JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Jurnal Eduscience (JES)
Volume 11, No. 2
August, Year 2024
Submit: 22 June 2024
Accepted: 13 August 2024

ITEM ANALYSIS OF THE CRITICAL THINKING TEST INSTRUMENT ON TEMPERATURE AND HEAT USING RASCH MODELLING

ARUM ALYAMUARI[1], WINNY LILIAWATI[2], DUDEN SAEPUZAMAN[3]

[1,2,3]Departement of Physics Education, FPMIPA, Universitas Pendidikan Indonesia
arumal24@upi.edu

Abstract

The purpose of this study was to analyse the test items of the critical thinking test instrument using Rasch modelling. This research is a quantitative research. The sample determination used simple random sampling method and the subjects obtained were class XII students at one of the high schools in Bandung city consisting of 36 students. The research instrument used was a critical thinking ability test question consisting of 10 description questions. The results showed that the value of the difficulty level of the items was in the range of -2.45 to 1.86 and the average SEM was 0.31. Meanwhile, item reliability was 0.93, person reliability was 0.90, and Cronbach's alpha was 0.92. Based on the results of this study, it can be concluded that overall the items are declared fit with the Rasch model. Therefore, the critical thinking test instrument can be used by researchers to measure students' critical thinking skills.

Keywords: Item Analysis; Rasch Model; Critical Thinking Test.

Abstrak

Tujuan penelitian ini adalah menganalisis butir soal uji coba instrumen tes kemampuan berpikir kritis menggunakan pemodelan rasch. Penelitian ini merupakan penelitian kuantitatif. Penentuan sampel menggunakan metode simple random sampling dan subjek yang diperoleh yaitu peserta didik kelas XII pada salah satu SMA di kota Bandung yang terdiri dari 36 peserta didik. Instrumen penelitian yang digunakan berupa soal tes kemampuan berpikir kritis terdiri dari 10 butir soal uraian. Hasil penelitian menunjukkan bahwa nilai tingkat kesukaran butir soal berada pada rentang -2.45 sampai dengan 1.86 dan rerata SEM sebesar 0.31. Sementara itu, item reliability senilai 0.93, person reliability senilai 0.90, dan Alpha Cronbach senilai 0.92. Berdasarkan hasil penelitian ini dapat disimpulkan bahwa secara keseluruhan butir soal dinyatakan fit dengan model rasch. Oleh karena itu, instrumen tes kemampuan berpikir kritis dapat digunakan oleh peneliti untuk mengukur kemampuan berpikir kritis peserta didik.

Kata Kunci: Analisis Butir Soal; Rasch Model; Tes Kemampuan Berpikir Kritis.

INTRODUCTION

The implementation of learning in the 21st century is expected to be able to contribute and realise human resources in order to survive in all dynamics by requiring human preparation to deal with it. Therefore, the world of education is faced with a number of challenges to produce human resources who have the skills to be ready to compete in the era of globalisation. In the 21st century, students are required

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

to have 4C abilities, namely critical thinking, creative thinking, communication, and collaboration (Septikasari & Frasandy, 2020). The goal of teaching and learning in the 21st century is to find effective teaching methods and create a curriculum that suits the demands of the 21st century (Ssemugenyi, 2023). One form of creating qualified human resources with the demands of the 21st century learning process is to train certain skills in students. The skills needed are in accordance with the 21st century framework, learners need to have the skills to reason that focus on determining what to believe and do, known as critical thinking skills (Suciono, 2021).

Critical thinking skills can be measured and trained to students through three stages of learning, namely the planning stage, the implementation stage, and the evaluation stage. Before carrying out the implementation stage, the teacher conducts the planning stage first. One of the teacher's activities at the planning stage is preparing test instruments for the evaluation stage. The test instrument is an assessment technique in which students are asked to complete various items that must be done to produce a score of the students' abilities (Sawaluddin & Muhammad, 2020). Therefore, to obtain accurate information related to the ability of students, a fit test is needed.

