Jurnal Elemen, 8(1), 175-186, January 2022
https:/doi.org/10.29408/jel.v8i1.3981

e- ISSN: 2442 - 4226

The development and validation of mathematical reflective thinking test for prospective mathematics teachers using the Rasch model

Muntazhimah [1] *, Reni Wahyuni [2]

[1] Mathematics Education Study Program, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia

[2] Mathematics Education Study Program, Universitas Islam Riau, Riau, Indonesia

* Correspondence: muntazhimah@uhamka.ac.id

Abstract

Research on developing a mathematical reflective thinking test instrument for prospective mathematics teachers using the Rasch model is limited. This study aims to develop a reflective thinking ability test for prospective mathematics teachers using the Rasch model. The development steps consist of preparing a blueprint, writing items, reviewing items, conducting trials, analyzing test results, and revising. The test developed is a description test validated by five experts, consisting of two expert professors in evaluating mathematical ability, one expert lecturer in mathematical ability, and two expert lecturers in group theory. The test was also piloted on 26 students. The study's results using the Rasch model analysis showed that this instrument was "very reliable," and 12 of the 13 test items were "valid." The instrument has fulfilled all the stages of instrument development and was declared valid and reliable; thus, this test can be used to examine prospective mathematics teachers' mathematical reflective thinking ability.

Keywords: mathematical reflective thinking; Rasch model; test

Received: 1 September 2021 | Revised: 27 September 2021
Accepted: 23 December 2021 | Published: 6 January 2022

Introduction

The Law of the Republic of Indonesia No. 14 of 2005 stipulates that a teacher or lecturer must possess pedagogic, personality, social, and professional competencies. Furthermore, Minister of National Education Regulation No. 16, the Year 2007 concerning Academic Qualification Standards and Teacher Competencies stated that one of the core competencies in pedagogic competence is to conduct assessments, evaluate learning processes and outcomes. One of the implementations to mastery of these competencies is that teachers must develop assessment instruments and evaluate processes and learning outcomes.

Various types and kinds of instruments can be used as tools to measure the achievement of learning objectives, one of these instruments is a test (Prabowo & Dahlan, 2020; Suharman, 2018; Wahyudi, 2012). A test is crucial in education, as mandated in the law above, which is also the basis that educators must be willing and able to develop test instruments. However, the data shows that the intensity of teachers in developing test instruments is minimal (Osnal et al., 2015), resulting in many test instruments not fulfilling the standard of a quality test (Prabowo et al., 2018; Wardhani & Putra, 2016). It is necessary to analyze the quality of the instrument to obtain an excellent test (Arifin, 2016). The instrument analysis in the future must fulfill the quality criteria in substance, construction, language and meet the elements of validity and reliability. These stages also apply when developing research instruments in mathematics education. The instruments used in the research must also go through the analysis process.

Mathematical reflective thinking ability is essential in various literature, especially in professional teacher education (Amidu, 2012; Yuen Lie Lim, 2011). Mathematical reflective thinking ability is included in Higher Order Thinking Skills (HOTS). 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. (Muin et al., 2018; Clarà, 2015). Mathematical reflective thinking ability is one of the mathematical abilities required by all students. Through this ability, students actively, earnestly, and carefully consider using the knowledge obtained on the given mathematical problem.

Various studies on mathematical reflective thinking skills in prospective teachers have been carried out. Among reseRachers are the reflective thinking ability of prospective teachers in Aceh based on gender and prior knowledge (Rahmadhani et al., 2020); adapting the Reflective Thinking Questionnaire (RTQ) into the Indonesian version, and investigating its quality in prospective chemistry teachers in Tanjungpinang (Sabekti et al., 2020); and the reflective thinking process of a female student who has an independent type of cognitive style in solving algebraic problems (Agustan et al., 2017). Some of the findings included research on the development of teaching materials to develop high school students' mathematical reflective thinking skills (Hendriana et al., 2019; Nindiasari et al., 2016), and Muntazhimah (2019) regarding the development of a mathematical reflective thinking test instrument for 8th-grade junior high school students. On the other hand, research on teaching materials and instruments is rare.

