Aligning the Attitude towards Math Inventory

Chue Kah Loong[1]

[1]National Institute of Education, Nanyang Technological University, Singapore
kahloong.chue@nie.edu.sg

Abstract. The aim of this study is twofold: 1) to align the Attitude towards Math Inventory so that it fits the construct of an attitude and 2) to investigate the psychometric properties and possible gender biasness in the inventory. The short version of the attitude towards math inventory was modified to fit the typical construct of an attitude by including a behavioral subscale and removing the subscale pertaining to self-confidence. The inventory was then administered to a sample of 419 tertiary level students (181 male, 238 female) enrolled in a mathematics course. Data was subsequently entered into WINSTEPS ver 3.92 for Rasch analysis. Results indicated that the three dimensions were consistent with the attitude construct of affect, behavior and cognition. Items in the inventory demonstrated low local dependency, targeted the sample reasonably well and were a relatively good fit for the data. Both person and item separation indicated a high reliability. There were no substantial differential item functioning effects for gender. Evidence from the study indicates that the modified scale is a valid instrument to assess attitude as defined by attitude researchers. Educators may wish to use the scale to measure a student's attitude for early interventions or for future research purposes.

Keywords: Rasch analysis, Differential item functioning, Attitude towards Mathematics Inventory, Scale validation.

1 Introduction

In many undergraduate programs such as engineering or computer programming, it is increasingly common for mathematics to be a core or elective subject. However, even after years of studying mathematics in secondary education, many students still perceive the subject as difficult and one that causes least enjoyment [1] [2]. Several reasons for this have been put forward but many educators would agree that a student's attitude towards mathematics plays a major role. The development of a positive attitude towards the subject is important as research has shown that students' attitude towards mathematics not only impacts their academic outcomes [3] [4] [5], it also has an impact on the climate of the class [6] [7].

Q. Zhang (ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)*,
Atlantis Highlights in Social Sciences, Education and Humanities 23,
https://doi.org/10.2991/978-94-6463-494-5_13

1.1 Attitude towards Mathematics Instruments

To measure the strength of a student's attitude towards mathematics, several instruments have been developed over the past decades. The more widely used instruments are the Fennema-Sherman Mathematics Attitude Scales (FSMAS) [8] [9] and the Attitude Towards Mathematics Inventory (ATMI) [10] [11]. The FSMAS is the most widely used instrument [12] and comprises of 108 items broken down into nine subscales of male domain, mother, father, teacher, confidence in learning, math anxiety, effectance motivation and usefulness. However, some researchers who have used the FSMAS reported poor validity and reliability [13], adding that there was not a suitable model fit for the original structure.

On the other hand, the ATMI is one of the latest instruments to measure students' attitude towards mathematics [12]. The original ATMI comprised of 40 items broken down into four subscales of security, value, motivation and enjoyment. The security subscale has 15 items and measures the anxiety and self-confidence level, the value subscale has 10 items and measures the belief that mathematics is a worthwhile subject to study, the motivation subscale has 5 items and measures the interest level and the enjoyment subscale has 10 items and measures the degree of satisfaction gained from studying mathematics. The instrument was tested on a sample of high school students in Mexico and demonstrated sound reliability and validity [10].

Subsequently, a shortened version was developed that comprised of 15 items broken down into three subscales of self-confidence, value and enjoyment [11]. The removal of one subscale was suggested due to the strong correlation between the enjoyment and motivation subscale. The shortened version was tested on a sample of 1601 students in Singapore and demonstrated sound psychometric properties. In addition, the researchers noted that the completion of the shortened version took less than 10 minutes for the participants, making it an attractive option for educators with limited time. The shortened version of the ATMI has been separately validated in other contexts [14].