Item analysis is diagnostic information related to whether students understand what has been learnt and to improve the quality of questions by correcting or eliminating invalid questions (Fauziana & Wulansari, 2021). This study aims to detect the quality of the test items of the critical thinking ability test instrument by analysing the validity, reliability, and difficulty level of the items tested in the form of information about student abilities and the quality of the questions given with the Rasch model. The use of Rasch modelling on raw data from the test results of the instrument test in this study has the aim of obtaining a rating scale using the same interval so that it can show accurate data, both the ability of students and the quality of the questions given to students (Erfan et al., 2020). The Rasch model is the result of the development of the classical test theory (CTT) and item response theory (IRT) models response theory). The Rasch model fulfils the 5 principles of measurement models, namely: 1) Provides linear data, predicts missing data; 2) Is precise in estimating data; 3) Detects the inaccuracy of a measurement model estimating data; 4) Detecting the inaccuracy of a model; 5) Producing replicable data (Chan, Ismail et al., 2011). The advantage of the Rasch model is that it is able to provide more precise estimates and detect model inaccuracies so as to produce replicable measurements (Sumintono & Widhiarso, 2015).

RESEARCH METHODS

This research is quantitative research. The research data were obtained from the results of the test questions of the research instrument by determining the sample using the simple random sampling method. The sample size was 36 students in class XII at one of the high schools in Bandung city consisting of 7 male students and 26 female students. The data collection instrument is in the form of 10 questions on the description of temperature and heat material that refers to the indicators of R.Ennis' critical thinking skills.

Analysis of test items in this study was carried out using Rasch modelling. Rasch modelling provides accurate diagnostic information when testing instruments (Fitri, 2017). This study did not use the classical model because the items of the classical model are inconsistent or can change depending on the ability of the respondents or participants (Yusuf et al., 2021).

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

The analysis in this study was carried out using winsteps software. Based on the results of the data on the test questions of the critical thinking ability test of students, then the data is processed into Microsoft Excel and then analysed with the help of winsteps software so that the criteria in the Rasch model analysis are obtained, starting from model fit, unidimensionality, item fit, difficulty level, and item parameters.

RESULT AND DISCUSSION

Result

The critical thinking test instrument test on temperature and heat material was analysed with the help of winsteps software so as to produce item parameter output and estimation of test takers' abilities. Unidimensionality can be shown if there are no problematic items then the raw variance> 20% and the observed value in unexplained < 15% to show the appropriate item (Sumintono & Widhiarso, 2015). Based on the results of the unidimensionality analysis, the raw variance is 65.1% and the observed value in the unexplained variance 1 contrast shows a value of 8.8%, so it can be concluded that there are no problematic items and each item is able to measure the range of abilities of students comprehensively. Data analysis using Rasch modelling on Winsteps summary statistic software, obtained important information presented in Table 1.

Table 1. Summary Statistics Analysis Results

	Logit Mean (SD)	Separation	Reliability	Alpha Cronbach	Outfit MNSQ	Outfit ZSTD
Person	0.57 (2.08)	3.08	0.90	0.92	0.97	-0.17
Item	0.00 (1.25)	3.70	0.93		0.97	-0.15

Based on the results of the instrument analysis in Table 1, it is known that the average logit person is 0.57, while the average logit item is 0.00. This shows that the average logit person is greater than the average logit item, thus proving that the ability of students in general is greater than the difficulty of the instrument items. Furthermore, regarding separation or grouping of person and item. Separation shows how good a set of items on the critical thinking ability instrument test is in spreading the logit ability range. The higher the separation value, the better the instrument is made. This shows that the items in it can reach individuals with high to low level abilities.

The consistency of answers from students (person reliability) has a value of 0.90 and the quality of instrument items (item reliability) with a value of 0.93. In addition, the Cronbach Alpha value for the critical thinking skills test has a value of 0.92. This shows that there is a match between the items and the person (students), so it can be concluded that the critical thinking ability test on temperature and heat material can be said to be reliable.

Regarding the average outfit Mean Squared (outfit MNSQ) value of 0.97 in the person and item columns, it shows that the value is classified as a fit category, which is located between 0.5 < MNSQ < 1.5. This proves that the test instrument used is in accordance with the model to measure students' critical thinking skills. In addition, it is known that the average outfit Z Standardised (outfit ZSTD) value is -0.17 for person and -0.15 for item. Both values are in the range -2 < ZSTD < 2, meaning that overall the items are in

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

accordance with the Rasch model and can be used as instruments to measure students' critical thinking skills in temperature and heat material.

