

Construct the Validity of STEM and Project-based Critical Thinking Skills Test Instruments Using the Rasch Model

Rina Dwi Setyawati¹, Agnita Siska Pramasdyahsari², Binod Prasad³, Sindi Nur Aini⁴

^{1,2,4}*Universitas PGRI Semarang, Indonesia*

³*Kathmandu University, Nepal*

Abstract

Students' critical thinking skills are very important abilities in the 21st century. To measure critical thinking skills, a valid and reliable instrument is needed. This article uses the Rasch model to construct an instrument for critical thinking skills on number pattern material whose learning uses STEM PjBL. The development method uses the analysis, design, development, implementation, and evaluation (ADDIE) stages. Data processing uses the Rasch model assisted by the Winstep program. The instrument was tested on 33 public junior high school students in Semarang. The research results include validity, namely the validity test by experts with an average percentage of 92.4% in the very valid category, the quality of the questions in the instrument reliability aspect is good, and the average critical thinking ability of students who use STEM-based and Project Based instruments Learning was 61.46, higher than other classes at 49.79. The critical thinking test instrument is based on STEM and project needs to be further developed on other materials.

Keywords: Construct Validity, STEM, Critical Thinking Skills, Rasch Model

Membangun Validitas Instrumen Tes Keterampilan Berpikir Kritis Berbasis STEM dan Proyek Menggunakan Model Rasch

Abstract

Kemampuan berpikir kritis siswa merupakan kemampuan yang sangat penting di abad 21 ini. Untuk mengukur kemampuan berpikir kritis diperlukan instrumen yang valid dan reliabel. Tujuan penelitian ini adalah mengembangkan instrumen tes kemampuan berpikir kritis yang berbasis STEM dan proyek. Soal terdiri dari 15 item materi aritmatika sosial yang memuat indikator *Focus, Reason, Inferent, Situation, Clarity dan Overview*. Metode pengembangan menggunakan tahapan *analysis, design, development, implementation, and evaluation* (ADDIE). Pengolahan data menggunakan model *Rasch* berbantuan program *Winstep*. Instrumen diujicobakan pada 33 siswa SMP Negeri di Semarang. Hasil penelitian

mencakup valid yaitu pada uji validitas oleh ahli dengan rata-rata persentase sebesar 92,4 % dalam kategori sangat valid, kualitas butir soal pada aspek reliabilitas instrumen baik, dan rata-rata kemampuan berpikir kritis siswa yang menggunakan instrument berbasis STEM dan proyek sebesar 61,46 lebih tinggi dari kelas lain sebesar 49,79. Instrument tes berpikir kritis didasarkan pada STEM dan proyek perlu dikembangkan lebih lanjut pada materi-materi lainnya.

Kata kunci: *Membangun Validitas, STEM, Kemampuan Berpikir Kritis, Model Rasch*

INTRODUCTION

Critical thinking skills are indispensable in 21st century learning (Basri, 2019; Lamb, 2017; Cahyono et al., 2021; Changwong, 2018). Critical thinking is a reflective thinking ability that focuses on patterns of decision making about what to believe, what to do, and what can be proven true. (Ennis, 2011; Cahyono, 2019). Critical thinking abilities can be identified through their main elements, namely; Focus, Reason, Inference, Situation, Clarity, and Overview (FRISCO) (Ennis, 1996; Cahyono, 2017). In the 21st century learning process, instruments are needed to measure critical thinking skills that are valid, reliable and innovative.

Instruments to measure critical thinking skills based on STEM (Science Technology Engineering and Mathematics) and projects based on open-ended problems are the right alternative choices. According to Sarlin's opinion (Sarlin et al., 2022), STEM is an approach that focuses on training students to engage in critical thinking, inquiry, problem-solving, collaboration, and engineering as design thinking. Instruments with the STEM approach are able to facilitate students in developing problem-solving skills, critical thinking skills, and creative thinking skills that support change in the 21st century (Sarican & Akgunduz, 2018; Saxton et al., 2014). Instruments with a STEM Approach can help students develop critical thinking skills, problem-solving abilities, and collaboration (Mufida et al., 2020). Question instruments with a STEM approach combined with project-style assignments can improve critical thinking skills. (Laboy-Rush, 2011; Fitriyah & Ramadani, 2021; Allanta & Puspita, 2021). An open-ended question instrument to measure critical thinking abilities has a positive impact on student's critical thinking abilities (Cahyono, 2021; Siswono, 2018).

The instrument is declared good if the instrument can function as a measuring tool for a variable that meets academic requirements (valid and reliable) and can measure what

you want to measure accurately. In accordance with the opinion of Djaali (2000) who defines an instrument as a tool that meets academic requirements and is used as a tool to measure an object or collect data from a variable. Nurkencana (1992: 141) states that a measuring device can be said to be a valid measuring device if the measuring device can measure what is intended to be measured precisely. Instruments that are suitable for use are instruments that meet the validity and reliability criteria so that they can carry out the measuring function correctly or provide measurement results that are in accordance with the measurement objectives (Sugiyono, 2014; Marfu'i et al., 2019). The instrument meets the criteria of validity and reliability if the research test instrument measures what it is supposed to measure (Wright, B. D., & Stone, 1979; Moleong, 2006). Analysis of the validity and reliability of research instruments can be carried out using the CTT (classical test theory) or the Rasch model (Rakkapao, 2022; Polat, 2022).