1.2 Attitude towards Mathematics Instruments

However, despite its stable factor structure and sound psychometric properties, several limits still exist in the ATMI that have not been addressed fully. First, the construct of an attitude towards mathematics may not be measured accurately. Currently, the most accepted and well-recognized attitude model proposes that attitude is made up of three components, namely affect, behaviour and cognitive [15] [16] [17]. Affect refers to the emotional response towards an attitude object, behaviour refers to the observable responses that a person has towards the attitude object and cognitive refers to the beliefs of a person about the attitude object. For example, a positive attitude towards math could mean that a person likes to study math, puts in effort into math problems and believes that math is an important subject to do well in. In the shortened ATMI, affect is measured by the enjoyment subscale and cognitive is measured by the value subscale. However, there are no items pertaining to the behavioural aspects of attitude towards math. Furthermore, the subscale of self-confidence may be a secondary dimension that is unrelated to attitude and researchers have argued for it to be measured separately [18] [19].

Second, there has been few investigations on whether the items in the ATMI scale are biased against a specific gender. Many researchers have concluded that there are gender differences in attitude towards mathematics. Generally, boys are reported to have more positive math attitudes than girls [20] [21], although there have been other studies reporting that girls hold a more positive attitude [22] or that there were no gender differences [23]. Whilst there are many variables (e.g. school environment, parental attitudes, teaching styles) that could have contributed to this disparities, one aspect that is seldom investigated concerns whether items in the scale are biased against females or males.

The purpose of this study is to address the two limitations above. To accurately measure the attitude construct, the self-confidence subscale of the ATMI will be replaced by a subscale that measures the behavioral aspects of an attitude. As the behavior of students towards mathematics are related to the amount of effort that they put in, these items will be obtained and modified from a Work Effort Scale [24]. The scale was designed to assess the direction, intensity and persistence as a measure of an individual's work effort. Since the direction of an individual's effort was related to an organization, only items that measured persistence and intensity of effort were taken and modified to fit a mathematical learning context. The scale comprised of 10 items with responses on a 7-point Likert scale. Items from the scale has been shown internally consistent and statistically valid in other contexts [25] [26]. Subsequently, the validity of the new scale will be analysed using Rasch modelling techniques. Gender biasness in the items will be investigated using differential item functioning. This paper will thus address the following research question: What are the psychometric properties of the modified ATMI?

2 Method

2.1 Sample and Procedure

The participants for the survey were 419 students enrolled in a diploma course that is a pre-requisite for entrance to an undergraduate program in business. The sample comprised of 181 male and 238 female students. The mean age was 19.0 years, SD = 1.98. English was the medium of instruction.The questionnaire was administered in the middle of the semester over two cohorts of students. The lesson was concluded half an hour earlier to provide students with time to complete the questionnaire. Administration time took approximately 15 minutes. They were informed that their responses would be kept confidential and had the option to withdraw from the survey at any time.

2.2 Instruments

Attitude towards mathematics was measured using the enjoyment and value subscales of the ATMI [11] combined with a modified version of the Work Effort scale [24]. In total, there were 3 subscales: Enjoyment (ENJ - 5 items), Value (VAL - 5 items) and Effort (EFF - 5 items). Each item was measured on a 7 point scale from 1 = strongly disagree to 7 = strongly agree.

Mathematics achievement was measured using their grades from the end of semester written examination. As the examination was a standardized test that assessed the content taught in the entire semester, the grades were a good indicator of their academic achievement. In both cohorts, the examination assessed students on both conceptual understanding (50%) and application to business and economics (50%). To ensure content validity, both examinations were each vetted by two external subject matter experts. The subject matter experts considered the difficulty level of the items, coverage of the content within the semester, the appropriateness of the language used, and the weightage of the content and skills assessed in the examination

3 Data Analysis and Results

Prior to analyzing the data, all responses to negatively worded statements were reversed. Data from the completed questionnaires were coded and descriptive statistics for each item are shown in Table 1. The data was subsequently entered into WINSTEPS ver 3.92 for Rasch analysis.

Table 1. Descriptive statistics for the attitude towards math scale.