From several previous studies mentioned above, research references on developing mathematical reflective thinking tests for prospective mathematics teachers are limited. Also, most research carried out employed classical test theory instead of Rasch Model. Rasch modeling has advantages over classical theory tests, including predicting missing data based on a systematic response pattern (Nur et al., 2020; An & Yu, 2021). The model makes statistical analysis results more accurate. More importantly, Rasch modeling can produce standard error measurement values for the instruments used and increase the accuracy of calculations (Sumintono & Widhiarso, 2015).

Therefore, this study aims to develop a reflective thinking ability test instrument for prospective mathematics teachers using the Rasch model to analyze the validity and reliability. This research was expected to measure the reflective thinking ability of prospective mathematics teachers and is beneficial for further research related to the learning process to improve teachers' mathematical reflective thinking.

Methods

This study applied the method and procedures developed by Spaan (2006), Inc (2006), and Prabowo and Dahlan (2020). Their procedures include preparing a blueprint, writing items, reviewing items, conducting trials, analyzing test results, and revising to produce an effective, efficient, and quality mathematical reflective thinking test instrument for prospective mathematics teachers.

The first step began with compiling a blueprint. Rowe (2001) stated that blueprints could be a practical guide for composing items. Thus, it can also be referred to as a test matrix guideline. This guideline will state the definition of mathematical reflective thinking ability, selecting relevant materials or courses, and adjusting sub-course achievements with mathematical reflective thinking indicators.

The next step was to write the items by deriving the indicator of reflective thinking ability into several items. Table 1 explains the definitions and indicators used in this study and examples of items designed.

Table 1. Definitions, indicators, and examples of items

Definition	Indicators	Examples of Items
Think thoughtfully by applying mathematical knowledge and experience obtained so that students can analyze, evaluate and get deep meaning in solving mathematical problems.	1. Analyzing the truth of the question/solution/analogy or generalization of mathematics, 2. Identifying mathematical concepts or formulas used in complex math problems, 3. Distinguishing between relevant and irrelevant data, 4. Evaluating the validity of arguments based on the	This sample item represents the fifth indicator Example: H= $\{[0], [3][6], [9]\}$. and $(\mathbb{Z}_{12}, +_{12})$ is a group. **Question:** a. Find how many strategies are there to prove that H is a subgroup of$(\mathbb{Z}_{12}, +_{12})$ b. Explain the concepts, principles,

Definition	Indicators	Examples of Items
	concepts or properties used, 5. Finding various strategies in solving math problems	characters, or rules used by these strategies c. Choose a strategy and prove that H is a subgroup of $(\mathbb{Z}_{12}, +_{12})$

After preparing a test instrument representing each indicator, the third step was to review the items. This step was done by asking for a judgment by five experts to validate the material's content. The five experts consisted of two lecturers who teach the same subject as the material on the test and come from different universities, two professors who are experts in the evaluation of mathematical ability, and one lecturer who is an expert in the field of mathematical ability. After revisions were made on input from the experts, the test instrument was also asked for a limited review by five prospective mathematics teachers to check their readability. Revisions were made following input and suggestions from experts and students in a limited trial.

After revision, the next step was instrument testing conducted on 26 students. The test results were analyzed using the Rasch model assisted by Winstep Rasch 4.4.3 software, and the results were used as the basis for revision. The item samples that have been revised were then compiled into a mathematical reflective thinking test instrument for prospective mathematics teachers.