Research instrument analysis using the Rasch model has several advantages and can overcome various limitations of classical test theory or CTT (classical test theory). According to expert opinion, the limitations of the validity and reliability analysis of instruments with classical theory tests can be overcome with the Rasch model (Muntazhimah, 2020). The Rasch model is a solution to the validity problem where the Rasch model provides useful statistics and offers a tremendous opportunity to investigate validity (Bond & Fox, 2007). In addition, the application of the Rasch model in a study will be able to facilitate and produce more efficient, reliable, and valid measurements in addition to increasing user convenience. The Rasch model can produce reliable and valid instruments (Guat & Hamidon, 2022; Setyawati, 2018). A study to identify the validity and reliability of the instrument is very important to maintain the accuracy of the instrument (Ariffin et al., 2010). The use of the Rasch model is the development of a measuring tool in the social sciences in response to various paradigm weaknesses from classical test theory and the Rasch model can be used as a method for returning data according to its natural conditions (Sumintono, Bambang & Widhiarso, 2015).

Research in the last ten decades states that the development of critical thinking ability test instruments is still very much needed and analysis of the validity and reliability of the device using the Rasch model is a recommended alternative. In accordance with Faradillah's research (2021) states that instruments for measuring the critical thinking skills of prospective mathematics teachers which are examined through indicators of

open-mindedness, curiosity, systematics, truth-seeking, analytics, and self-confidence are declared valid and reliable after being analyzed with the help of the Rasch model using software Winstep and Confirmatory Factor Analysis (CFA) using JASP. The validity and reliability of test instruments need to be done to ensure that the instrument measures what should be measured and can draw good conclusions from trials on research samples (Bernardi & Pazinato, 2022; Wijayanti, 2019). Development of critical thinking skills at the secondary school level through indicators of (1) interpretation, (2) analysis, (3) evaluation, (4) inference, and (5) explanation and analysis through the Rasch model results in the conclusion that all questions developed are valid and reliable (Harjo, 2019). Rosnawati (2015) developed an instrument for critical thinking skills in learning mathematics in junior high schools which was analyzed through expert validity and classical test theory or CTT (classical test theory). Sumarni and Kadarwati (2018) states that 15 critical thinking skills questions developed through a contextual problem approach to chemistry and analyzed through expert validity and classical test theory or CTT (classical test theory) are declared valid and reliable. Facione (1990) developed the California Critical Thinking Skill Test (CCTST) which is an instrument to measure the critical thinking skills of nursing students. Based on the analysis of the research conducted, no research has been found regarding the development of instruments to measure critical thinking skills in the preparation of questions using the STEM approach and project assignments. So this study aims to develop test instruments that are used to measure critical thinking skills with the STEM approach and project assignments that are valid and reliable.

RESEARCH METHODS

The assessment instruments were developed based on the Research and Development using ADDIE (Branch, 2009) with analysis, design, development, implementation, evaluation. Critical thinking skills test tested on 33 students and has met the number of respondents who have met the requirements where based on. (Tan & Vicente, 2019) the number of respondents for the pilot study is between 25 to 100 people. (Johanson & Brooks, 2010) suggests the minimum number of respondents is 30 people. The analysis method of research data used is the Rasch model with the Winstep Software version 3.92.1 developed by (Linacre, 2006). The data was collected by using a test that was constructed by the researcher which contained 15 questions. The data analysis technique

used RASCH modeling analysis. This research identifies several things which include; 1) reliability, 2) validity, 3) analysis of the difficulty level of the questions, 4) the distribution of items, and 5) a description of the projection of respondents' answers in the scalogram. These five things are sufficient to describe the measuring power of the instrument on students' critical thinking skills. In addition, this analysis will show the reliability of respondents in taking tests which will later be used as data to measure students' critical thinking skills.

(Boone et al., 2014) suggest the following criteria to check whether an item is fit (item fit) or not fit (outlier or misfit). The Rasch model analyzes the suitability of items and the criteria used are: (1) Acceptable Outfit Mean Square (MNSQ) value = $0.5 < \text{MNSQ} < 1.5$; (2) Accepted Outfit Z-standard (ZSTD) value = $-2.0 < \text{ZSTD} < +2.0$; (3) Measure Correlation Value (Pt Mean Corr): $0.4 < \text{Pt Measure Corr} < 0.85$. Analyzing the data obtained using Winsteps software according to the Rasch model if the value of the sum of the middle quadrants (mean square) is 1.0 while the standardized value (Z-standardized value) is 0.0. This analysis can also identify questions that are too easy or too difficult and respondents who are outliers. The three criteria must be met by the item, so that the item is considered fit. In Perspective, Rasch sees that the item difficulty level and the testee's ability level are on the same scale, namely the logit scale (moving between -4 to 4). A rough idea of the logit number is the Testee (ability level). This taste will show low ability (-4.00 to -2.00), medium ability (-1.99.00 to 1.99), and high ability (2.00 to 4.00). The items indicating the level of difficulty (measures) are the easy items are in the MEASURE value of -4.00 to -2.00), the medium items are in the MEASURE value of 1.99 to 1.99, and the difficult items are in the MEASURE value of 2.00 to 4.00. Item Difference Power/PT-measure (Point Measure/Total Item Correlation) is shown if a negative value is obtained, meaning that item is problematic. If the item is below 0.20 it needs further inspection/not good.