Item label	Full statement	Mean	SD
ENJ1	I usually enjoy studying math in school.	4.57	1.70
ENJ2	I like to solve new problems in math.	4.49	1.67
ENJ3	I like math.	4.54	1.76
ENJ4	I am happier in math class than in any other class.	4.13	1.80
ENJ5	Math is a very interesting subject.	4.64	1.71
VAL1	Math is a very worthwhile and necessary subject.	5.09	1.50
VAL2	Math is important in everyday life.	5.04	1.51
VAL3	Math is one of the most important subjects for people to study.	5.03	1.57
VAL4	Math lessons would be very helpful no matter what I decide to study in future.	4.97	1.54
VAL5	A strong math background could help me in my professional life.	5.24	1.49
EFF1	I do not give up quickly when solving math problems.	5.00	1.43
EFF2	I do my best in my math assignments.	5.51	1.24
EFF3	When I start a math problem, I pursue it to the end.	5.04	1.35
EFF4	I think of myself as a hard worker	4.56	1.28
EFF5	I put a lot of energy into the math problems that I start.	4.83	1.25

3.1 Rasch Analysis

The Rasch model is a logistic model first developed by George Rasch to analyse dichotomously scored items [27]. In the model, item difficulty and person ability are placed on a single ratio scale whereby the units in terms of logits (i.e. log-odds). In general, a higher logit value indicates greater item difficulty or higher person ability. Extending on his work, two further models, namely the Rating Scale model and the Partial Credit model, were developed to analyse polytomously scored items. The Rating Scale model assumes that all items on the scale has the same rating scale structure [28] whereas the Partial Credit model assumes that each item has a unique rating scale structure [29].

In this study, the Rating Scale model was used to analyse the instrument as all items were represented by a 7-point Likert scale. The rating scale model specifies the probability that a person n will choose a given category k on an item i. The mathematical expression is given by

$$P_{nik}\left(B_n\right)=\frac{e^{\left(B_n-D_i-F_k\right)}}{1+e^{\left(B_n-D_i-F_k\right)}}$$

where B_n refers to the person ability, D_i is the difficulty of the entire item and F_k is the difficulty estimate of the k^{th} threshold. In this instance, B_n is represented by students' attitude towards mathematics, D_i indicates how likely the entire item would be endorsed and F_k is the estimate at which a person has an equal chance of scoring category k and category $k - 1$. The Rating Scale model assumes that threshold estimate F_k is the same across all items.

Prior to performing a Rasch analysis, two major assumptions of unidimensionality and local independence must be met. Unidimensionality refers to the concept that the items measure a single latent trait or ability. Local independence means that a person's response on one item is independent of other items. Once the two assumptions are met, the procedure estimates the various parameters based on the frequency of responses for each item. Tests are conducted to assess item fit as well as item and person reliability. Potential differential item functioning (DIF) between gender were also investigated. Finally, the predictive validity of the scale was examined by a correlational analysis between the ability measures predicted by the Rasch model and their actual achievement results.

3.2 Dimensionality and Local Independence

Local independence between items was assessed by using the correlations between the standardized residuals of item pairs to compute the distribution of Fisher's Z. Item pairs with Fisher's Z values within 2 standard deviations of the mean are considered to be locally independent [30]. The dimensionality of the scale was assessed using standardized residual principal components analysis (PCA). The procedure looks for patterns within the residuals in the data to determine whether they share similar patterns. The strength of each PCA component of residuals, i.e. the contrasts, is provided by their eigenvalues. If the size of eigenvalues exceeds 2 (strength of 2 items), it is indicative of a secondary dimension. Items in secondary dimensions can be observed from the patterns in the contrast plot and their contrast loadings [31].

The distribution of Fisher’s Z had a mean of -.07 and a standard deviation of 0.25. Thus the critical values for significant Fisher’s Z were -.57 and .43. Examining the statistics for item pairs revealed 5 residual correlations beyond these critical values. However, as the identified residual correlations represent less than 5% of all the residual correlations, this indicates that item and person estimates would not be adversely impacted [32] and thus no corrections were made.

The eigenvalue of the first and second contrast was 3.4 and 3.1 respectively, indicating that there were possibly three dimensions within the scale. An inspection of the contrast plot and contrast loadings revealed that there were three clusters of items that were consistent with the three dimensions of Enjoyment, Value and Effort (See Table 2).

Table 2. Standardized residual loadings for items.