Analysis of each item declared not fit or misfit can be seen in Figure 3. The item is said to fit the model if the item that is said to be suitable is an item that has an outfit mean square value of 0.5 < MNSQ < 1.5, an outfit Z Standardized value of -2.0 < ZSTD < 2.0, and the Point Measure Correlation (Pt Mean Corr) value is not negative (Bambang Sumintono, Wahyu Widhiarso, 2015). If the item does not meet the three requirements of the fit model, the instrument can be used with a note that it is replaced or revised. The following results of the item parameter analysis can be seen in Table 2.

Table 2. Results of Question Item Parameter Analysis

Item	Measure	SEM	Outfit MNSQ	Outfit ZSTD	Point Measure Correlation
S1	-0.98	0.31	1.45	1.75	0.87
S2	0.12	0.30	0.69	-1.53	0.83
S3	0.39	0.30	1.11	0.56	0.67
S4	0.12	0.30	0.51	-2.69	0.80
S5	1.86	0.31	1.05	0.28	0.68
S6	1.02	0.30	1.00	0.09	0.80
S7	0.93	0.30	0.83	-0.73	0.81
S8	-2.45	0.34	0.54	-1.49	0.84
S9	0.66	0.30	1.42	1.78	0.60
S10	-1.68	0.32	1.10	0.46	0.75

Based on Table 2, there is one item that does not fulfil all three model fit requirements. Outfit ZSTD on question number 4 shows a value of -2.69 (outside the range of outfit ZSTD requirements) which means question number 4 is not compatible with the expected model. However, item number 4 met the MNSQ outfit and Pt Measure Corr requirements, with an MNSQ outfit value of 0.51 and a Pt Measure Corr value of 0.80. By considering other item fit requirements, question number 4 will still be used and also overall the items on the validity of the instrument have an average infit outfit ZSTD value that is in accordance with the fit requirements put forward by Sumintono & Widhiarso (2015). Therefore, all test questions in the critical thinking test instrument will still be used with a note that question number 4 is corrected so that it can be compatible with the expected model. Items with negative Pt Measure Corr values indicate misleading items because low-ability test takers are able to answer items correctly and high-ability test takers answer incorrectly (Bambang Sumintono, Wahyu Widhiarso, 2015). Items with negative correlation values should be checked to see if the answer key is incorrect, needs to be revised, or removed. Because Pt Measure Corr of all items on the critical thinking test instrument is not negative, the research will use all items to test students' critical thinking skills.

The level of accuracy of item measurement or standard error measurement (SEM) shows a good discrimination value if SEM < 0.5 logits. In column 3 of Table 2, it shows that overall the items have SEM < 0.5 logits with an average of 0.31 so it can be said that the items are "good" in measurement accuracy. In addition, the measure is shown in Table 2 column 2. According to Sumintono & Widhiarso (2015) the level of difficulty of the items has a range of -2 < measure < 2. Items are categorised as easy if they have a value close to -2.00 logits, items are categorised as medium if -1.00 logits < measure < +1.00 logits, and items are categorised as difficult if the value is close to +2.00 logits (Ayala, 1993).

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Based on the results of logit analysis of critical thinking ability instruments, it is known that there are 20% of items that fall into the category of difficult difficulty, 50% of items are in the medium category, 20% of items are in the easy category, and 10% of items are in the very easy category. The average difficulty level of all items is 0.00 logits with a standard deviation of 1.25 logits. Measure on item number 8 has a logit of -2.45 which indicates that item number 8 is in the category of very easy questions and item number 5 has the highest logit among other items, which is 1.86 which indicates that item number 5 is in the difficult category.