The following are indicators of critical thinking that are used as the basis for making questions (B. Cahyono et al., 2019; Fisher, 2000; Fitriana et al., 2018; Norris, S. & Ennis, 1989), which can be seen in table 1:

Table 1. Description of Critical Thinking Indicator

Critical Thinking Indicator (FRISCO)	Indicators needed in competence
F (Focus)	Students understand the problem in a given problem. Added understanding of what information is known about the problem
R (Reason)	Students provide reasons based on relevant facts/evidence at each step in making decisions and conclusions. In this step, students should make steps to solve the problem
I (Inference)	(1) Students make appropriate completion plans; (2) Students choose the right reason (R) to support the completion plan that has been made.
S (Situation)	Students use all the information that is appropriate to the problem.
C (Clarity)	(1) Students use further explanations about what is in the conclusions made; (2) If there is an unusual (new) term in the question, students can explain it; (3) Students give examples of similar cases.
O (Overview)	Students research or re-examine thoroughly from start to finish (produced by FRISCO)

In this study the data obtained by processing and analyzing using descriptive statistical techniques. This analysis includes the average, standard deviation, maximum value and minimum value. The difference in the average value of critical thinking skills is used independent t-test.

RESULTS AND DISCUSSION

The results of product validation carried out by two experts obtained an average of 92.4% with a very good category so that the category can be used with slight revision. The instrument used is expected to be able to measure what it should measure (critical thinking ability) after carrying out minor revisions according to suggestions given by experts and being declared valid. In accordance with the opinion of Suyatman et al. (2021) which states that valid instruments can be used to measure what should be measured.

Table 2. Example questions to measure critical thinking skills using STEM and project approaches

<p>1. Julia membeli sebidang tanah, sebuah rumah dan sebuah toko dengan harga Rp. 500.000.000,-. Tiga tahun kemudian pemuda tersebut menjual tanahnya dengan harga Rp. 300.000.000,- dan toko dengan harga Rp. 250.000.000,-. Jika Julia ingin memperoleh keuntungan sebesar 25%, apakah kalian setuju jika harga beli tanahnya adalah Rp. 240.000.000,- ? berapakah harga rumah Julia?</p>	<p>1. Julia bought a plot of land, a house and a shop for Rp. 500,000,000,-. Three years later the young man sold his land for Rp. 300,000,000, - and a shop with a price of Rp. 250,000,000,-. If Julia wants to get a profit of 25%, do you agree that the purchase price of the land is Rp. 240,000,000,- ? How much is Julia's house worth?</p>
---	---

2. Susi akan berbelanja ke toko “Berkah Rejeki” untuk membeli tunik dan celana. Susi menemukan tunik yang ia sukai dengan harga Rp. 350.000,-, sebelumnya ia telah memiliki voucher senilai Rp. 85.0000,-. Voucher tersebut dapat digunakan dengan minimal pembelian Rp. 300.000,-. Lalu ia beranjak ke bagian lain untuk mencari celana. Susi tertarik untuk membeli celana berwarna hitam seharga Rp. 200.00,- yang bertuliskan diskon 20%. Sesuai dengan aturan toko, Susi hanya bisa menggunakan salah satu jenis potongan, tidak bisa keduanya. Susi memilih voucher untuk mendapatkan harga termurah. Apakah kalian setuju dengan keputusan susi? Berikanlah alasannya!

2. Susi is going shopping at the "Berkah Fortune" store to buy a tunic and pants. Susi found a tunic she liked for Rp. 350,000, -, before he had a voucher worth Rp. 85.0000,-. The voucher can be used with a minimum purchase of Rp. 300.000,-. Then he moved to another section to look for pants. Susi is interested in buying black pants for Rp. 200.00, - which says 20% discount. According to the rules of the shop, Susi can only use one type of cut, not both. Susi chose a voucher to get the cheapest price. Do you agree with Susi's decision? Give reasons!

3. Perhatikan tabel di bawah ini!



Batik Semarangan (Credit by Batik Bedjo)

Toko A	Toko B	Toko C	Toko D
Harga Rp200.000	Harga Rp120.000	Harga Rp130.000	Harga Rp102.500
Diskon 25%	Diskon 15%	Diskon 28%	Cashback Rp20.000

Suatmoko ingin membeli Batik Semarang. Ia dapat membeli di toko A, toko B, toko C, maupun toko D. Harga dan diskon yang diberikan keempat toko tersebut berbeda-beda. Perhatikan skema diskon tabel diatas. Suatmoko ingin mengeluarkan uang seminimal mungkin.

- Apakah kalian bisa menemukan toko mana yang akan dipilih Suatmoko?
- Apakah pilihanmu untuk suatmoko sudah menjawab permasalahan di soal tersebut? Jelaskan!

