit was used to determine whether the item is operating normally when performing the claimed measures and to evaluate the item's appropriateness. Table 4 shows that the respondents had a misconception regarding the item if the item shows misfit.

Table 4. Misfit Order of the Items in Mathematical Reflective Thinking Scale

Item	MEASURE	Outfit MNSQ (0.50-1.50)	Outfit ZSTD (-2.0-2.0)	PTMEA-CORR (0.40-0.85)	Results
6	1.00	**1.54**	**9.4**	**.29**	*Misfit*
8	1.68	**1.53**	**9.6**	**.15**	*Misfit*
7	.92	1.42	7.5	.31	Fit
29	.47	1.28	5.0	.45	Fit
17	.19	1.21	3.5	.46	Fit
15	.36	1.18	3.2	.36	Fit
18	.20	1.17	3.0	.47	Fit
16	-.47	1.12	1.8	.35	Fit
22	.26	1.11	2.1	.49	Fit
10	-.40	.93	-1.1	.45	Fit
14	-.24	.95	-.8	.46	Fit
26	-.06	.93	-1.4	.55	Fit
2	-.53	.92	-1.4	.45	Fit
19	-.01	.89	-2.1	.56	Fit
9	-.30	.92	-1.3	.47	Fit
11	-.42	.91	-1.8	.49	Fit
27	.21	.92	-1.6	.58	Fit
5	-.21	.90	-1.8	.50	Fit
4	-.45	.88	-2.1	.49	Fit
3	-.43	.87	-2.3	.49	Fit
25	-.24	.88	-2.0	.56	Fit
12	-.47	.89	-2.0	.51	Fit
13	-.20	.87	-2.4	.53	Fit
20	-.04	.86	-2.7	.57	Fit
28	.03	.89	-2.1	.58	Fit
24	-.26	.86	-2.3	.56	Fit
1	-.87	.86	-2.9	.51	Fit
21	.07	.82	-3.6	.62	Fit
23	.22	.77	-4.7	.65	Fit

Source: Boone et.al (2014) MNSQ Outfit index is 0.50-1.50, ZSTD Outfit index is -2.00-2.00, PTMEA-COOR index is 0.40-0.85

In Table 4, the outfit MNSQ statistic for items 6 and 8 is very high at 1.54 and 1.53, respectively. Because 1.5 and 1.53 < 2.00, it may be concluded that the information provided by off-variable noise for these items is not very helpful. It should be updated to address the mismatch since these components degrade measurement. More investigation may be necessary to eliminate the off-variable noise in these two items and enhance their model fit because they could be problematic. Additionally, given that each item's infit and outfit MNSQ measurements fall within the permissible range of 0.50 to 1.50, this suggests that all 27 of the remaining products are productive of measurement. Last but not least, point biserial correlations are positive, showing that the items have great discriminatory powers and that the orientation of the score on each item is compatible with the orientation of the latent variable (Linacre, 2018, pp. 526-532). Items 6 and 8 are generally beyond the range, according to the statistics. Yet, each item satisfies at least one requirement within the acceptable range. As a result, 2 components in this MRTS instrument were mismatched.

Table 5 identifies the individual (in this case, the student) whose response was most inconsistent with the Rasch analysis, or, to put it another way, whose response deviated from the estimate provided by the Rasch model. The students were arranged in order of Outfit ZSTD's highest value.

