e-ISSN: 2355-620X
DOI: 10.12928/jrkpf.v12i1.1161

Aiken's V analysis and Rasch modeling to determine the quality of physics creative thinking test instruments

Kurniawan, Parlindungan Sinaga, Duden Saepuzaman

Physics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Email: kurniawan@upi.edu

Abstract

Creative thinking ability is one of the important high-level thinking skills. To assess this ability, a measurement instrument in the form of a test is needed. Therefore, the study aims to evaluate the validity of the content and constructs that measure creative thinking ability using Aiken's V index analysis and Rasch modeling. The study used a quantitative descriptive method by assessing 12 test items on the measurement material, climate change, and renewable energy. Five experts assessed material, construction, and language, and a trial was conducted on 134 student samples. The results showed that all questions were valid. Undimensionality showed results of less than 15%, the level of item suitability was stated as appropriate, and the difficulty level varied in 4 categories. The Cronbach Alpha value was 0.77, and the person and item reliability was 0.76 and 0.92, respectively. These results indicate that the test is a reliable evaluation tool and a basis for developing similar instruments. It also has practical implications in designing learning to develop students' creative thinking.

Keywords: Aiken's V, creative thinking, Rasch model, test instrument,

Article submitted 2024-12-20. Revision uploaded 2025-01-05.
Accepted for publication 2025-01-10.
Published on 2025-04-01.
https://doi.org/10.12928/jrkpf.v12i1.1161

I. Introduction

Education is a fundamental pillar in the development of individuals and societies. According to the OECD [1], the global education system continues to transform with the integration of technology and dynamic education policies to meet the needs of students in the 21st century. In the context of high school education, there are various subjects studied by students that are considered difficult and uninteresting, one of which is physics [2]. Physics is considered difficult because it often stems from the abstract nature and complexity of the mathematical concepts involved [3].

The difficulties experienced by students in solving physics problems indicate that students are still less able to solve more complex problems [4]. If students can develop their creative thinking skills well, they will be able to solve physics problems and identify problems correctly from the start [5]. By thinking creatively, students can develop the ability to see problems from various perspectives, produce innovative solutions, and apply physics concepts in different contexts [6]. In addition, students with high creativity play a significant role in solving physics problems by tending to be more able to find various alternative solutions and apply physics concepts effectively in different situations [7].

Creative thinking is generating new and useful ideas to solve problems or face challenges. According to Runco and Acar [8] in their book The Creative Self, creative thinking involves two main processes, namely divergent thinking, which produces various different ideas, and convergent thinking, which chooses the best idea. Creative thinking is an important ability that must be developed in order to be able to compete amidst global challenges [9]. According to the identification results of the Partnership for 21st Century Skills (P21) from the United States based in the US, there are four skills known as 4C, namely communication, collaboration, critical thinking, and creativity, which are considered important to have in the 21st century [10].

Elis Paul Torrance is one of the researchers who has conducted research on creative thinking skills. According to Torrance in Munandar [11], creative thinking indicators include originality, fluency, flexibility, and elaboration. This category is also in line with the views of Hue and Adey [12], who use the characteristics of scientific creativity, namely fluency, flexibility, and originality, to evaluate creative thinking skills.

Teachers are one of the factors that play an important role in developing student's creative thinking skills at school. However, not all teachers know the extent of students' creative thinking skills. Research by Tumangger, Kartono, & Ridlo [13] shows that teachers do not yet know students' creative thinking skills because they do not yet have an instrument to assess creative thinking skills. Likewise, Widyatiningtyas [14], in their research, highlighted that some teachers have not developed tests that specifically measure students' creative thinking skills. As conveyed by Solihat [15], it was revealed that the limited instruments that can specifically measure students' creative thinking skills make it difficult for teachers to make objective assessments; the development of valid and reliable instruments is still a challenge to support the assessment of student's creative thinking skills. The researcher also conducted a questionnaire to a group of physics tutoring students who stated that the physics teacher at their school gave a test with a single answer that did not require them to think creatively freely.