3. Pay attention to the table below!



Batik Semarangan (Credit by Batik Bedjo)

Store A	Store B	Store C	Store D
200 IDR	120 IDR	130 IDR	102,5 IDR
Discount 25%	Discount 15%	Discount 28%	Cashback Rp20.000

Suatmoko wants to buy Batik Semarang. He can buy at shop A, shop B, shop C, or shop D. Prices and discounts offered vary widely. Look at the discount scheme table above. Suatmoko wants to spend as little money as possible.

- Can you tell Suatmoko which store to choose?
- Does your chosen moko answer the questions in the question? Explain!

Table 2 is an example of a valid instrument in the form of questions with a STEM approach and open-ended type projects which are able to stimulate the thinking process to get a variety of different but correct forms of answers (creativity) in solving problems so as to improve critical thinking skills. In accordance with research by Parno (2021) which states that the STEM approach combined with the 7E LC learning model is effective in improving students' critical thinking skills in completing fluid statistics projects at the high school level. Monrat (2022) states that open-ended questions are effective for improving students' performance of critical thinking skills in learning mathematics. Sapta (2019) Penerapan model problem posing dengan soal open end mampu meningkatkan kemampuan berpikir kritis siswa. Novitasari (2022) dalam penelitiannya mengungkapkan bahwa LKS berbasis etnomatematika dan pendekatan STEM efektif dalam meningkatkan kemampuan berpikir kritis siswa.

The Measure shows that the item difficulty level and the testee's ability level are on the same scale, which is meeting the logit scale, the value is between -4 to 4. From the test questions, it shows a moderate level of difficulty. No one goes on a high or low

level of difficulty. The usefulness of the item difficulty level has two things, namely usefulness for teachers, testing, and teaching (Hadzhikoleva et al., 2019). Its uses for teachers are (1) as an introduction to the concept of relearning and providing input to students about learning outcomes, and (2) obtaining information about curriculum emphasis or reviewing the usual items (Rangkuti, 2011). Figure 1 shows that the level of difficulty of the test questions is in the medium category seen in “Measure”.

TABLE 3.1 DATA.XLSX Z00M34M5.TXT Aug 25 23:16 2022
 INPUT: 32 PERSON 15 ITEM REPORTED: 32 PERSON 15 ITEM 11 CATS WINSTEPS 3.73
 PERSON: REAL SEP: 2.07 REL: .91 --- ITEM: REAL SEP: 1.81 REL: .94

ITEM STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBSV	EXPS	ITEM
15	5	32	-.94	.24	.85	-.12	-.6	-.45	-.29	93.8	91.8	4d	
13	10	32	-.74	.18	1.94	1.4	-.34	-.2	-.45	38	87.5	4b	
14	10	32	-.74	.18	1.94	1.4	-.34	-.2	-.45	38	87.5	4c	
12	23	32	.47	.12	1.28	0	.42	-.8	.65	49	89.4	4a	
11	68	32	.27	.10	.86	-.3	.26	-1.0	.78	54	37.6	3d	
9	54	32	-.15	.09	.73	-1.0	-.63	-.8	.64	58	25.0	3b	
10	64	32	-.08	.08	1.24	1.0	1.38	1.0	-.57	59	21.9	3c	
8	80	32	-.02	.08	1.57	2.1	1.84	2.2	.53	61	25.0	3a	
4	125	32	-.26	.07	.97	-2.1	-.75	-1.1	.91	64	21.9	1d	
2	128	32	-.27	.07	.91	-.3	.98	-.1	.51	64	12.9	1b	
3	159	32	-.42	.07	1.19	-.9	1.35	1.3	.66	63	6.3	1c	
6	166	32	-.46	.07	1.02	2.1	1.26	1.0	.63	63	15.6	2b	
3	175	32	-.50	.07	.90	-.4	1.25	-.9	.60	62	11.9	1a	
5	206	32	-.65	.07	1.00	1.1	1.03	-.2	.58	60	12.5	2a	
1	234	32	-.79	.07	1.27	1.1	1.16	-.8	.38	57	9.4	1a	
MEAN	98.6	32.0	-.00	.00	1.13	-.3	.88	-.2		35.8	34.0		
S.D.	74.2	.0	-.52	.05	.40	1.1	.48	-.9		29.6	28.7		

Figure 1. Rasch's perspective for the Differentiating Power of Items in the test questions

Figure 1 also shows Rasch's perspective for the Differentiating Power of Items in the test questions, all items have high discriminating power at PT. Measure Corr. get more than 0.2. The discriminatory power of a question is the ability of the question to distinguish between students who have a high level of ability and students who have low abilities. Hambleton et.al (1978) stated that the main purpose of Item Response Theory is to provide similarities between the statistical questions and ability estimates. Subekti and Khoirunnisa (2018) said that the use of the Rasch model is highly recommended to develop instruments to measure critical thinking skills in order to get valid and reliable questions