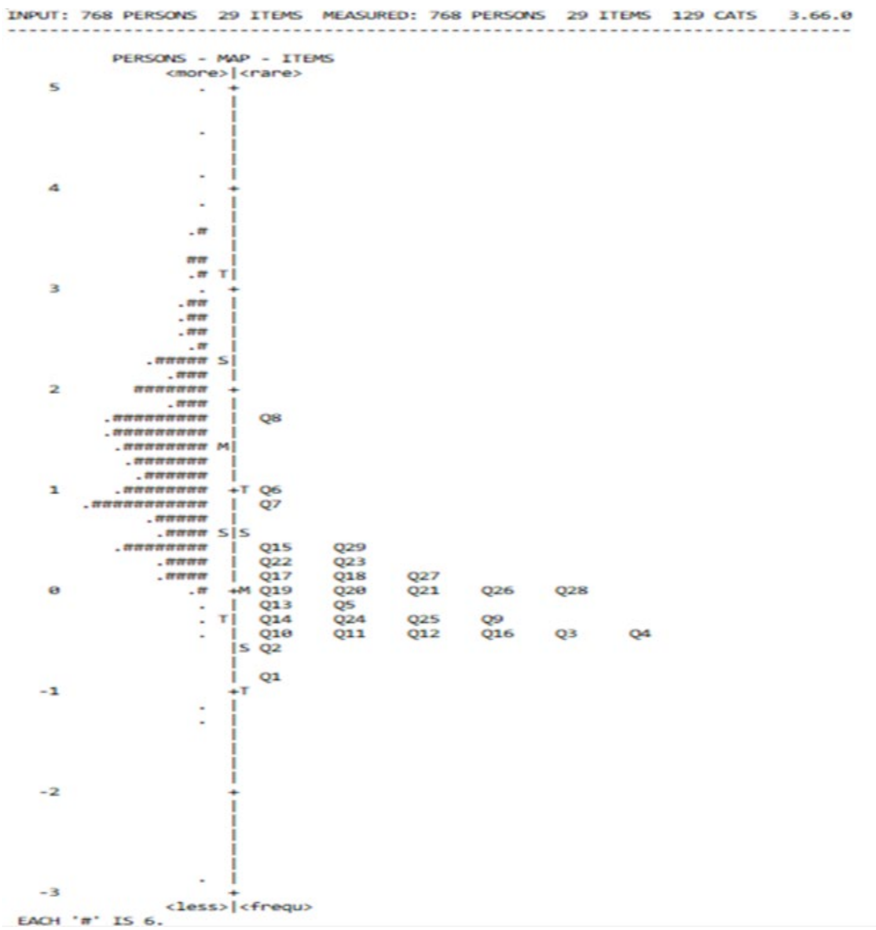
Table 5. Misfit Order of the Person in Mathematical Reflective Thinking Scale

Person	Total Score (/130)	MEASURE	Outfit MNSQ (0.50-1.50)	Outfit ZSTD (-2.0-2.0)	PTMEA-CORR (0.40-0.85)
P057	35	-2.82	7.19	5.5	-0.5
P232	97	0.86	4.07	6.6	0.15
P279	98	0.93	4	6.6	-0.12
P449	91	0.49	3.78	5.6	-0.02
P501	86	0.2	3.29	4.8	0.18
P626	84	0.08	3.24	4.7	-0.04
P631	81	-0.08	3.07	4.6	0.18
P454	90	0.43	2.93	4.3	0.25
P331	83	0.03	2.85	4.2	0.25
P619	88	0.31	2.84	4.1	0.05
P554	104	1.33	2.76	5.1	-0.05
P359	102	1.19	2.7	4.8	0.57
P654	104	1.33	2.61	4.8	0.23
P470	74	-0.44	2.59	4.2	0.13
P101	77	-0.29	2.56	4	0.26
P170	89	0.37	2.52	3.6	0.14
P649	90	0.43	2.48	3.6	0.3
P374	90	0.43	2.47	3.5	0.15
P136	113	2.04	2.38	4.2	0.19
P733	111	1.86	2.31	4.2	0.26
P696	120	2.82	2.28	3	0.15
P262	113	2.04	2.17	3.7	0.31

Twenty-two students received an Outfit ZSTD value greater than 2.0, according to Table 5. The Outfit ZSTD value for the remaining students falls within the allowed range. This shows that virtually all the students in the study (97.14%) could use the items, and the analysis done on those students produced high-quality results for the assessment using the Rasch analysis. The top four responses that didn't fit were from the four students (P279, P232, P449 & P57). Student P279's results, which included a very high total score and measure, suggest that the person most likely answered simple questions incorrectly. This was in fact the case as item Q1 ('I think of a better way of doing things') received just a "3" from student P279 even though item Q1 is an easy-to-answer item based on the Rasch analysis. Student P57, on the other hand, has a low measure but an Outfit ZSTD score higher than 2.0, which may imply that student P57 correctly answered a challenging question but erroneously for other questions. This is true since for student P57, the student scored "3" for a quite difficult item Q8 ('when I repeatedly do things, I start to do them without thinking about it'). A negative PTMEA-CORR result indicated that the respondents made typical decisions.