The Guttman scalogram shows the pattern of learners' responses to the item in terms of item difficulty. The pattern of learners' responses is sorted based on the highest to the lowest learner ability. In addition, the item difficulty is shown with the left to the right showing the items that have the easiest difficulty to the most difficult items.

```
Person |Item
       |  1
       |8012439765
       |----------
    32 +4444433443  32P
     2 +4444333433  02P
     8 +4444334333  08P
    33 +4443333433  33L
    34 +4443333433  34L
    36 +4333344343  36L
    29 +4443433332  29P
    31 +4443333342  31P
    16 +4443333332  16L
     7 +4443332332  07P
    30 +4443333232  30P
    15 +3443333232  15P
    17 +4443333222  17P
    26 +3343333233  26P
     4 +4443232223  04P
    13 +3443332322  13L
    12 +3343332322  12P
    19 +4333333222  19P
    18 +4432222332  18P
    20 +4332332322  20P
    23 +3243332223  23P
     1 +3214314341  01P
     3 +4332223223  03P
     5 +4233323222  05P
    10 +3333231232  10P
     9 +3433221213  09P
    24 +3411213222  24P
    35 +3322222221  35P
    11 +3222231212  11L
    14 +3222232121  14P
     6 +3222231111  06L
    28 +2212212122  28P
    21 +2211222211  21P
    22 +2211222211  22P
    27 +2212212112  27P
    25 +2211212111  25P
       |----------
       |  1
       |8012439765
```

Figure 1. Guttman scalogram

Based on Figure 1, it can be seen that question number 8 which is included in the category of very easy questions can be answered by all students. However, there are some error responses from students, one of which is students with code 9P. Where these learners can answer the easiest question (question number 8) and answer the most difficult question (question number 5) with the same score, but on questions that have a medium level of difficulty, namely question number 9, these learners get a low score. Previously, based on

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

the results of summary statistical analysis, the ability parameters of students who had conducted a critical thinking ability instrument trial test on temperature and heat material were obtained. Measuring the estimated ability of students using the critical thinking ability test instrument can be seen based on the results of the output analysis of students' abilities in Table 3 and the distribution of students' abilities in the form of a histogram presented in Figure 2.

Table 3. Description of Learners' Ability

Deskripsi	Nilai
Mean	0.57
Maksimal	4.26
Minimal	-3.78
Standar Deviasi	2.08

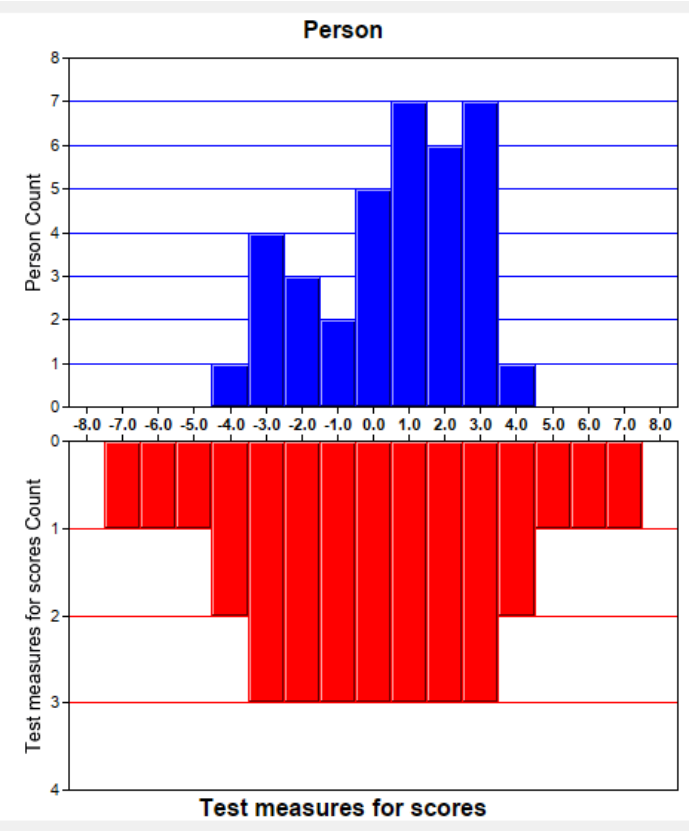

Figure 2. Distribution Histogram of Learners' Ability

Based on Figure 2, it can be seen that the distribution of students' ability to test critical thinking skills on temperature and heat material is close to a normal curve. The measurement scale of students' abilities can also be seen on the wright map. The right part of the Wright map shows the distribution of the difficulty level of each item, while the left part shows the distribution of students' abilities. In addition, the wright map can analyse the level of ability of students in working on critical thinking skills trial questions. The following winsteps wright map output results are presented in Figure 3.