SUMMARY OF 32 MEASURED PERSON													SUMMARY OF 15 MEASURED ITEM				
TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	Z	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	Z		
MEAN	46.3	15.0	-.42	.12	1.00	.0	.88	MEAN	98.6	32.0	-.00	.00	1.13	-.3	.88		
S.D.	22.3	.0	-.19	.02	4.98	4.6	5.09	S.D.	74.2	.0	-.52	.05	.40	1.1	.48		
MAX.	305.0	15.0	-.19	.18	4.98	4.6	5.09	MAX.	234.0	32.0	-.94	.24	1.94	2.1	1.04		
MIN.	38.0	15.0	-1.87	.11	-.38	-2.1	-.1	MIN.	5.0	32.0	-.79	.07	.97	-2.1	-.12		
REAL RMSE	.13	TRUE SD	.38	SEPARATION	2.40	PERSON RELIABILITY		REAL RMSE	.13	TRUE SD	.52	SEPARATION	4.46	ITEM RELIABILITY			
S.E. OF PERSON MEAN	.088							S.E. OF ITEM MEAN	.14								
PERSON RAW SCORE-TO-MEASURE CORRELATION	-.89							ITEM RAW SCORE-TO-MEASURE CORRELATION	-.97								
CROHBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY	.84							480 DATA POINTS, LOG-LIKELIHOOD CHI-SQUARE: 790.31 with 425 d.f., p=.0000									
								Global Root-Mean-Square Residual (excluding extreme scores): 2.1375									

Figure 2. Summary Statistic Reliability Coefficient

From Figure 2 show that the value of Cronbach's Alpha (KR-20) is the reliability coefficient obtained based on the classical test theory approach. This value is the

interaction between people and goods as a whole. The alpha value is 0.84. This shows that the reliability of the test in general is very satisfactory (Reliable). The person reliability value is 0.81 and the item reliability value is 0.94. This shows that the consistency of the answers from the subjects is good and the quality of the items on the reliability aspect of the instrument is good. Reliability is the extent to which a research test instrument can be expected to obtain consistent results when the test is repeated. Reliability can provide consistency of validity (Yasin et al., 2015).

One of the features of Rasch's analysis with Winstep is that there is a map that depicts the distribution of the subject's ability and the distribution of item difficulty with the same scale. This map is called the Wright Map which is nothing but a person-item map. This shows that the average individual ability is slightly higher than the item difficulty level. Therefore, this study will look further at one or two items that have a low index of difficulty to be changed and adjusted so that the instrument items will be a little more difficult for research respondents to answer.

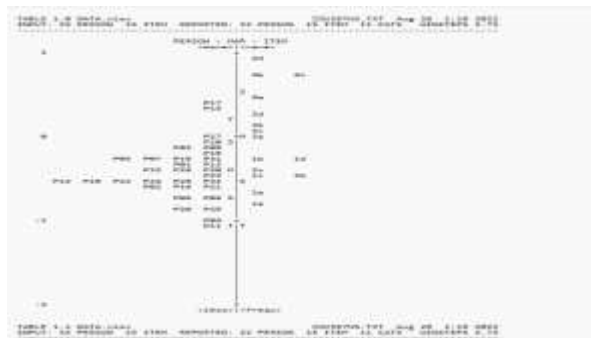


Figure 3. Wright Map

Figure 3 depicts the distribution of respondents' abilities and the distribution of difficulty levels of questions on the same scale which can be seen from the Wright Map. The Wright map on the left describes the students' abilities, and it can be seen that students with code P17 (have the highest ability compared to other students. Even though they have the highest ability, this student's logit score is less of +1 logit Student P17 is outside the limit of two standard deviations (T) indicating a different high intelligence (outlier). The student with the lowest ability is P11, with a logit value of more than -1, which also indicates a very low ability (outlier), because it is outside the boundary of T. Wright's map on the right explains the distribution of the logit value of the item, number 4 (4d) is the problem with the highest level of difficulty (+1 logit), which means the probability that all students do this problem correctly. very small. As for question 1a is the question

with the lowest logit value (close to -1 logit). In this case, the more students are able to work on this question I correctly. Problems 1b and 1d have the same level of difficulty because the logit value is the same, as in questions 1c and 2b. The average logit value of the item is always set in logit 0.0 which indicates the initial reference point of the scale. From the Wright Map, the average logit person was found to be -1 logit (below 0.0 logit). This shows that the average ability of students is below the average level of difficulty of standard questions. Wright's map also shows that 2 students had logit scores below -1 logit, reflecting that more than half of the sample had low critical thinking skills.

The ability of the test instrument is designed to differentiate critical thinking skills between students who practice critical thinking skills who receive the STEM-PjBL learning process and those who do not receive treatment. Test data on students who are normally distributed and homogeneous, then on the different hypothesis test the average post-test results are obtained as follows:

Table 3. Z-Test Results for Critical Thinking Ability

Group	Average	Deviasi Standart	Z _{hitung}	Z _{0,05}	Interpretation
The Critical Thinking ability	49,79	28,84	12,03	1,69	hypothesis accepted
Critical Thinking Ability with a project, STEM and open-ended approach	61,46	25,79			

From Table 3, information is obtained that Z_{count} is greater than Z_{table} , so it can be stated that the average critical thinking ability of the experimental class which uses a project, STEM, and an open-ended approach is better than the control class. The experimental average was 61.46 which showed that students' critical thinking skills were still low. In accordance with research by Rahmawati (2019) revealed that critical thinking and creative thinking skills can be improved through project-based learning with a STEAM approach integrated with chemistry concepts. Novitasari (2022) in his research revealed that worksheets based on ethnomathematics and the STEM approach are effective in improving students' critical thinking skills. Retnowati (2020) revealed that the development of rectangular modules with a STEM approach was effective in improving the critical thinking skills of students in the medium category. Yulianti et al., (2021) and Awad (2023) The results of his research stated kemampuan

berpikir siswa dapat dikembangkan melalui penerapan model pembelajaran dengan pendekatan STEM.