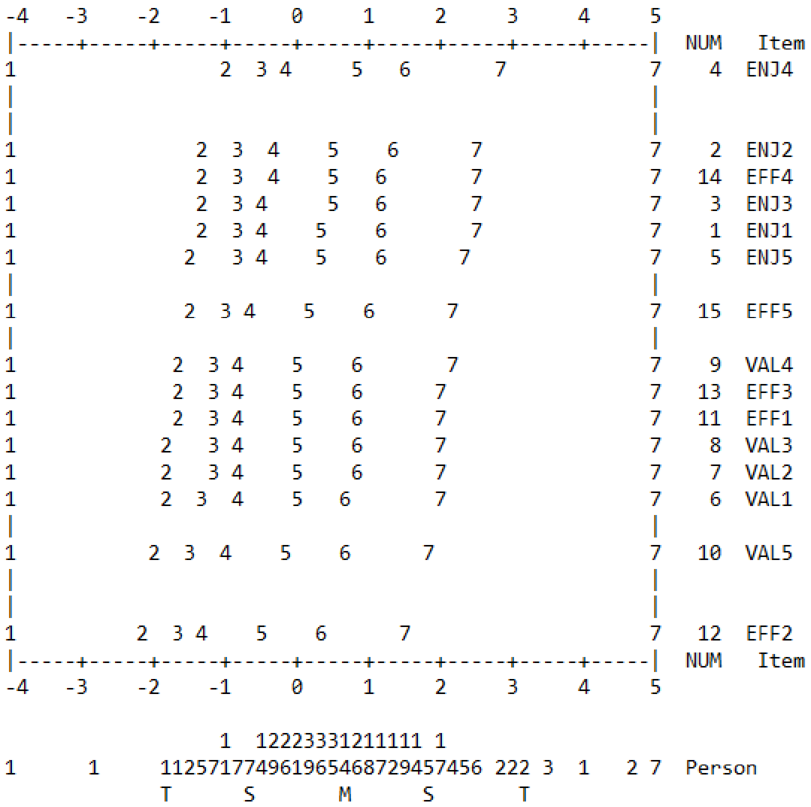
Contrast	Item	Loading	Contrast	Item	Loading	Contrast	Item	Loading
1	ENJ2	.70	2	EFF3	.14	3	V1	-.47
1	ENJ3	.66	2	EFF1	.11	3	V5	-.54
1	ENJ4	.49	2	EFF5	-.01	3	V4	-.63
1	ENJ1	.46	2	EFF2	-.07	3	V2	-.63
1	ENJ5	.46	2	EFF4	-.16	3	V3	-.66

To cater for multi-dimensionality, three possible methods have been suggested: the unidimensional approach where all items are treated as representative of one dimension, the consecutive approach where each dimension are treated separately or the multi-dimensional approach which is a compromise between the unidimensional approach and the consecutive approach [33]. The unidimensional approach was adopted in this study because of the following reasons: 1) All the items can be projected onto a single dimension of attitude towards mathematics, 2) A lesser number of items in each dimension would result in a larger standard error of the estimates and 3) the disattenuated correlation between the person’s measures on the various dimensions are substantial ($R_{\text{Enj-Val}} = 0.73$, $R_{\text{Enj-Eff}} = 0.64$, $R_{\text{Eff-Val}} = 0.62$), indicating a high inter-relation between the dimensions.

3.3 Item Threshold Estimates

The Rasch model places item difficulty and person ability on the same scale. In the current study, the Rasch-Thurstone thresholds are used for item difficulty which are the points of equal cumulative probability above and below the boundary of two options [34]. To check whether items are suitably targeted at the sample, a 50% cumulative category probability figure is produced. The item difficulty at each threshold is shown in the table and the number of persons at that particular ability level is shown at the bottom. For example, if an item difficulty at the threshold of 4 is –1, a person at an ability level of –1 would have a 50% chance of scoring below 4 and a 50% chance of scoring 4 and above.