Results

The study results contain the instrument testing steps analyzed in-depth using the Rasch model. Winstep Rasch 4.4.3 software was used to see the reliability and validity of this mathematical reflective thinking test instrument for prospective mathematics teachers. Validity and reliability are essential aspects that must be considered in developing a new instrument in research (Alavi et al., 2020). Validity is the accuracy and precision of a measuring instrument in carrying out its function. While reliability is originated from the words rely on and ability. Although reliability has various other names, such as trustworthiness, constancy, stability, consistency, and so on, the main idea contained in the concept of reliability is the extent to which the results of a measurement can be trusted.

Reliability is intended to analyze whether this instrument can be used to measure the mathematical reflective thinking ability of prospective mathematics teachers whenever and wherever it is used. This reliability analysis is based on Table 3 (Summary statistics) in the Winsteps program, shown in Figure 1 below.

```
     SUMMARY OF 26 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      27.2      13.0        -.67     .30      1.02   -.04   1.15    .21 |
|  SEM       1.5        .0         .13     .00       .16    .24    .21    .20 |
| P.SD       7.6        .2         .65     .01       .78   1.21   1.04   1.02 |
| S.SD       7.8        .2         .66     .01       .79   1.23   1.06   1.04 |
| MAX.      53.0      13.0        1.40     .33      3.54   3.34   4.87   3.55 |
| MIN.      16.0      12.0       -1.69     .29       .23  -1.86    .34  -1.05 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .34 TRUE SD     .55  SEPARATION  1.64  Person RELIABILITY  .73 |
|MODEL RMSE    .30 TRUE SD     .57  SEPARATION  1.91  Person RELIABILITY  .79 |
| S.E. OF Person MEAN = .13                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .71  SEM = 4.11

     SUMMARY OF 13 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      54.5      25.9         .00     .25      1.01    .04   1.15    .07 |
|  SEM       9.5        .1         .23     .02       .09    .37    .18    .37 |
| P.SD      32.9        .3         .80     .07       .33   1.27    .64   1.28 |
| S.SD      34.3        .3         .83     .07       .34   1.33    .66   1.33 |
| MAX.     143.0      26.0        1.14     .34      1.84   2.73   3.08   2.47 |
| MIN.       9.0      25.0       -1.52     .11       .47  -2.78    .47  -2.74 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .27 TRUE SD     .75  SEPARATION  2.77  Item   RELIABILITY  .88 |
|MODEL RMSE    .26 TRUE SD     .76  SEPARATION  2.93  Item   RELIABILITY  .90 |
| S.E. OF Item MEAN = .23                                                     |
-------------------------------------------------------------------------------
```

Figure 1. Output summary statistics Winstep 4.4.3

Figure 1 shows the Cronbach's alpha of 0.71, indicating the "Good" criteria, as presented in Table 2 (Sumintono & Widhiarso, 2015).

Table 2. The Interpretation of reliability based on Cronbach's alpha value

Score	Interpretation
$a > 0.8$	Excellent
$0,7 < a \leq 0.8$	Good
$0.6 < a \leq 0.7$	Acceptable
$0.5 < a \leq 0.6$	Poor
$a < 0.5$	Unacceptable

Furthermore, for the Person Reliability score listed in Figure 1, the output of 0.73 means that the respondents' answers are in the "good" category. The Item Reliability of 0.88 means the quality of the instrument is in the "Excellent" criteria. Hence, the test instrument for the mathematical reflective thinking ability of the prospective mathematics teachers studied can be concluded as an instrument with a high level of reliability. The detailed description can be seen in Table 3.