CONCLUSION

Rasch model analysis on the measure that shows the level of difficulty of the questions has a moderate level of difficulty. At PT-Measure Corr. indicates that the discriminating power of the grains has a high discrepancy. Furthermore, checking whether an item is appropriate (item fit) or not fit (outlier or misfit) is shown in the MNSQ OUTFIT value, ZSTD OUTFIT value, PT MEASURE CORR value. From the questions tested, items 1b, 1c, 1d, 2a, 2b, 2c, 2d, 3b, 3c, 3d met the fit items. Cronbach's Alpha value (KR-20) is the reliability coefficient obtained based on the classical test theory approach. This value is the interaction between the person and the item as a whole. Alpha value is 0.72. This shows that the reliability of the test in general is very satisfactory (Reliable). The value of person reliability is 0.76 and the value of item reliability is 0.84. This shows that the consistency of the answers from our subjects is good and the quality of the items in the instrument's reliability aspect is good. The result of its application is that there is an average difference in pretest and posttest scores using instruments based on open ended and ethnomathematics. The average pretest score was 49.79 and the posttest score was 61.46.

REFERENCE

- Allanta, T. R., & Puspita, L. (2021). Analisis keterampilan berpikir kritis dan self efficacy peserta didik: Dampak PjBL-STEM pada materi ekosistem. *Jurnal Inovasi Pendidikan IPA*, 7(2). <https://doi.org/10.21831/jipi.v7i2.42441>
- Ariffin, S., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability Multiple Intelligent item using Rasch measurement model. *Procedia - Social and Behavioral Sciences*, 9, 729–733. <https://doi.org/10.1016/j.sbspro.2010.12.225>
- Awad, N. (2023). Exploring STEM integration: assessing the effectiveness of an interdisciplinary informal program in fostering students' performance and inspiration. *Research in Science and Technological Education*, 41(2), 675–699. <https://doi.org/10.1080/02635143.2021.1931832>
- Basri, H., Purwanto, As'ari, A. R., & Sisworo. (2019). Investigating Critical Thinking Skill of Junior High School in Solving Mathematical Problem. *International Journal of Instruction*, 12(3), 745-758. <https://doi.org/10.29333/iji.2019.12345a>
- Bernardi, F. M., & Pazinato, M. S. (2022). The Case Study Method in Chemistry

- Teaching: A Systematic Review. *Journal of Chemical Education*, 99(3), 1211–1219. <https://doi.org/10.1021/acs.jchemed.1c00733>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences Second Edition University of Toledo.an Sciences Second Edition University of Toledo.*
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. *Rasch Analysis in the Human Sciences*, 1–482. <https://doi.org/10.1007/978-94-007-6857-4>
- Changwong, K. (2018). Critical thinking skill development: Analysis of a new learning management model for Thai high schools. *Journal of International Studies* 11(2), 37–48. doi: 10.14254/2071-8330.2018/11-2/3.
- Cahyono, B., Kartono, Waluyo, B., & Mulyono. (2019). Analysis critical thinking skills in solving problems algebra in terms of cognitive style and gender. 5th International Conference on Mathematics, Science and Education 2018, ICMSE 2018, 1321(2). <https://doi.org/10.1088/1742-6596/1321/2/022115>
- Cahyono, B. (2017). Analisis Ketrampilan Berfikir Kritis Dalam Memecahkan Masalah Ditinjau Perbedaan Gender. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika* 8 (1), 50-64. <https://doi.org/10.26877/aks.v8i1.1510>
- Cahyono, B., Kartono, K., Waluya, B., Mulyono, M., & Setyawati, R. D. (2021). Problem-based learning supported by arguments scaffolding that affect critical thinking teacher candidates. *Cypriot Journal of Educational Sciences*, 16(6). <https://doi.org/10.18844/cjes.v16i6.6480>
- Branch Djaali (2000). *Metode Penelitian Untuk Ilmu-ilmu Perilaku*. Jakarta : Bumi Akasara.
- Ennis, R. (1996). *Critical Thinking*. New Jersey: Prentice-Hall.
- Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions*. University of Illinois.
- Faradillah, A., & Adlina, S. (2021). Validity of critical thinking skills instrument on prospective Mathematics teachers. *Jurnal Penelitian dan Evaluasi Pendidikan*, 25(2), 126-137. doi:<https://doi.org/10.21831/pep.v25i2.40662>
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction—The Delphi report*. Millbrae, CA: California Academic Press.
- Fitriyah, A., & Ramadani, S. D. (2021). Pengaruh Pembelajaran Steam Berbasis Pjbl (Project-Based Learning) Terhadap Keterampilan. *Journal Of Chemistry And Education (JCAE)*, X(1), 209–226.
- Fisher, A. (2000). Book Reviews: *Critical Thinking* Prentice Hall). Robert H.Ennis, 1996. *Argumentation*, 40(1), 51–56. <https://doi.org/https://doi.org/10.1023/A:1007850227823>
- Guat, C. L., & Hamidon, Z. (2022). *A case study of problem-based learning approach in a mathematics online learning course* (I. H., Z. J., A. N., D. M.C.B., B. Y. A.M., K. S., & B. J.B.M. (eds.); Vol. 2472). American Institute of Physics Inc. <https://doi.org/10.1063/5.0092735>