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

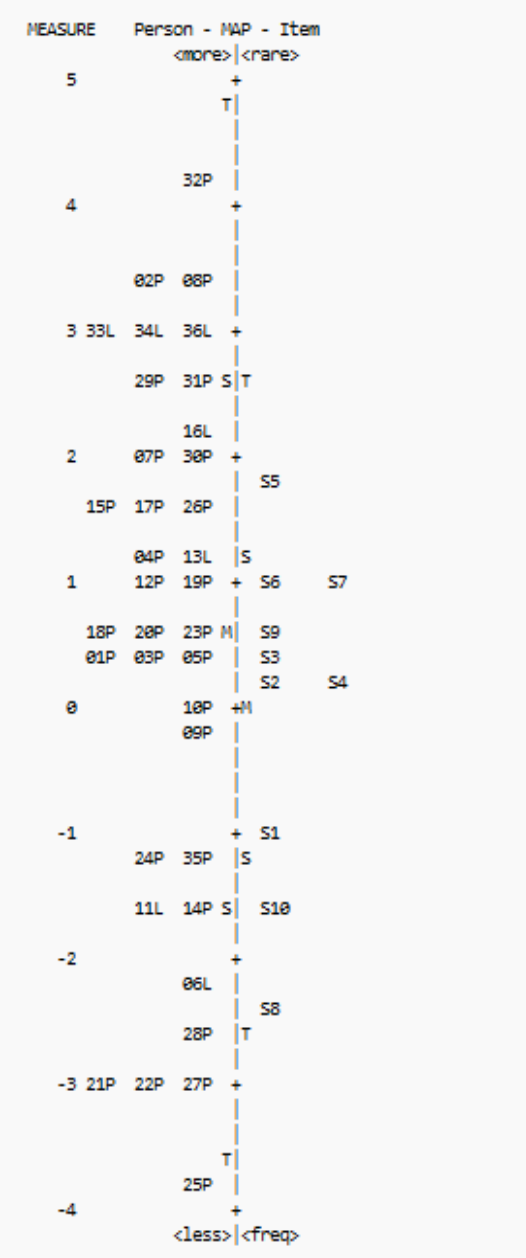

Figure 3. Wright Map

Figure 3 shows that learners with code 32P have a higher ability level than other learners and learners with code 25P have a lower ability than other learners. In addition, it can be seen in Figure 5 that learners with codes 30P, 7P, 16L, 31P, 29P, 36L, 34L, 33L, 8P, 2P, and 32P can answer all test questions because the ability of these learners is above the item that is included in the difficult category, namely question number 5. The characteristics of the critical thinking ability test items are presented in the form of an Item Characteristic Curve (ICC) graph in Figure 4.