Based on the problems above, it is necessary to create a test instrument that can reveal students' creative thinking abilities. The test is designed with open-ended questions that can challenge students to think more deeply and creatively, so it is expected to encourage students to develop various alternative solutions, explore new ideas, and combine their knowledge and experience in solving problems. By using this test, researchers can analyze how students apply creative thinking skills in completing the tasks given.

The development of test instruments must go through several stages to ensure that the test creates an accurate measure of what should be measured. Among the forms of validation that can be done are expert validation and construct validation. Several previous studies have done this with expert validation, such as the validation of test instruments designed to measure high school students' problem-solving skills in physics [16] and validation of questions designed to measure problem-solving abilities with validation by experts covering aspects of content, construct, and appearance, as well as the practicality of using the instrument [17]. Several studies have also conducted construct validation using Rasch modeling with Winstep. Such as the analysis of the mental health screening tools instrument using the Rasch model [18], analysis of the reliability and validity of the self-determination questionnaire [19], Construct validity of the ICS (Indonesian Creativity Scale) [20], and analysis of the distribution map of physical education learning motivation through rasch modeling in elementary schools [21]. Unfortunately, no research has been found that investigates or reveals the analysis of special creative thinking test instruments in physics learning in terms of content with Aiken's V and in terms of construct using Rasch modeling.

This research is important to be conducted with the aim of knowing the validity of the content and construct of the test instrument because before being distributed as a tool for evaluating students' creative thinking, a validation process is needed to ensure the quality of the test instrument created. The validation process consists of content validation carried out by several experts using Aiken's V. After being declared valid; a limited field trial is conducted to determine the quality of the test instrument in terms of construct using Rasch modeling. This research contributes to providing a valid and reliable test instrument to measure students' creative thinking abilities, which can be used as a reference in educational research and practice.

II. Methods

This study uses a descriptive method with a quantitative approach. Yakin [22] explains that the purpose of descriptive research is to identify the value of one or more independent variables without comparing or relating them to other variables. Meanwhile, Agustini [23] emphasizes that the quantitative approach involves the use of numbers in the process of collecting, analyzing, and presenting data.

From the description above, it can be concluded that quantitative descriptive research aims to collect information about existing phenomena, clearly explain the objectives to be achieved, plan the approach to be applied, and collect various types of data as a basis for compiling reports. In this study, the author aims to evaluate the quality of creative thinking test instruments in physics learning, especially on descriptive questions about measurement materials, climate change, and renewable energy.

The creative thinking ability targeted in this instrument is in the form of descriptive questions that focus on measuring four aspects of creative thinking: fluency, flexibility, originality, and elaboration with assessment techniques ranging from 0-4 points based on assessment techniques according to creative thinking indicators. Researchers use four aspects along with indicators and assessment techniques as described by Torrance [24] in Table 1.

Table 1. Aspects and indicators of creative thinking ability

No.	Creative Thinking Aspects	Indicator	Assessment Techniques
1	Fluency	Providing many relevant ideas or thoughts in the form of words	Based on the total number of relevant responses
2	Flexibility	Produces varied answers with different points of view	Based on the number of different categories of relevant responses
3	Originality	Generate ideas according to one's own thoughts	Based on the original response answer
4	Elaboration	Develop, embellish, or elaborate on ideas	Based on the amount of detail in the response

Participants in this study were students from two Senior High Schools (SMA) who had received learning about measurement, climate change, and renewable energy, namely SMA Negeri 1 Rongga and SMA Pasundan 8 Bandung. Sampling was carried out using the purposive sampling method with a total of 134 11th-grade students consisting of 77 males and 57 females.

The data obtained from the assessment of 5 experts consisting of 3 lecturers of UPI Physics Education and two physics teachers at the school will be analyzed using Aiken's V to assess the validity of the test instrument content from the aspects of material, test construction, and language use. This V value will be interpreted in the range of 0.00 to 1.00, which indicates whether or not the content validity is good and provides an indication of support for the overall content validity. Aiken's V validity threshold value with five assessors and four assessment scales between 0-3 is 0.87 [25].