- Hadzhikoleva, S., Hadzhikolev, E., & Kasakliev, N. (2019). Using peer assessment to enhance Higher Order thinking skills. *TEM Journal*, 8(1), 242–247. <https://doi.org/10.18421/TEM81-34>
- Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. (1978). Criterion-Referenced Testing and Measurement: A Review of Technical Issues and Developments. *Review of Educational Research*, 48(1), 1–47. <https://doi.org/10.3102/00346543048001001>
- Harjo, B., Kartowagiran, B., & Mahmudi, A. (2019). Development of Critical Thinking Skill Instruments on Mathematical Learning High School. *International Journal of Instruction*, 12(4), 149166. <https://doi.org/10.29333/iji.2019.12410a>
- Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394–400. <https://doi.org/10.1177/0013164409355692>
- Laboy-Rush, D. (2011). *Integrated STEM Education through Project-Based Learning*. [https://doi.org/http://rondoutmar.sharpschool.com/UserFiles/Servers/Server_719363/File/12-13/STEM/STEM-White-Paper-101207-final\[1\].pdf](https://doi.org/http://rondoutmar.sharpschool.com/UserFiles/Servers/Server_719363/File/12-13/STEM/STEM-White-Paper-101207-final[1].pdf)
- Lamb, S. (2017). Key skills for the 21st century: An evidence-based review. Melbourne Australia: State of New South Wales.
- Linacre, J. M. (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, 7(1), 129–139.
- Marfu'i, L. N. R., Ilfiandra, & Nurhudaya. (2019). The analysis of critical thinking skills test in social-problems for physics education students with Rasch Model. *Journal of Physics: Conference Series* 1280 (2019) 052012 IOP Publishing doi:10.1088/1742-6596/1280/5/052012
- Mason, M. (2007). Critical thinking and learning. *Educational Philosophy and Theory*, 39(4), 339–349. <https://doi.org/10.1111/j.1469-5812.2007.00343.x>
- Mufida, S. N., Sigit, D. V., & Ristanto, R. H. (2020). Integrated project-based e-learning with science, technology, engineering, arts, and mathematics (PjBeL-STEAM): its effect on science process skills. *Biosfer*, 13(2), 183–200. <https://doi.org/10.21009/biosferjpb.v13n2.183-200>
- Muntazhimah, M., Putri, S., & Khusna, H. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, 6(1): 65-74. <https://journal.lppmunindra.ac.id/index.php/jkpm/article/view/8144>
- Moleong, j, Lexy. 2006. Metodologi Penelitian Kualitatif. Bandung: PT. Remaja Rosdakarya.
- Monrat, N., Phaksunchai, M., & Chonchaiya, R. (2022). Developing students' mathematical critical thinking skills using open-ended questions and activities based on student learning preferences. *Education Research International*, 2022. <https://www.hindawi.com/journals/edri/2022/3300363/>
- Norris, S., & Ennis, R. (1989). *Evaluating critical thinking*. Critical Thinking Press and Software.