Figure 1 shows the 50% cumulative category probability table. The threshold estimates range from –2.25 to 2.81 whereas person ability (excluding extreme data) ranged from –2.80 to 4.67. The mean person ability was .55 logits with a standard deviation of 1.03 logits, suggesting that the scale was targeted at individuals who are slightly less able, i.e. slightly negative attitude. The scale did not suitably target 7 persons whose ability measure was at a maximum extreme score of 5.89 logits and 1 person with a minimum extreme score of –4.93. The 8 persons constituted 1.9% of the overall sample.



```
-4   -3    -2    -1     0     1     2     3     4     5
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM   Item
1                 2  3 4     5   6       7            7    4  ENJ4
|                                                     |
|                                                     |
1               2  3  4    5    6      7              7    2  ENJ2
1               2  3  4    5  6        7              7   14  EFF4
1               2  3 4     5  6        7              7    3  ENJ3
1               2  3 4    5   6        7              7    1  ENJ1
1              2   3 4    5   6       7               7    5  ENJ5
|                                                     |
1              2  3 4    5    6      7                7   15  EFF5
|                                                     |
1             2  3 4    5    6       7                7    9  VAL4
1             2  3 4    5    6      7                 7   13  EFF3
1             2  3 4    5    6      7                 7   11  EFF1
1            2   3 4    5    6      7                 7    8  VAL3
1            2   3 4    5    6      7                 7    7  VAL2
1            2  3  4    5   6       7                 7    6  VAL1
|                                                     |
1           2  3  4    5    6      7                  7   10  VAL5
|                                                     |
|                                                     |
1          2  3 4    5    6      7                    7   12  EFF2
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM   Item
-4   -3    -2    -1     0     1     2     3     4     5

                  1  12223331211111 1
1      1     112571774961965468729457456 222 3  1   2 7  Person
             T      S       M      S       T
```

M: Mean, S: 1 SD from mean, T: 2 SD from mean

Fig. 1.50% cumulative category probability table.

3.4 Item Fit Statistics

Item fit were assessed using two statistics. One, the observed point-measure correlation is compared with the expected point-measure correlation. The point-measure correlation is a correlation between the observed/expected observations with the person's measure and is an item discrimination index. A higher observed point-measure correlation corresponds to a better item. Comparing the observed correlation with the expected correlation gives an indication of the item fit as the difference should be not too significant [35]. Two, Rasch analysis provides mean-square and z-standardized values as fit statistics. Mean square values are based on chi-square statistics and indicate whether the data fits the model whereas z-standardized values are z-scores that whether the data fit is significant. These fit statistics are further divided into outfit and infit values. Outfit refers to outlier-sensitive fit and places equal weightage onto all responses. Infit refers to inlier-sensitive fit and places more weightage on items targeted on the person or group. Ideally, mean square fit statistics should lie between 0.6 and 1.4 although values less than 2 would not degrade the measurement. Z-standardized statistics should lie between –1.96 and 1.96 [36].

Table 3 shows the item fit statistics for the scale. The absolute differences between the observed and expected point-measure correlation ranged from .00 to .17. The Outfit mean square values range from .78 to 1.64 and the Infit mean square values range from .80 to 1.39. There were 4 items in which the absolute z-standardized values were greater than 1.96. However, as z-standardized statistics can be overly sensitive when there are more than 300 observations, these values are only taken into consideration when the mean square values indicate a misfit. All items but one satisfied the criteria for the mean square fit statistics. The item EFF4 ("I think of myself as a hard worker") had an outfit mean square of 1.64 (z-standardized = 7.5) and an infit mean square of 1.39 (z-standardized = 4.9). Since the infit mean square value was within the criteria, the item was not removed from the scale.

Table 3. Item fit indices.

Item	Pt. measure correlation		Outfit		Infit	
	Obs	Exp	Mnsq	Z-std	Mnsq	Z-std
ENJ1	.65	.69	1.39	4.9	1.21	2.8
ENJ2	.73	.69	.85	-2.2	.88	-1.8
ENJ3	.74	.69	.91	-1.2	.90	-1.5
ENJ4	.72	.71	.99	-.1	.99	-.2
ENJ5	.75	.68	.78	-3.3	.80	-2.9
VAL1	.69	.66	.87	-1.7	.90	-1.4
VAL2	.66	.66	1.07	.9	1.05	.6
VAL3	.69	.66	.95	-.7	.99	-.1
VAL4	.68	.66	.97	-.4	.98	-.3
VAL5	.66	.65	.98	-.3	1.00	.0
EFF1	.67	.66	1.01	.2	.91	-1.2
EFF2	.62	.62	.97	-.4	.93	-1.0
EFF3	.67	.66	.95	-.7	.97	-.4
EFF4	.52	.69	1.64	7.5	1.39	4.9
EFF5	.63	.67	1.08	1.1	.96	-.5