After that, a trial of the instrument was conducted to ensure that the sample had studied the physics material tested in the test consisting of 5 physics classes from two different schools. The trial was conducted with the same stages on November 26, 2024, at SMA Negeri 1 Rongga and November 28, 2024, at SMA Pasundan 8 Bandung, with each test duration in each class 90 minutes. The trial results will be analyzed using the Rasch model with the help of Winstep 3.73 software to assess Undimensionality, Item Fit levels, Item Difficulty levels (Item Measure), Cronbach Alfa values, and Person & Item Reliability. The Rasch model was chosen because this approach produces a consistent measurement scale, so it can provide more precise information about test participants and the quality of the questions they answer [26].

III. Results and discussion

Content validity using Aiken's V

Determination of the validity of the instrument content in this study was carried out by involving expert agreement. The use of expert agreement in assessing content validity is important because a test instrument is considered valid if experts believe that the instrument can measure the intended ability. Aiken's validity index can be used to assess expert agreement [27]. The results of Aiken's validation are shown in Table 2.

According to Aiken, the validity of a question item is said to be good if it is assessed by five assessors and obtains an Aiken index equal to or greater than 0.87 [25]. Aiken's V index is a measure of agreement to measure the level of agreement between assessors regarding the suitability of the question item with the indicator to be measured [28]. The calculation results show that the twelve questions in the creative thinking test instrument are declared valid. Furthermore, the instrument was tested on high school students to assess the construct validity of the creative thinking test.

Table 2. Aiken's V Results

Question Items	Average V Value	Conclusion
1b	0.93	Valid
1c	0.92	Valid
1d	0.92	Valid
2a	0.93	Valid
2b	0.95	Valid
2c	0.91	Valid
2d	0.93	Valid
3a	0.94	Valid
3b	0.95	Valid
3c	0.92	Valid
3d	0.93	Valid

Construct Validation with Rasch Model

Testing the validity and reliability of test instruments to assess creative thinking skills can be done by applying the Rasch Model. The following table contains the results of the validity and consistency analysis that have been processed using the Rasch Model. These results provide an overview of the extent to which the research instrument used is able to measure the validity of the intended construct consistently and accurately. It is hoped that these results can provide confidence in the quality of the instruments used in this study.

1. Undimensionality

Undimensionality analysis aims to identify various aspects measured by the instrument. This process is carried out using the output menu table 23 in the Winsteps application version 3.73 by considering the raw variance explained by measures and the unexplained variance from the first to the fifth contrast [29]. Undimensionality can be said to exist if the raw variance explained by the measurement reaches ≥ 20%, with general interpretation categories: sufficient if in the range of 20-40%, good if 40-60%, and very good if more than 60%. And the unexplained variance value in the 1st to 5th residual contrast must be less than 15% [30]. The complete results of the Undimensionality analysis are described in Table 3.

Table 3. Results of Undimensionality Analysis

Table of Standardized Residual Variance (in Eigenvalue units)		
Raw variance explained by measures	=	33.4%
Unexplained variance in 1st contrast	=	14.2%
Unexplained variance in 2nd contrast	=	9.4%
Unexplained variance in 3st contrast	=	7.9%
Unexplained variance in 4th contrast	=	7.4%
Unexplained variance in 5th contrast	=	5.9%

Based on Table 3, the raw variance explained by the measures instrument is 33.4%, which is included in the sufficient category. Meanwhile, the unexplained variance in the first contrast is 14.2%; in the second, 9.4%; in the third, 7.9%; in the fourth, 7.4%; and in the fifth, 5.9%. All of these values indicate that the results are below 15%. Thus, it can be concluded that the instrument really measures what it should measure: students' creative thinking abilities.