- Novitasari, N., Febriyanti, R., & Wulandari, I. A. (2022). Efektivitas LKS berbasis etnomatematika dengan pendekatan STEM terhadap kemampuan berpikir kritis. *Vygotsky: Jurnal Pendidikan Matematika Dan Matematika*, 4(1), 57-66. <https://jurnalpendidikan.unisla.ac.id/index.php/VoJ/article/view/521/385>
- Nurkencana, Wayan., PPN. Sunartana. Evaluasi Hasil Belajar, Surabaya: Usaha Nasional, 1992.
- Parno., Supriana, E., Widarti, A. N., & Ali, M. (2021). The effectiveness of STEM approach on students' critical thinking ability in the topic of fluid statics. The effectiveness of STEM approach on students' critical thinking ability in the topic of fluid statics. *Journal of Physics: Conference Series IOP Publishing 1882 (2021) 012150* doi:10.1088/1742-6596/1882/1/012150
- Polat M, Turhan NS, Toraman Ç (2022). Comparison of Classical Test Theory vs. Multi-Facet Rasch Theory. *Pegem Journal of Education and Instruction*, Vol. 12, No. 2, 2022, 213-225. DOI: 10.47750/pegegog.12.02.21
- Rahmawati, Y., Ridwan, A., & Hadinugrahaningsih, T. (2019). Developing critical and creative thinking skills through STEAM integration in chemistry learning. *IOP Conf. Series: Journal of Physics: Conf. Series 1156 (2019) 012033 IOP Publishing* doi:10.1088/1742-6596/1156/1/012033
- Rakkapao, S., Nualtong, K., & Sumathakulawat, W (2022). Classical test theory and Rasch analysis of test of understanding of vectors (TUV). *Siam Physics Congress (SPC) 2020 Journal of Physics: Conference Series 1719 (2021) 012089 IOP Publishing* doi:10.1088/1742-6596/1719/1/012089
- Rangkuti, Freddy. 2011. SWOT Balanced Scorecard Teknik Menyusun Strategi Korporat yang Efektif plus Cara Mengelola Kinerja dan Risiko. Jakarta: Gramedia.
- Retnowati, S., Riyadi, & Subanti, S. (2020). The STEM approach: The development of rectangular module to improve critical thinking skill. *International Online Journal of Education and Teaching (IOJET)*, 7(1). 2-15. <http://iojet.org/index.php/IOJET/article/view/704>
- Rosnawati, R., Kartowagiran, B., & Jailani (2015). A Formative Assessment Model Of Critical Thinking In Mathematics Learning In Junior High School. *Research and Evaluation in Education Journal*. Volume 1, Number 2, (pages 186-198). <https://doi.org/10.21831/reid.v1i2.6472>
- Sarican, G., & Akgunduz, D. (2018). The impact of integrated STEM education on academic achievement, reflective thinking skills towards problem solving and permanence in learning in science education. *Cypriot Journal of Educational Sciences*, 13(1), 94–113. <https://doi.org/10.18844/cjes.v13i1.3372>
- Sarlin, M., Arsyad, A., & Haris, I. (2022). Identifying Key Component of Collaborative Problem Solving in Teaching and Learning Process: The Challenges Ahead in Preparing for 21st Century Skills. *Journal of Higher Education Theory and Practice*, 22(5), 207–214. <https://doi.org/10.33423/jhetp.v22i5.5215>
- Sapta, A., Pakpahan, S. P., & Sirait, S. (2019). Using the problem posing learning model based on open ended to improve mathematical critical thinking ability. *Journal of Research in Mathematics Trends and Technology*, 1(1), 12-15. <https://doi.org/10.32734/jormtt.v1i1.752>

- Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, E. A. (2014). A Common Measurement System for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation*, 40, 18–35. <https://doi.org/10.1016/j.stueduc.2013.11.005>
- Setyawati, R. D. (2018). Instrumen Angket Self-Esteem Mahasiswa Ditinjau Dari Validitas Dan Reliabilitas. *Phenomenon : Jurnal Pendidikan MIPA*, 7(2), 174–186. <https://doi.org/10.21580/phen.2017.7.2.1932>
- Siswono, Tatag Yuli Eko. 2018. Pembelajaran Matematika Berbasis Pengajaran dan Pemecahan Masalah. Bandung: PT Remaja Rosdakarya.
- Subekti, A. W., and Khoirunnisa, F. (2018). Penggunaan Rasch Model Untuk Mengembangkan Instrumen Pengukuran Kemampuan Berpikir Kritis Siswa Pada Topik Ikatan Kimia. *Jurnal Zarah*, VI (2), 74.
- Sugiyono. (2014). Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Sumarni, W. & Kadarwati, S. (2020). Ethno-Stem Project-Based Learning: ITS Impact to Critical And Creative Thinking Skills. *Jurnal Pendidikan IPA Indonesia*. (JPII) 9 (1) (2020) 11-21. DOI: 10.15294/jpii.v9i1.21754
- Sumintono, Bambang & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan, 1–127
- Suyatman, Saputro, S., Sunarno, W., & Sukarmin. (2021). Of student analytical thinking skills in the natural sciences by implementing problem-based learning model . *Journal of Educational, Cultural and Psychological Studies*, 2021(23), 89–111. <https://doi.org/10.7358/ecps-2021-023-suya>
- Tan, T. A. G., & Vicente, A. J. (2019). An innovative experiential and collaborative learning approach to an *undergraduate* marketing management course: A case of the Philippines. *International Journal of Management Education*, 17(3). <https://doi.org/10.1016/j.ijme.2019.100309>
- Wijayanti, M.D., Rahardjo, S.B., Saputro, S., & Mulyani, S. (2019). Item Analysis Of Critical Thinking Skills Instrument To Measure Effectiveness Of Scientific Group Inquiry Learning (SGIL) Model. *Jurnal Pendidikan IPA Indonesia* (JPII) 8 (4) (2019) 538-546. <https://doi.org/10.15294/jpii.v8i4.20794>
- Wright, B. D., & Stone, M. H. (1979). *Test Design*. <https://doi.org/Chicago>, IL: Mesa Press.
- Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences*, 204(November 2014), 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>
- Yulianti, E., Hamimi, E., & Rahmagusviana, R. (2021). Pengembangan Buku Ajar IPA Berpendekatan STEM untuk Mengembangkan Kreativitas Siswa pada Tema Tekanan. *Phenomenon : Jurnal Pendidikan MIPA*, 11(2), 169–188. <https://doi.org/10.21580/phen.2021.11.2.7775>