3.5 Item and Person Reliability

The reliability of the scale was assessed using item and person separation and reliability indices. Person reliability is analogous to Cronbach alpha and is indicative of how well the scale can differentiate individuals according to their ability. Person separation refers to the number of statistically different strata of individuals in the sample that can be identified, e.g. if person separation is 3, then 3 levels of people of differing ability can be identified in the sample. Likewise, item reliability refers to how well the sample can separate items in the scale and item separation refers to the number of different difficulty strata in the items. Acceptable values for reliability are similar to those for Cronbach alpha.

The person separation was 2.95 with a reliability of .90 and the item separation was 6.26 with a reliability of .98. This implies that the scale can identify approximately 3 strata in the sample, e.g. high, average, low attitude, and items can be split into approximately 6.3 levels of difficulty. Both person and item separation demonstrate high reliability.

3.6 Differential Item Functioning

Differential item functioning (DIF) by gender was assessed by examining the value of the DIF contrast, that is, the difference in measures in logits between two groups [37]. An item is considered biased when the DIF contrast is significant and the effect size is substantial. A *t*-test is conducted to examine significance ($p < .05$) and effect size is estimated by the absolute value of the DIF contrast. The following criteria is used: < .43 logits is negligible, between .43 to .64 logits is moderate, > .64 logits is large [38].

Table 4 shows the DIF statistics for the 15 items. There was only one item that exhibited a significant DIF: ENJ4 ("I am happier in math class than in any other class") had an absolute DIF contrast of .25 logits ($|t(385)| = 2.83$, $p = .005$). For this item, the difficulty measure for males was .74 compared to .50 for females, indicating that this item is biased against males. However, its absolute DIF contrast was .25 which implies that the DIF effect is negligible.

Table 4.DIF statistics.

Item	|DIF contrast|	|*t*|	*p*
ENJ1	.08	.91	.36
ENJ2	.00	.00	1.00
ENJ3	.13	1.43	.15
ENJ4	.25	2.83	.005
ENJ5	.08	.87	.39
VAL1	.07	.68	.50
VAL2	.17	1.73	.08
VAL3	.00	.00	1.00
VAL4	.14	1.41	.16
VAL5	.11	1.08	.28
EFF1	.10	1.07	.29
EFF2	.00	.00	1.00
EFF3	.00	.00	1.00
EFF4	.10	1.14	.25
EFF5	.05	.54	.59

3.7 Predictive Validity

As students' attitudes towards mathematics has been shown to be positively related to academic outcomes [5], the predictive validity of the scale was assessed by computing the correlation of persons' ability measures as predicted by the Rasch model and their actual achievement results on the semester written examination. The correlation coefficient was .38 ($p < .001$), indicating that the scale has a good predictive validity.

4 Discussion

The main aims of this study are twofold: 1) to validate a scale that accurately measures the attitude towards mathematics construct by replacing the self-confidence subscale of an existing scale with the behavioural aspect of an attitude construct and 2) to investigate whether there were DIF concerns in the scale that might have caused biasness against a specific gender. Rasch techniques were used to assess the psychometric properties of the scale.

Initial analysis indicated that most of the items were locally independent of each other, and that the small percentage of local dependencies would not have any adverse impact on measurements. The scale could also be partitioned into three dimensions of affect, behavioural and cognitive, which was consistent with the attitude construct. However, as the three dimensions together could be construed as a single construct of attitude and since they were highly correlated with each other, it was decided to analyse the scale as an unidimensional construct.