Table 3. Reliability test summary

Cronbach Alpha	Interpretation	Item Reliability	Interpretation	Person Reliability	Interpretation	Conclusion
0.71	Good	0.88	Excellent	0.73	Good	Reliable

The instrument's validity is intended to analyze whether the mathematical reflective thinking ability test instrument can be used as a tool to measure mathematical reflective thinking skills for prospective mathematics teachers. The output of the Winstep software can use Table 23 (Item unidimensionality) and Table 10 (Item Fit Order). The output will present which items are appropriate to measure and what is supposed to be measured. Analysis of the validity of the Rasch approach with Winsteps software is called the fit and misfit test (valid and invalid items). The output can be seen in Figure 2.

```
         Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units
                                                  Eigenvalue   Observed     Expected
     Total raw variance in observations      =       42.5612 100.0%         100.0%
       Raw variance explained by measures    =       29.5612  69.5%          70.5%
         Raw variance explained by persons   =        6.9255  16.3%          16.5%
         Raw Variance explained by items     =       22.6358  53.2%          54.0%
       Raw unexplained variance (total)      =       13.0000  30.5% 100.0%   29.5%
         Unexplned variance in 1st contrast =         2.7831   6.5%  21.4%
         Unexplned variance in 2nd contrast =         2.3464   5.5%  18.0%
         Unexplned variance in 3rd contrast =         2.1617   5.1%  16.6%
         Unexplned variance in 4th contrast =         1.2525   2.9%   9.6%
         Unexplned variance in 5th contrast =         1.0658   2.5%   8.2%
```

Figure 2. Output table 23 item unidimensionality Winstep 4.4.3

The item unidimensionality criterion in the Rasch model is seen from the raw variance explained by measures score in Figure 2 (69.5%). If the raw variance explained by measures score > 20% is acceptable; the score > 40% means good, and the score > 60% indicates the excellent criteria. So, this instrument meets the excellent criteria.

Furthermore, the eigenvalue and observed scores in unexplained variance contrast can be rechecked to see if question items are still problematic or inappropriate. The item is not problematic if the eigenvalue is less than 3, and the observed score must be less than 15% for an appropriate item. A results summary of the analysis of the validity test with Winstep software version 4.4.3 is listed in Table 4.

Table 4. Validity test summary

Raw variance explained by measures	**Interpretation**	**unexplained variance 1st contrast**		**Interpretation**
		Eigenvalue	**Observed**	
69.5%	Excellent	2.7831	6.5%	There are no problematic items

Table 4 shows that the score of raw variances explained by measures is 69.5%, meaning that all the mathematical reflective thinking test items meet the "Excellent" criteria. More profoundly, the score observed in the unexplained variance contrast is 6.5%, indicating that all items are appropriate and can be used. The eigenvalue score is 2.7831, showing that all items are good and not problematic. However, further analysis needs to be done, namely item fit order because the eigenvalues are close to 3.

In the item fit order analysis stage, it is essential to look at Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and the Point Measure Correlation (Pt Mean Corr) scores. The complete scores can be seen in Figure 3. The first line shows that for item number 1, the MNSQ outfit score is 3.08, the ZSTD is 1.44 and the PTMEASURE Corr is -0.50. The following line shows that for item number 13, the scores are 1.73 and 2.47 and 0.68 and so on until the last line.

```
        Item STATISTICS:  MISFIT ORDER

-------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PTMEASUR-AL|EXACT MATCH|           |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item   G |
|------------------------------------+----------+----------+-----------+-----------+----------|
|     1      9     26      .64   .29| .95   .21|3.08  1.44|A .50   .61| 84.6  86.5| E1     A |
|    13     40     26      .76   .22|1.84  2.73|1.73  2.47|B .68   .49| 26.9  41.7| E13    E |
|     9     43     26     -.27   .31|1.25   .82|1.32   .94|C .30   .41| 57.7  57.6| E9     C |
|    10     31     26      .80   .29|1.25  1.02|1.24   .98|D .56   .39| 30.8  50.8| E10    C |
|     2    143     26    -1.06   .11|1.08   .39| .90  -.06|E .66   .68| 19.2  19.7| E2     B |
|     4     94     26    -1.52   .18|1.08   .37|1.04   .23|F .57   .50| 23.1  32.8| E4     D |
|     7     64     26     -.45   .23|1.08   .34|1.07   .31|G .14   .57| 50.0  47.6| E7     E |
|    11     42     26      .70   .34| .88   .25| .98   .16|f .64   .66| 65.4  72.0| E11    F ||
|     5     76     26    -1.00   .16| .91  -.30| .89  -.33|e .56   .56| 38.5  31.1| E5     D |
|     3     39     26      .10   .30| .90  -.27| .89  -.28|d .63   .39| 73.1  55.4| E3     C |
|    12     52     25     -.20   .31| .84  -.41| .82  -.45|c .54   .31| 68.0  63.2| E12    G |
|     6     48     26      .36   .22| .57 -1.84| .56 -1.78|b .47   .51| 57.7  43.0| E6     E |
|     8     27     26     1.14   .29| .47 -2.78| .47 -2.74|a .26   .39| 84.6  50.4| E8     C |
|------------------------------------+----------+----------+-----------+-----------+----------|
| MEAN    54.5   25.9      .00   .25|1.01    .0|1.15    .1|           | 52.3  50.1|          |
| P.SD    32.9     .3      .80   .07| .33   1.3| .64   1.3|           | 21.9  17.0|          |
-------------------------------------------------------------------------------------------------
```