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

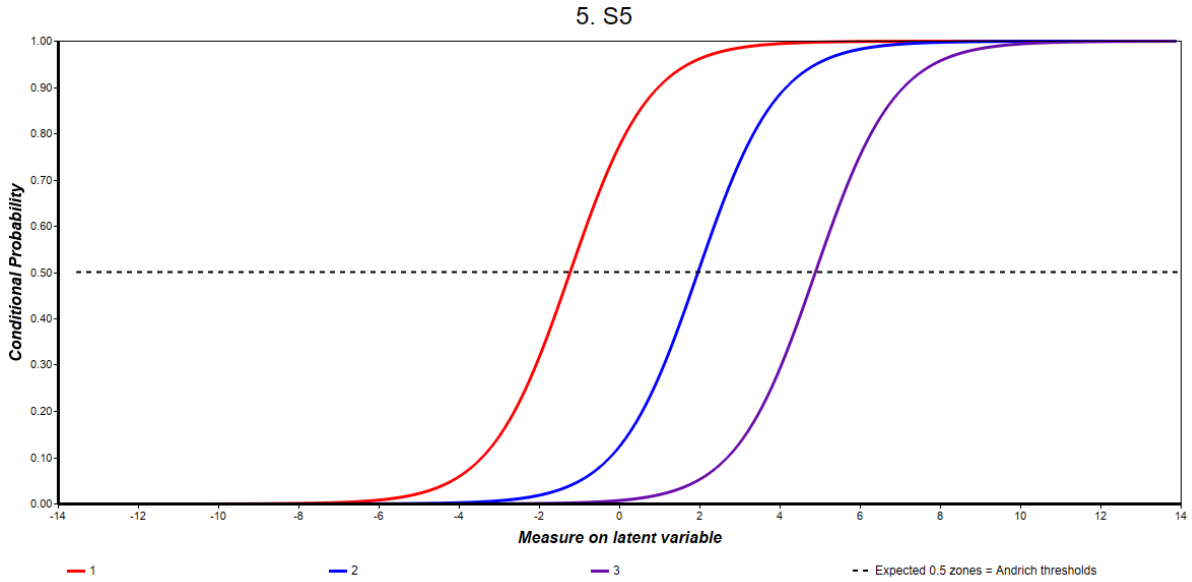

Figure 4. ICC Item Number 5 Critical Thinking Ability Test Instrument

Figure 4 shows the relationship between the ability of students and the probability of students answering correctly on item number 5. It can be seen that the shape of the ICC graph is S-shape which shows the relationship between the probability of answering an item correctly in the range of abilities measured (Santoso, 2018). Curves 1, 2, and 3 move further to the right indicating the higher the ability of the learners, the higher the chance of answering correctly. Item parameters for scoring can be described by a Category Characteristic Curve (CCC). The following CCC on item number 5 is presented in Figure 5.

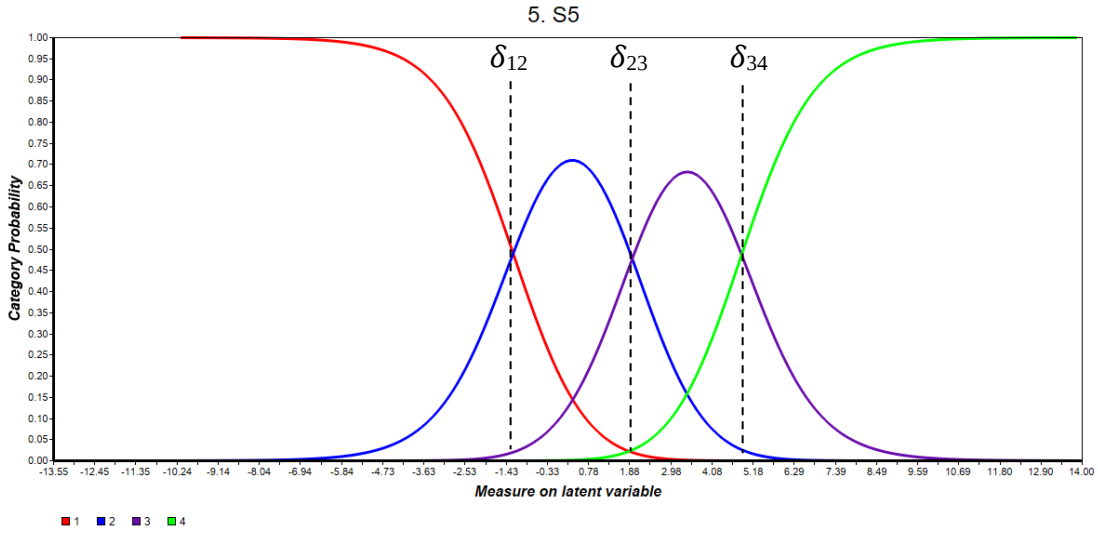

Figure 5. CCC of Item Number 5 Critical Thinking Ability Test Instrument

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Based on Figure 5, the intersection point between the red line and the blue line, shows the step parameter ($\delta 12$) of the test taker to get a score of 1. In Figure 5 there is a dotted black line showing the ability value of -1.0175 with a chance of 0.50. This means that test takers who have abilities below -1.0175 have a great chance of getting a score of 0 with opportunities ranging from 0.50 to 1 and the chance of getting a score of zero can decrease as the ability increases. Test takers who have an ability of -1.0175 are able to get a score of 0 or 1 with a chance of 0.50. Someone whose ability is above -1.0175 to 1.88 has the largest chance of 0.50 to get the value 1. Likewise for the second step parameter ($\delta 23$) and the third step parameter ($\delta 34$).