2. Item Suitability Level

Item suitability refers to the extent to which each item in the instrument fits the measurement model used. Item suitability analysis ensures that each item in the test functions well in measuring construct validity. The assessment of item suitability is based on three main parameters, namely the Outfit mean square (MNSQ) value, Outfit z-standardized value (ZSTD), and point measure correlation value (PT-MEASURE CORR) with ideal values for the three main parameters Outfit Mean Square (Outfit MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PT-MEASURE-CORR) is in the range of 0.5 - 1.5, -2.0 - 2.0, and 0.4 - 0.85, respectively [26]. The output results of Item statistics: Misfit order on the Winstep application are shown in Table 4.

Table 4. Item Fit Output Results

Item	Outfit		PT Measure Corr
	MNSQ	ZSTD	
1D	1.25	1.6	0.50
2D	1.25	2.0	0.51
1C	1.27	2.1	0.39
3D	1.03	0.3	0.54
1A	0.93	-0.4	0.46
3C	0.96	-0.3	0.51
2A	0.90	-0.8	0.52
1B	0.86	-1.1	0.49
2C	0.89	-1.0	0.54
3A	0.85	-1.3	0.53
3B	0.79	-1.8	0.64
2B	0.76	-2.2	0.62

Table 4 presents the results of the item fit analysis sorted by the degree of misfit. Evaluation of the Outfit Mean Square (Outfit MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PT-MEASURE-CORR) values shows that all items are within the expected value range. Thus, it can be concluded that the twelve items on this creative thinking ability test instrument are declared fit. Based on item statistics: misfit order, the results show that the twelve items have met the validity criteria. However, there is one item that does not meet one of the criteria. Outfit ZSTD on question number 2B does not meet the criteria because it has a value of (-2.2). Although the item is still declared valid because it meets the criteria in the MNSQ and PT Measure Corr, to ensure the overall reliability of this test instrument, a revision is needed to question item 2B.

3. Item Difficulty Level

The difficulty level of each question item can be analyzed through Table 5. This shows the measurement sequence of question items in the Winsteps application. From the existing table, the standard deviation value is 0.38. By linking the standard deviation value to the average logit, the level of difficulty of the question items can be grouped into several categories, namely: very difficult (more than +1 SD), difficult (between 0.0 logit to +1 SD), easy (between 0.0 logit to -1 SD), and very easy (less than -1 SD). Thus, the value limit for the very difficult category is above 0.38, the difficult category is in the range of 0.0 to 0.38, the easy category is between 0.0 and -0.38, and the very easy category is below -0.38.

Based on the level of difficulty, the questions are sorted from the most difficult to the easiest. There is one question that falls into the very difficult category, namely question 3C. In addition, there are six questions that fall into the difficult category, namely questions 2B, 2C, 2D, 3A, 3B, and 3D. Three other questions fall into the easy category, namely questions 1C, 1C, and 2A. Finally, two questions that fall into the very easy category are questions 1A and 1D. However, the results of Table 3 found a gap between items categorized as "very difficult" and "very easy," which can cause the instrument to be less sensitive in measuring students' abilities at certain ranges so that it does not provide accurate information about students' abilities across the expected skill spectrum. Therefore, adjustments are needed in the preparation of items to achieve a more even distribution of difficulty levels and ensure that the instrument can evaluate various levels of student skills effectively.

Table 5. Item Statistics: Measurement Order

Item	Total Score	Measure
3C	303	0.65
2C	344	0.27
3D	345	0.26
2D	352	0.19
3A	354	0.17
2B	355	0.16
3B	367	0.04
1C	372	-0.01
1B	387	-0.16
2A	387	-0.16
1A	427	-0.62
1D	440	-0.79
	Mean	0.00
	SD	0.38

4. Instrument Reliability Using Rasch Model

The following are the results of the reliability test analysis carried out using the Rasch Model based on Table 6 and Table 7.