The items targeted the sample reasonably well, with only 1.9% of the sample located beyond the boundaries of the item thresholds. Nevertheless, further improvements to the scale could be made by inserting items that are able to target the higher end of the ability measures. Most items demonstrated a reasonable to good fit although one item had a relatively high outfit mean square value (EFF4: "I think of myself as a hard worker"), suggesting that the item could be influenced by some other variable other than attitude towards math. This may be due to the phrasing of the item that may have caused overlaps between working hard in mathematics with the general characteristics of students. For example, EFF4 could be interpreted as a measurement of general conscientiousness. On hindsight, it might have been better to paraphrase the item to reflect conscientiousness exclusively in the domain of mathematics.

The item and person separation and reliability indices indicated that items in the scale were able to differentiate people into 3 strata of ability reasonably well. Likewise, individuals in the sample were also able to separate the items into approximately 6.3 difficulty levels very well. This suggests that in future, the number of points for each item in the scale may be trimmed down to 6 options instead to improve reliability.

To address the issue of gender biasness, a DIF test indicated that most items had no significant and substantial DIF effects with the exception of one item (ENJ4: "I am happier in math class than in any other class") that had a significant effect. Males tend to find this item harder to agree and this may be due to females expressing their emotions more readily than males (Simon & Nath, 2004). Even so, the DIF effect was not substantial and the item was left in the scale.

4.1 Limitations and Future Directions

A number of limitations should be noted in this study. First, the Rasch model only includes item difficulty as a single parameter in its analysis. Although this may be sufficient, analyzing future datasets with more complex models will also be useful. Second, the sample was solely from one tertiary institution. There is a need to expand the sample size to include other students for more generalizable results. Third, it is possible that DIF effects are culturally dependent and may not replicate itself in a different context. Future research could conduct a similar DIF analysis in a different cultural setting.

5 Conclusion

In summary, the evidence in the current study indicates that the modified scale would be a valid instrument to measure attitude towards mathematics. Theoretically, it addresses the attitude construct as defined by attitude researchers and the items have no substantial gender biasness. Practically, the scale is easy to administer and takes a short time to complete. Educators may wish to use the scale to assess a student's attitude before the start of a module. This will aid them in providing early interventions to address future problems that students may face. For example, a student may perceive the value of mathematics and enjoy lessons but does not put in the necessary work. Others may also adapt this scale in attitudinal or mathematical research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Nardi, E., & Steward, S. (2003). Is mathematics TIRED? A profile of quiet disaffection in the secondary mathematics classroom. British Educational Research Journal, 29(3), 345-366.
2. Wong, N.-Y., Lam, C.-C., Wong, K.-m. P., Leung, F.-S., & Mok, I.-C. (2001). Students' views of mathematics learning: A cross-sectional survey in Hong Kong. Educational Journal-Hong Kong-Chinese University of Hong Kong, 29(2), 37-60.
3. Grootenboer, P., & Hemmings, B. (2007). Mathematics performance and the role played by affective and background factors. Mathematics Education Research Journal, 19(3), 3-20.
4. Samuelsson, J., & Granstrom, K. (2007). Important Prerequisites for Students' Mathematical Achievement. Online Submission, 3(2), 150-170.
5. San Pedro, M. O. Z., Snow, E. L., Baker, R. S., McNamara, D. S., & Heffernan, N. T. (2015). Exploring Dynamical Assessments of Affect, Behavior, and Cognition and Math State Test Achievement. Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.
6. Kiwanuka, H. N., Van Damme, J., Van Den Noortgate, W., Anumendem, D. N., Vanlaar, G., Reynolds, C., & Namusisi, S. (2017). How do student and classroom characteristics

affect attitude toward mathematics? A multivariate multilevel analysis. School Effectiveness and School Improvement, 28(1), 1-21.