Figure 3. Output item fit order Winstep 4.4.3

The last analysis is the Item fit test to understand whether the items function normally to measure. The scores to consider are outfit means-square, outfit z-standard, and point measure correlation (Boone et al., 2014). The criteria for assessing item fit are shown in Table 5 (Sumintono & Widhiarso, 2015).

Table 5. Item fit criteria

Criteria	Score
Outfit mean square (MNSQ)	0.5 < MNSQ < 1.5
Outfit Z-standard (ZSTD)	-2.0 < ZSTD < +2.0
Point Measure Correlation	0.4 < PT Measure Corr < 0.85

If the three criteria in Table 5 are met, the items are "appropriate", and the quality of the items is good and can be used. However, if only two criteria or one criterion are met, the items can be maintained and not be changed to be called "appropriate" items and can be used. Meanwhile, if it does not meet the criteria set in Table 5, the items are "inappropriate" and must be replaced or redesigned. The interpretation of the Item Fit test results is listed in Table 6.

Table 6. Item fit order test interpretation

Number	Item Code	Outfit		PT Measure Corr.	Status	Interpretation
		MNSQ	ZFTD			
1	P1	*3.08*	1.44	0.50	1 criterion	appropriate
2	P2	0.90	-0.06	0.66	-	appropriate
3	P3	0.89	-0.28	0.63	-	appropriate
4	P4	1.04	0.23	0.57	-	appropriate
5	P5	0.89	-0.33	0.56	-	appropriate
6	P6	0.56	-1.78	0.47	-	appropriate
7	P7	1.07	0.31	*0.14*	1 criterion	appropriate
8	P8	*0.47*	*-2.74*	*0.26*	3 criteria	inappropriate
9	P9	1.32	0.94	*0.30*	1 criterion	appropriate
10	P10	1.24	0.98	0.56	-	appropriate
11	P11	0.98	0.16	0.64	-	appropriate
12	P12	0.82	-0.45	0.54	-	appropriate
13	P13	*1.73*	*2.47*	0.68	2 criteria	appropriate

Table 6 shows that only one item does not meet any of the criteria (item 8), so it can be concluded that item 8 is invalid (Misfit) and must be removed. Meanwhile, the remaining items meet at least one criterion and thus are valid. Overall, the results show that prospective mathematics teachers' mathematical reflective thinking test instrument in this study was declared reliable and valid with "very reliable" criteria, and 12 out of the 13 test items are "valid."

After analyzing the test results, the final stage is revising based on the analysis results of the trial conducted. This final stage produces a final product of a mathematical reflective thinking test instrument which is ready to be socialized and used as an instrument in research.