Table 6. Summary Results of Person Statistics

	Infit		Outfit	
	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	0.98	-0.1	0.98	-0.1
SD	0.52	1.4	0.53	1.4
MAX.	2.81	3.4	3.01	3.8
MIN.	0.17	-3.5	0.18	-3.3
Separation				1.80
Person reliability				0.76
Person raw score-to-measure correlation				0.97
Cronbach Alfa (KR-20) person raw score ”test” reliability				0.77

Table 7. Summary Results of Item Statistics

	Infit		Outfit	
	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1.01	0.0	0.98	-0.2
SD	0.21	1.7	0.18	1.4
MAX.	1.46	3.2	1.27	2.1
MIN.	0.75	-2.4	0.76	-2.2
Separation			3.39	
Item reliability			0.92	

Person statistics show the average score obtained by all respondents when working on the creative thinking test instrument. Suppose the average person value is greater than the average item value (where the average item value is 0.00 logit). In that case, this indicates that the respondent's overall ability is better than the difficulty level of the items contained in the instrument. The Cronbach Alpha reliability value is categorized into four levels, namely very good (0.80 - 1.00), good (0.70 - 0.80), sufficient (0.60 - 0.70), and poor (0.00 - 0.60) [31]. The Cronbach Alpha value indicating the interaction between the individual and the item as a whole is 0.77, which is included in the good category. In addition, the individual reliability value of 0.76 indicates the consistency of the respondent's answers and is also classified as good. Meanwhile, the item reliability reaching 0.92 indicates the quality of the items in the instrument and is classified as very good.

Based on the data in Table 4. the average value of Infit MNSQ was recorded at 0.98, while the value of Outfit MNSQ was also at 0.98. On the other hand, the Item Table shows that the average value of Infit MNSQ was 1.01 and Outfit MNSQ reached 0.98. Values closer to 1 indicate better quality, with 1 as the ideal value. Therefore, the average for persons and items is quite close to the expected criteria. Furthermore, the average value of Infit ZSTD for person is -0.1, and the average value of Outfit ZSTD for person is also -0.1. Meanwhile, the Infit ZSTD value for items is recorded at 0.0, and the Outfit ZSTD for items is -0.2. The ideal ZSTD value is 0, and the closer it is to that number, the better the quality. Thus, it can be concluded that the quality of individuals and items in this study is classified as good.

Finally, we will discuss the separation or grouping of people and items. Person separation indicates how effectively a series of items in a creative thinking test instrument can be spread across the logit ability range. A higher person separation value indicates better instrument quality because the test items are able to reach individuals with a wide range of abilities, from the highest to the lowest. On the other hand, item separation indicates how widely the sample used in the measurement is spread along a linear interval scale. A higher item separation value indicates that the measurement is better. This index is also useful for describing the meaning of the construct being measured. From Table 4, it can be seen that the separation value for individuals is 1.80, while in Table 5, the separation value for items is 3.39. The higher the separation value, the better the quality of the individual and the instrument as a whole. The separation value can be calculated using the formula H = ((4 x separation) + 1)/3. Based on the formula, the separation value for a person is 2.73, which is rounded to 3, while the separation value for an item is 4.85, which is rounded to 5. This shows that the respondents in this study have varying abilities, which can be grouped into three categories, while the level of difficulty of the items is divided into five groups, starting from the easiest to the most difficult.

Instrument validation is carried out by involving experts and using Aiken's V index to ensure that the instrument has covered all aspects of material, constructs, and language that are relevant to the measurement objectives. Good content validity is an indicator that this instrument can accurately reflect students' creative thinking skills. In addition, content validity also functions to assess the extent to which the questions in the developed instrument and the scores produced can measure the skills to be assessed [32]. Content validity also plays an important role in assessing the validity of questions that function as evaluation tools [33].

In this case, the research instrument has gone through a thorough validation process and meets the specified criteria, so it can be trusted and used with confidence by teachers and researchers. The validation results from five experts using Aiken's V formula showed that all questions were declared valid. Aiken's V values ranged from 0.91 to 0.95, which exceeded the minimum V table value of 0.87 (for five assessors with four assessment categories).

Based on the field test data, a construct validity test was conducted to ensure that the instrument used was in accordance with the expected theoretical construct. This construct validity aims to test whether the data obtained in the field can reflect the established construct [34]. In addition, a reliability test is very important to assess the extent to which this instrument can be trusted to measure student abilities consistently. An instrument is said to be reliable if it is able to produce the same results when tested on the same group at different times or situations. If the results obtained from the measuring instrument show significant variations, then the measuring instrument cannot be said to be reliable [35].