Related to the information function related to stating the strength of an item on a test instrument. The information function and Standard Error of Measurement (SEM) have a quadratically inverse relationship, namely the greater the information function, the smaller the SEM it has. The cut between the information function curve and SEM is seen as the feasibility of the instrument in measuring the ability of students (Retnawati, 2017). Outside this range, the critical thinking ability test on temperature and heat material will not be suitable for students because the SEM value is greater than the information function value. The following is a graph of the information function of the critical thinking test instrument, presented in Figure 6.

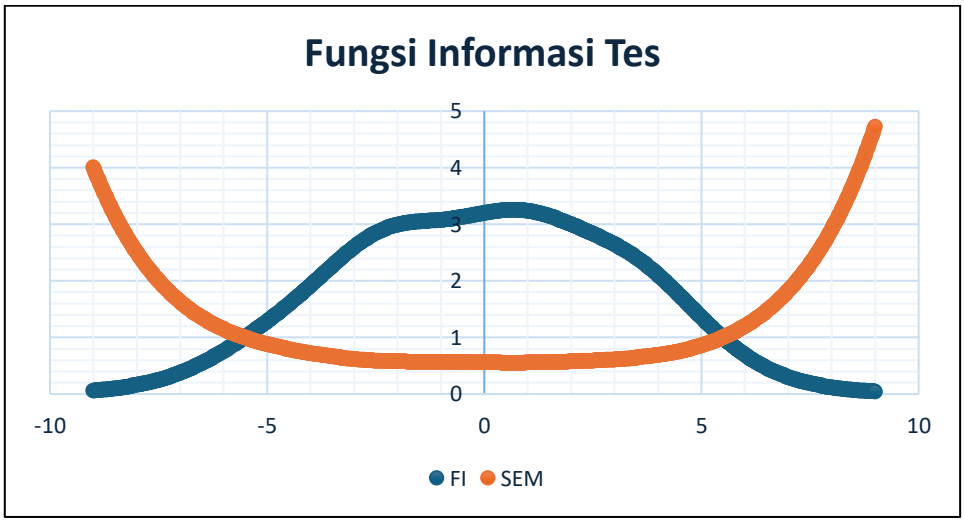

Figure 6. Information function of Critical Thinking Ability Instrument

Based on Figure 6, the Test Information Function (TIF) curve has a peak point or the highest information at around 0.70 logits. This indicates that the instrument test information is greatest when used or tested on students who have abilities around 0.70 logits.

Discussion

Items are said to fit the model if the items that are said to be suitable are items that have an outfit mean square value of 0.5 < MNSQ < 1.5, an outfit Z Standardised value of -2.0 < ZSTD < 2.0, and the Point Measure Correlation (Pt Mean Corr) value is not negative (Sumintono & Widhiarso, 2015). If the item does not fulfil the three fit model requirements, the instrument can still be accepted by considering other fit model requirements.

JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

There is one item that does not fulfil all three fit model requirements. Outfit ZSTD on question number 4 shows a value of -2.69 (outside the range of outfit ZSTD requirements) which means question number 4 is not compatible with the expected model. However, item number 4 met the MNSQ outfit and Pt Measure Corr requirements, with an MNSQ outfit value of 0.51 and a Pt Measure Corr value of 0.80. By considering other item fit requirements, question number 4 will still be used and also overall the items on the validity of the instrument have an average infit outfit ZSTD value that is in accordance with the fit requirements put forward by Sumintono & Widhiarso (2015). Therefore, all test questions in the critical thinking test instrument will still be used with a note that question number 4 is corrected so that it can be compatible with the expected model.

Based on the results of the logit analysis of the critical thinking ability instrument, it is known that there are 20% of the items that fall into the category of difficult difficulty, 50% of the items are in the medium category, 20% of the items are in the easy category, and 10% of the items are in the very easy category. The average difficulty level of all items is 0.00 logits with a standard deviation of 1.25 logits. Measure on item number 8 has a logit of -2.45 which indicates that item number 8 is in the very easy question category and item number 5 has the highest logit among other items, which is 1.86 which indicates that item number 5 is in the difficult category.