Figure 3 presented the Wright map which shows the distribution of persons (students) and items in a logit measurement scale. The Wright map offers helpful data on how the spread of item difficulty corresponds to the person's skill (Sumintono & Widhiarso, 2015). Wright Map, also known as an item-person map, shows the distribution of students' ability and item difficulty on a single logit scale. This enables the researcher to determine whether the items match the students' abilities (Bond & Fox, 2007).

Figure 3. Wright Map of Person and Item



Based on the item difficulty side of the Wright map, 5 items (Q19, Q20, Q21, Q26 and Q28) were calculated as being at the average of the item difficulty estimates with a value of 0.00 logit, these are more difficult to endorse than all the easy items but are easier to endorse than difficult items. Ten items (Q6, Q7, Q8, Q15, Q17, Q18, Q22, Q23, Q27, Q29) spread above the average which indicate that the items are more difficult, while fourteen items (Q1, Q2, Q3, Q4, Q5, Q9, Q10, Q11, Q12, Q13, Q14, Q16, Q24, Q25) spread below it which considered as easy items. It was discovered that item Q8, with a value of +1.68 logit, was the most challenging item, while item Q1, with a value of -.87 logit, was the one that was the simplest for the SHS students to respond to in the study. The Wright map shows that Q1 and Q2 are the items that students most readily approve, indicating that they greatly value thinking of a better way to accomplish things and enjoy learning how I handle problems. The students then readily agree with statements Q10, Q11, Q12, Q16, Q3, and Q4. This shows that students consider and seek criticism from others to better comprehend, improve their prior performance, and incorporate it into their new assignment. Students also consider what they have done to advance in their actions. Students also think that in order to figure out why they can't solve a problem, they need to ask themselves questions.

The variable map's left side displays the students' abilities. The students were found to have an average ability below the 2.0 logit, or more precisely, 1.4 logit (denoted by M in the line). In addition, two students (P275 & P700) had the highest ability, with a logit score of +3.13. Incidentally, one student breached the T lower boundary with the lowest student (P57) having recorded the result of -2.82 logit, which implies that this student held the lowest ability among the rest of the students.

Discussion

To establish a baseline reference for the prescription of standardized instruments to support 21st century learning, the study's objective is to investigate the psychometric features of the MRTS for Filipino SHS students using the Rasch Model. According to the analyses, the MRTS for Filipino SHS students corroborated the scales' general unidimensional structure.

The results closely matched the Rasch partial credit model for all scales. The primary findings indicate that, from a statistical and content standpoint, two items (under the mastery construct), item Q6 and item Q8, were not fitting within the content of their respective scales. The results suggest that items Q6 and Q8 should be eliminated from their respective scales even if they may be crucial to measure because they do not add to the measurement of the components evaluated by these scales. Overall, the MRTS for Filipino SHS students instrument has both a very high Cronbach's alpha, and item and person reliability based on the study from the Rasch Model.

Given that each of the 27 items met at least one of the fit requirements for Outfit MNSQ, Outfit ZSTD, and PTMEA-CORR, the researcher chose to keep them for purposes of validity. Also, every item has a positive PTMEA-CORR score, indicating that all of the things travel in the same direction (Bond & Fox, 2015). Moreover, the Outfit MNSQ value for 27 out of 29 items is within the permitted range, demonstrating that the items are consistent with the item measurement. As a result, it can be concluded that the MRTS for Filipino SHS students is a highly accurate and trustworthy tool for evaluating students' mathematical reflective thinking.