7. Vandecandelaere, M., Speybroeck, S., Vanlaar, G., De Fraine, B., & Van Damme, J. (2012). Learning environment and students' mathematics attitude. Studies in Educational Evaluation, 38(3), 107-120.
8. Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. Journal for Research in Mathematics Education, 7(5), 324-326.
9. Mulhern, F., & Rae, G. (1998). Development of a shortened form of the Fennema-Sherman Mathematics Attitudes Scales. Educational and Psychological Measurement, 58(2), 295-306.
10. Tapia, M. (1996). The Attitudes toward Mathematics Instrument. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, AL.
11. Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. Educational Studies in Mathematics, 82(1), 145-164.
12. Chamberlin, S. A. (2010). A review of instruments created to assess affect in mathematics. Journal of Mathematics Education, 3(1), 167-182.
13. Melancon, J. G., Thompson, B., & Becnel, S. (1994). Measurement integrity of scores from the Fennema-Sherman Mathematics Attitudes Scales: The attitudes of public school teachers. Educational and Psychological Measurement, 54(1), 187-192.
14. Lin, S.-H., & Huang, Y.-C. (2016). Development and Application of a Chinese Version of the Short Attitudes toward Mathematics Inventory. International Journal of Science & Mathematics Education, 14(1), 193-216.
15. Ajzen, I. (1989). Attitude structure and behavior. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function (pp. 241-274): L. Erlbaum Associates.
16. Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. Gilbert, S. Fiske, & G. Lindsay (Eds.), Handbook of Social Psychology (pp. 269-322): McGraw-Hill.
17. Van den Berg, H., Manstead, A. S., van der Pligt, J., & Wigboldus, D. H. (2006). The impact of affective and cognitive focus on attitude formation. Journal of Experimental Social Psychology, 42(3), 373-379.
18. Evans, J. (2006). Affect and emotion in mathematical thinking and learning. In J. Maass & W. Schlöglmann (Eds.), New Mathematics Education Research and Practice (pp. 233-255): Sense Publishers.
19. Yáñez-Marquina, L., & Villardón-Gallego, L. (2016). Attitudes towards mathematics at secondary level: Development and structural validation of the Scale for Assessing Attitudes towards Mathematics in Secondary Education (SAT-MAS). Electronic Journal of Research in Educational Psychology, 14(3), 557-581.
20. Asante, K. O. (2012). Secondary students' attitudes towards mathematics. IFE Psychologia: An International Journal, 20(1), 121-133.
21. Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis: American Psychological Association.
22. Amirali, M. (2010). Students' conceptions of the nature of mathematics and attitudes towards mathematics learning. Journal of Research and Reflections in Education, 4(1), 27-41.
23. Orhun, N. (2007). An investigation into the mathematics achievement and attitude towards mathematics with respect to learning style according to gender. International Journal of Mathematical Education in Science and Technology, 38(3), 321-333.

24. De Cooman, R., De Gieter, S., Pepermans, R., Jegers, M., & Van Acker, F. (2009). Development and validation of the Work Effort Scale. European Journal of Psychological Assessment, 25(4), 266-273.
25. De Cooman, R., Stynen, D., Van den Broeck, A., Sels, L., & De Witte, H. (2013). How job characteristics relate to need satisfaction and autonomous motivation: Implications for work effort. Journal of Applied Social Psychology, 43(6), 1342-1352.
26. Sakurai, K., & Jex, S. M. (2012). Coworker incivility and incivility targets' work effort and counterproductive work behaviors: The moderating role of supervisor social support. Journal of Occupational Health Psychology, 17(2), 150-161.
27. Rasch, G. (1980). Some probabilistic models for intelligence and attainment tests. Chicago, IL: University of Chicago.
28. Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43(4), 561-573.
29. Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47(2), 149-174.
30. Shen, L. (1997). Quantifying Item Dependency by Fisher's Z. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.
31. Linacre, J. M. (2008). Variance in data explained by Rasch measures. Rasch Measurement Transactions, 22(3), 1164.
32. Smith Jr, E. (2004). Effect of item redundancy on Rasch item and person estimates. Journal of Applied Measurement, 6(2), 147-163.
33. Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. Journal of Applied Measurement, 4(1), 87-100.
34. Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. Educational and Psychological Measurement, 72(4), 547-573.
35. Linacre, J. M., & Rasch, G. (2008). The expected value of a point-biserial (or similar) correlation. Rasch Measurement Transactions, 22(1), 1154-1157.
36. Linacre, J. M. (2003). Size vs. significance: Standardized chi-square fit statistic. Rasch Measurement Transactions, 17(1), 918.
37. Bond, T., & Fox, C. M. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences: Routledge.
38. Zwick, R., Thayer, D. T., & Lewis, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. Journal of Educational Measurement, 36(1), 1-28