Regarding the information function related to stating the strength of an item on a test instrument, the Test Information Function (TIF) curve has a peak point or the highest information at around 0.70 logits. This indicates that the instrument test information is greatest when used or tested on students who have abilities around 0.70 logits. The intersection points of the two curves are at teta -5.51 and +5.53. In this range of teta, the value of the test information function is greater than the standard error. A test can be trusted when the value of the information conveyed is more dominant than the error (Arlinwibowo et al., 2021). So it can be concluded that the instrument is reliable for students who have abilities of -5.51 logits to +5.53 logits.

CONCLUSION

The results of the analysis of the test items of the critical thinking ability test instrument fit the Rasch model. This is indicated by the suitability of the items with the three fit model requirements, namely the average MNSQ output value of 0.97, Outfit ZSTD -0.15, and all Pt Measure Corr is not negative. Meanwhile, item reliability is 0.93, person reliability is 0.90, and Cronbach's Alpha is 0.92. Based on these values, it shows that the critical thinking ability test instrument shows that there is a match between the items and the person (students), so it can be concluded that the critical thinking ability test on temperature and heat material can be said to be reliable and can be used as a test instrument to measure students' critical thinking skills on temperature and heat material.

REFERENCES

Arlinwibowo, J., Retnawati, H., & Kartowagiran, B. (2021). Item Response Theory Utilisation for Developing the Student Collaboration Ability Assessment Scale in STEM Classes. Ingenierie Des Systemes d'Information, 26(4), 409-415. https://doi.org/10.18280/isi.260409.

Chan, S. W., Ismail, Z., &Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. Procedia -Social and Behavioral Sciences, 129.

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Tursina, R. (2020). Quality analysis of the ability to distinguish series and parallel circuits through classical test theory and Rasch model. Indonesian Journal of Educational Research and Review, 3(1). https://doi.org/10.23887/ijerr.v3i1.24080.

Fauziana, A., & Wulansari, A. D. (2021). Analysis of the Quality of Daily Test Items in Elementary Schools with the Rasch Model. Journal of Science-Based Islamic Primary Education. https://doi.org/10.21154/ibriez.v6i1.112.

Fitri. (2017). Analysis of Validity and Reliability of Accountant Performance Instruments Using the Rasch Model Approach. Scientific Journal of Accounting Civilisation.

Retnawati, H. (2017). Reliability of Research Instruments. Journal of Mechanical Engineering Education Unnes. Journal of Mechanical Engineering Education Unnes, 12(1).

Santoso, A. (2018). Test Item Characteristics of Introduction to Social Statistics. Journal of Maths and Science Education, 6(2). https://doi.org/10.21831/jpms.v4i1.10111.

Sawaluddin, & Muhammad, S. (2020). Steps and techniques for evaluating learning outcomes in Islamic Religious Education. Journal of PTK and Education, 6(1). https://doi.org/10.18592/ptk.v6i1.3793.

Septikasari, R., & Frasandy, R. N. (2020). 4C 21st Century Skills in Basic Education Learning. Journal of Basic Islamic Education, 75(20), 2635-2638. https://doi.org/10.1016/j.jacc.2020.04.015.

Ssemugenyi, F. (2023). Teaching and learning methods compared: A pedagogical evaluation of problem-based learning (PBL) and lecture methods in developing learners' cognitive abilities. Cogent Education, 10(1). https://doi.org/10.1080/2331186X.2023.2187943.

Suciono, W. (2021). Critical Thinking (Kodri (ed.); Pe mould). Adab Publisher.

Sumintono, B., & Widhiarso, W. (2015). Application of Rasch Modelling in Educational Assessment. Trim Komunikata.

Yusuf, S., Budiman, N., Yudha, E. S., Suryana, D., & Yusof, S. M. (2021). Rasch Analysis of the Indonesian Mental Health Screening Tools. The Open Psychology Journal, 14(1). https://doi.org/10.2174/1874350102114010198.