The MRTS for Filipino SHS students is a fine-tuned item adopted from the Reflective Thinking for Filipino (De Leon & Prudente, 2018) and Reflective thinking towards problem-solving (Kızılkaya & Aşkar, 2009) had surpassed the validity and reliability using exploratory factor analysis and confirmatory factor analysis. However, based on the Rasch analysis model, the initial 29-item was reduced to 27-items and led to 7 constructs namely self-reflection, mastery, feedback, realization, questioning, evaluation, and reasoning. This study proposes the following seven (7) constructs with twenty-seven (27) items to be the psychometric properties as standard tools in evaluating the Mathematical reflective thinking of Filipino senior high school students as shown in Table 6.

Table 6. Proposed Mathematical Reflective Thinking Scale for Filipino Senior High School Students

Items	Strongly Agree	Agree	Disagree	Strongly Disagree
Part I. Reflective Thinking for Filipino SHS Learners				
Self-reflection				
I think of a better way of doing things.				
I like to know how I do things.				
I think of what I have done so that I can improve on it further.				
I reflect on my actions to see whether I could have improved what I did.				
I reflect on what I think.				
Mastery				
I can work on things without thinking about what I am doing.				
Feedback				
I take feedback into consideration because it will help me improve on what I am doing.				

I think others' feedback is important as it will help me understand better.				
I seek feedback from others in order to enhance my tasks.				
I take into consideration my past performance and integrate it with what I am doing.				
Realization				
I correct my mistakes upon reflection.				
I immediately realize my mistakes when I reflect on what I have done.				
I reflect on what others have said.				

Part II. Reflective Thinking towards Problem Solving	Always	Often	Sometimes	Rarely	Never
Questioning					
When I cannot solve a problem, I ask questions to myself to understand why I cannot solve it.					
I try to find a better way of solving by questioning the paths followed by my peers to solve the problem.					
While solving a problem, I ask questions to myself to find different ways of solving the problem.					
When I read a problem, I think about which information I need for a solution.					
When I read a problem, I ask questions to myself to determine what is given and required					
Evaluation					
After solving a problem, I think about a better solution.					
I evaluate the possible solutions one by one to find a better solution to the next problem.					
When I have solved a problem, I go over and evaluate the operations I have performed.					
After solving a problem and finding the result, I check the operations I have performed.					
After solving a problem, I compare my solution with the solutions of my peers and evaluate my solution.					
Reasoning					
While solving a problem, I carefully think about why I perform which operation.					
While solving a problem, I think about the reasons for the operations and try to establish a connection with the result I have found.					
When I read a problem, I think about the problem I have previously solved and create connections between them based on similarities and differences.					
While solving a problem, I perform each operation by thinking previous and next stages.					

The proposed standardized instrument has two major sections: a) Reflective Thinking for Filipino SHS learners, and b) Reflective Thinking towards Problem Solving. The first section has 13-item inventory subscales to gauge agreement to factors on students' reflective thinking with four constructs namely self-reflection with 5 items, mastery with 1 item, feedback with 4 items, and realization with 3 items. The items are evaluated using a 4-point Likert scale from strongly agree to strongly disagree. The second section has 14 items measuring students' Reflective Thinking Ability towards Problem Solving with three sub-dimensions: questioning with 5 items, evaluation with 5 items and reasoning with 4 items. The students need to rate their agreement on items using a 5-point Likert scale (5 points for "Always", 4 for "Often", 3 for "Sometimes", 2 for "Rarely" and 1 for "Never").