Discussion

The previously described research shows that prospective mathematics teachers' mathematical reflective thinking test has gone through all the development steps. It has been declared valid and reliable based on the Rasch model analysis. The test was expected to produce a quality mathematical reflective thinking test instrument for prospective mathematics teachers. In the early stages of preparing the blueprint, the selection of material or courses relevant to the mathematical ability to be tested by this test instrument was carried out. The group theory course was chosen because it is compulsory for all prospective mathematics teachers. Furthermore, studying group theory material or algebraic structures could develop reflective mathematical thinking skills (Yenni & Sukmawati, 2019). The preparation of the items was done by previously setting operational definitions and indicators of mathematical reflective thinking skills used in this study. Furthermore, these indicators were converted into items so that they can be used to measure the mathematical reflective thinking ability of prospective mathematics teachers.

Based on the assessment results of two lecturers who teach group theory courses at different universities, some items were confusing or inaccurate and needed improvement. After the revisions, an assessment was carried out by an expert in mathematical ability and two professors who evaluated mathematical abilities. The assessment result showed that several sentences in the questions are not sharp and need improvement as an instrument for mathematical reflective thinking skills. In addition, a limited review was also carried out by five prospective mathematics teachers for the readability test. Revisions were made following input and suggestions from experts and students in a limited trial. After the appropriate results, the test instrument was feasible to be tested.

The test results were analyzed using the Rasch model. Wibisono (2018) stated that instruments validated with the Rasch model better meet the definition of measurement and produce more holistic information. The data analysis results showed the Cronbach's alpha (KR-20) of 0.71, with person reliability of 0.73 and items up to 0.88. It showed that the instrument met the reliable criteria. However, item P8 should be eliminated because it was considered redundant (similar to item P9 that was easier to understand). This item asked for an explanation of choice between relevant and irrelevant data.

The advantages of this research are that all the planned development steps have been carried out, the tests obtained have passed the expert judgment process and were tested on mathematics education students, and were analyzed using the Rasch model approach so that a valid and reliable test of mathematical reflective thinking skills for prospective mathematics teachers has been obtained. On the other hand, the drawback of this research is that the test produced is limited by certain mathematical material, in this case, the material in the group theory course, so that the use of this test is also limited to prospective mathematics teacher students who have taken the course or have studied group theory material.

This study has similarities with research conducted by Nindiasari et al. (2016) regarding the development of teaching materials and instruments to improve mathematical reflective thinking based on a metacognitive approach in high school students (SMA); Hendriana et al. (2019) who also developed an instrument for the reflective thinking ability of high school students, as well as Muntazhimah's research (2019) regarding the development of a mathematical reflective thinking test instrument for 8th-grade junior high school students. Based on the several studies that have been mentioned, it is still very difficult to find references to previous research on the development of mathematical reflective thinking tests for prospective mathematics teacher students; and the overall research on developing test instruments that have been carried out is still using classical test theory. Therefore, this study aims to develop a reflective thinking ability test instrument for prospective mathematics teacher students whose validity and reliability use the Rasch model.

The implications of the results of this study theoretically can add to the theory or literature related to how to develop a test instrument that can measure the reflective thinking ability of prospective mathematics teacher students. Practically, the test produced from this research can measure the reflective thinking ability of prospective mathematics teacher students. Learning mathematics is not only the reflective thinking ability that needs to be measured. Critical thinking skills, creative, geometric thinking skills, and other mathematical thinking skills also need to be measured. Therefore, the development of other mathematical thinking test instruments can then be carried out by referring to the current situation of a disruption of education.

Conclusion

The research on developing the mathematical reflective thinking test instrument for prospective mathematics teachers has gone through development procedures: preparing a blueprint, writing items, reviewing items, conducting trials, analyzing test results, and making revisions. The Rasch model approach was used to test the reliability and validity through the Winstep 4.4.3 software. After conducting an in-depth analysis, the results showed that the mathematical reflective thinking test instrument was declared reliable and valid with "very reliable" criteria, and 12 of the 13 test items were declared "valid." So, the mathematical reflective thinking test instrument for prospective mathematics teachers can be used in further research.