The MRTS addresses concerns associated with SHS students' reflective thinking in dealing with Mathematics problem-solving. The instrument is likely to be relevant to various stakeholders including students, instructors, administrators, DepEd curriculum developers, guidance counsellors, researchers, and others. For example, DepEd curriculum developers may use the instrument data to revisit the SHS curriculum, as a guide in their work to improve their designs, and identify the necessity of addressing concerns and issues in achieving quality of basic education in the Philippines through the development of critical thinking and problem solving as reflective thinking skills to learners which facilitate reflective learning. DepEd, school administrators, and guidance counselors may also utilize the test as a diagnostic tool to gauge students' critical thinking skills and identify those whose capacity for reflective thought is likely to be subpar before modifying course content to increase student achievement. On the other hand, the data gathered using the tool will help teachers, administrators, and researchers better understand how students' reflective thinking traits may relate to their ability to solve problems and finish their mathematics coursework, thereby significantly enhancing their mathematical competency.

The MRTS is limited to just 27 items for SHS school students. This constraint points to the need for other studies that may require more items particularly on the area of mastery with only one indicator. Future studies may take into account more dimensions and assess how much data is consistent across different subgroups (e.g. by gender, by strand).

Bibliography

1. Mcleod, Saul (2019). Constructivism as a theory for teaching and Learning. Available online: https://www.simplypsychology.org/constructivism.html
2. Kızılkaya, G., & Aşkar, P. (2009). The development of a reflective thinking skill scale towards problem solving. Education and Science, 34(154), 82-92. Online: http://eb.ted.org.tr/index.php/EB/article/view/550/44 . Kubiszyn,

T. & Borick, (2004) Educational testing and measurement Classroom Application and Practice 7thed. New York: John Wiley& Sons, Inc. ,2004.

3. Can, S. (2015). Pre-service science teachers reflective thinking skills toward problem solving. Educational Research and Reviews, 10(10), 1449-1457.
4. De Leon-Pineda, Josephine Luz & Prudente, Maricar. (2018). Development and Validation of Reflective Thinking Questionnaire for Senior High School Students. Advanced Science Letters. 24. 8072-8075. 10.1166/asl.2018. 12494
5. Hendriana, Heris & Putra, Harry & Hidayat, Wahyu. (2019). How to Design Teaching Materials to Improve the Ability of Mathematical Reflective Thinking of Senior High School Students in Indonesia?. Eurasia Journal of Mathematics, Science and Technology Education. 15. 10.29333/ ejmste/112033.
6. Boone, W. (2016). Rasch Analysis for Instrument Development: Why, When, and How?. CBE Life Sci Educ. 2016 Winter; 15(4): rm4. doi: 10.1187/cbe.16-04-0148
7. Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch Pada Assessment Pendidikan. Cimahi: Trim Komunikata Publishing House.
8. Royal, K. D., & Gonzalez, L. M. (2016). An evaluation of the psychometric properties of an advising survey for medical and professional program students. Journal of Educational and Developmental Psychology, 6(1), 195–203. https://doi.org/10.5539/jedp.v6n1p195
9. Linacre, J. M. (2018). A user's guide to Winsteps® Ministeps Rasch model measurement computer program user's guide. Beaverton, OR: Winsteps.com. Retrieved from https://www.winsteps.com/a/Winsteps-Manual.pdf
10. Bond, T. G., & Fox, C. M. (2007). Applying The Rasch Model: Fundamental Measurement in the Human Science (2nd Ed). New Jersey: Lawrence Erlbaum
11. Linacre, J. M. (2003). Dimensionality: Contrasts & variances. Help for Winsteps Rasch Krishnan, S. & Idris, N. (2014). Investigating Reliability and Validity for the Construct of Inferential Statistics. International Journal of Learning, Teaching and Educational Research, 4(1), 51-60.
12. Linacre, J. M. (2018). A user's guide to Winsteps® Ministeps Rasch model measurement computer program user's guide. Beaverton, OR: Winsteps.com. Retrieved from https://www.winsteps.com/a/Winsteps-Manual.pdf
13. Yang, Hongwei; Su, Jian; and Bradley, Kelly D., "Applying the Rasch Model to Evaluate the Self-Directed Online Learning Scale (SDOLS) for Graduate Students" (2020). Educational Policy Studies and Evaluation Faculty Publications. 19. https://uknowledge.uky.edu/epe_facpub/19
14. Bond, T.G. & Fox, C. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences (3rd Ed).