Acknowledgment

The authors would like to thank the research institute of Universitas Muhammadiyah Prof. Dr. Hamka who has provided funding for this research through an internal funding scheme, with the decree No. 695/F.03.07/2019.

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely by the authors.

References

Agustan, S., Juniati, D., & Siswono, T. Y. E. (2017). Reflective thinking in solving an algebra problem: A case study of field independent-prospective teacher. *Journal of Physics: Conference Series*, *893*(1), Article 012002. https://doi.org/10.1088/1742-6596/893/1/012002

Alavi, K., Isa, K., & Palpanadan, S. T. (2020). Application of Rasch model on resilience in higher education: An examination of validity and reliability of Malaysian academician happiness index (MAHI). *International Journal of Higher Education*, *9*(4), 261–271. https://doi.org/10.5430/ijhe.v9n4p261

Amidu, AR. (2012). Exploring real estate students' learning approaches, reflective thinking and academic performance. *48th ASC Annual International Conference Proceedings*, *June 2012*. http://ascpro0.ascweb.org/archives/cd/2012/paper/CEUE214002012.pdf

An, M., & Yu, X. (2021). A Rasch analysis of emerging adults' health motivation questionnaire in higher education context. *PLoSONE*, *16*(3), 1-13, Article e0248389. https://doi.org/10.1371/journal.pone.0248389

Arifin, Z. (2016). *Evaluasi pembelajaran: Prinsip, teknik, dan prosedur [Learning evaluation: Principles, techniques and procedures]*. PT Remaja Rosdakarya.

Boone, W. J., Staver, J. R., Yale, M. S., Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Item measures. In *Rasch analysis in the Human Sciences* (pp. 93–110). Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4_5

Clarà, M. (2015). What is reflection? Looking for clarity in an ambiguous notion. *Journal of Teacher Education*, *66*(3), 261–271. https://doi.org/10.1177/0022487114552028

Hendriana, H., Putra, H. D., & Hidayat, W. (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*(12), Article em1790. https://doi.org/10.29333/ejmste/112033

Muin, A., Novianti, L., & Musyrifah, E. (2018). Analysis of mathematical reflective thinking skills based on learning model and mathematical prior knowledge. *3rd International Conferences on Education in Muslim Society (ICEMS 2017)*. https://doi.org/10.2991/icems-17.2018.5

Muntazhimah. (2019). Pengembangan instrumen kemampuan berpikir reflektif matematis siswa kelas 8 SMP [Development of mathematical reflective thinking instruments for 8th grade middle school students]. *Imajiner: Jurnal Matematika dan Pendidikan Matematika*, *1*(5), 237–242. https://doi.org/10.26877/imajiner.v1i5.4551

Nindiasari, H., Novaliyosi, N., & Pamungkas, A. S. (2016). Pengembangan bahan ajar untuk meningkatkan tahapan kemampuan berpikir matematis [Development of teaching materials to improve the stages of mathematical thinking skills]. *JPPM*, *9*(1), 109–114.

Nur, L., Nurani, L. A., Suryana, D., & Ahmad, A. (2020). Rasch model application on character development instrument for elementary school students. *International Journal of Learning, Teaching and Educational Research*, *19*(3), 437–459. https://doi.org/10.26803/ijlter.19.3.24

Osnal, Suhartoni, & Wahyudi, I. (2015). Meningkatkan kemampuan guru dalam menyusun tes hasil belajar akhir semester melalui workshop di KKG Gugus 02 Kecamatan Sumbermalang tahun 2014/2015 [Improving teacher ability in preparing final semester learning outcomes tests through workshops at KKG Cluster 02 Sumbermalang District 2014/2015]. *Pancaran Pendidikan*, *5*(1), 67–82. https://jurnal.unej.ac.id/index.php/pancaran/article/view/2604

Prabowo, A., & Dahlan, J. A. (2020). Pengembangan tes matematika dengan konteks COVID-19 untuk siswa SMP/MTs kelas VIII [Development of mathematics tests with the context of COVID-19 for class VIII SMP/MTs students]. *Jurnal Elemen*, *6*(2), 302–317. https://doi.org/10.29408/jel.v6i2.2115

Prabowo, A., Kusdinar, U., & Rahmawati, U. (2018). Pelatihan pengembangan instrumen tes mata pelajaran matematika SMP [Training for development of middle school mathematics subject test instruments]. *International Journal of Community Service Learning*, *2*(3), 141–148. https://doi.org/10.23887/ijcsl.v2i3.14189

Professional, T. I. (2006). *What are the steps in the development of an exam program?* http://www.proftesting.com/test_topics/steps.php

Rahmadhani, E., Wahyuni, S., Mandasari, L., & Setiawan, Y. (2020). Reflective thinking ability of teacher candidate students based on ability level and gender. *Companion Proceedings of the SEADRIC 2019 (2020)*, 165–172. https://doi.org/10.24071/seadr.2019.23

Rowe, S. E. (2001). *Development of a test blueprint for the National Association of Industrial Technology certification exam [Unpublieshed Doctoral Dissertation]*. Iowa State University. https://lib.dr.iastate.edu/rtd/668

Sabekti, A. W., Khoirunnisa, F., & Liliasari, L. (2020). Validating the Indonesian version of reflective thinking questionnaire and investigation of the relationship between pre-service teachers’ reflective thinking and academic achievement. *Companion Proceedings of the SEADRIC 2019 (2020)*, 138–144. https://doi.org/10.24071/seadr.2019.19

Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly*, *3*(1), 71–79. https://doi.org/10.1207/s15434311laq0301_5

Suharman. (2018). Tes sebagai alat ukur prestasi akademik [Test as a measuring tool for academic achievement]. *Jurnal Ilmiah Pendidikan Agama Islam*, *10*(1), 93–115.

Sumintono, B. & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan [Rasch modeling applications for educational assessment]*. Trim Komunikata.

Wahyudi, W. (2012). Assesment pembelajaran berbasis portofolio di sekolah [Portfolio-based learning assessment in schools]. *Jurnal Visi Ilmu Pendidikan*, *2*(1), 288–297. https://doi.org/10.26418/jvip.v2i1.370

Wardhani, D., & Putra, A. (2016). Pengembangan instrumen tes standar kognitif pada mata pelajaran IPA kelas 7 SMP di Kabupaten Banjar [Development of cognitive standard test instruments for science subjects for grade 7 junior high schools in Banjar regency]. *Proceeding Biology Education Conference: Biology, Science, Enviromental, and Learning*, *13*(1), 75-82. https://jurnal.uns.ac.id/prosbi/article/view/5658/5026

Wibisono, S. (2018). Aplikasi model Rasch untuk validasi instrumen pengukuran fundamentalisme agama bagi responden muslim [Rasch model application for validation of religious fundamentalism measurement instruments for muslim respondents]. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia (JP3I)*, *5*(1), 1-29. https://doi.org/10.15408/jp3i.v5i1.9239

Yenni, Y., & Sukmawati, R. (2019). Analisis kemampuan berpikir reflektif matematis berdasarkan minat belajar pada mata kuliah struktur aljabar [Analysis of mathematical reflective thinking ability based on learning interest in algebra structure course]. *Teorema: Teori dan Riset Matematika*, *4*(2), 75-82. https://doi.org/10.25157/teorema.v4i2.2283

Yuen Lie Lim, LA. A. (2011). A comparison of students' reflective thinking across different years in a problem-based learning environment. *Instructional Science*, *39*(2), 171–188. https://doi.org/10.1007/s11251-009-9123-8