In this study, the Rasch Model is used as one of the statistical methods that allows researchers to obtain in-depth information about the characteristics and size of the test items. This model was first developed by Georg Rasch in the 1960s and has been used to assess various psychological aspects such as ability, attitude, and interest [33]. The study's results showed that the person reliability value reached 0.76, indicating that the consistency of the respondents' answers was at a good level. Meanwhile, the item reliability value of 0.92 provided information that the quality of the test items was at a very good level.

Furthermore, using the Rasch Model, the researcher ensured that the data obtained were accurate, objective, and consistent. The analysis showed that the dimensions of the creative thinking test instrument items were below 15%, indicating that all items met the suitability criteria. Of the twelve items analyzed, all were declared appropriate, with varying difficulty levels between very difficult, difficult, easy, and very easy. In addition, the Cronbach Alpha value indicating the interaction between respondents and the items as a whole was 0.77, indicating that this instrument has good reliability.

With the help of Winsteps version 3.7.3, the Rasch Model analysis produces accurate information about the instrument's reliability in providing consistent results in each measurement. Thus, thanks to the adequate reliability of the test results, this instrument can be used as a tool to assess students' creative thinking skills, especially in the context of physics learning.

IV. Conclusions

Validation from five experts with Aiken's V formula can be said that the results show that all questions are valid. The non-dimensionality of the questions based on the Rasch model shows that all results are less than 15%, the level of suitability of the twelve questions is stated as appropriate, the level of difficulty is spread across four categories of difficulty, the Cronbach Alpha value is 0.77, the person reliability value and item reliability are each 0.76 and 0.92 as indicators of the quality of the questions in the instrument are at a special level.

The research has implications for producing a creative thinking ability test in physics learning that is proven to be valid in terms of content and construct so that it can be used for research or references for the construction of similar test instruments and has practical implications in designing learning in developing students' creative thinking. However, not all physics materials and categories in Rasch modeling are displayed in this article, so in the future, it can analyze other topics in physics learning and the use of all assessment categories of instruments in Rasch modeling.

Acknowledgments

The author would like to thank all instrument validators, namely Mrs. Dra. Hj. Heni Rusnayati, M.Sc., Mrs. Dr. Hera Novia, MT, Mrs. Nurdini, M.Pd., Mr. Wildan Ade Gunawan, S.Pd., Mrs. Nisa Alyaumusyifa, S.Pd., and the schools where the test trials were conducted, namely SMA Negeri 1 Rongga and SMA Pasundan 8 Bandung.

References

[1] OECD, *PISA 2022 Results (Volume III): Creative Minds, Creative Schools*. in PISA. Paris: OECD Publishing, 2024. doi: 10.1787/765ee8c2-en.

[2] N. Podungge, D. D. Paramata, and A. H. Odja, "Penerapan E-Learning Berbantuan Media Pembelajaran Edmodo Terhadap Hasil Belajar Siswa pada Konsep Gejala Pemanasan Global," *Konstan - Jurnal Fisika dan Pendidikan Fisika*, vol. 5, no. 2, pp. 77–85, Dec. 2020, doi: 10.20414/konstan.v5i2.55.

[3] D. K. Fitra, "Pembelajaran Berdiferensiasi dalam Perspektif Progresivisme pada Mata Pelajaran IPA," *Jurnal Filsafat Indonesia*, vol. 5, no. 3, pp. 250–258, Oct. 2022, doi: 10.23887/jfi.v5i3.41249.

[4] L. Charli, A. Amin, and D. Agustina, "Kesulitan Siswa dalam Menyelesaikan Soal Fisika pada Materi Suhu dan Kalor di Kelas X SMA Ar-Risalah Lubuklinggau Tahun Pelajaran 2016/2017," *Journal of Education and Instruction (JOEAI)*, vol. 1, no. 1, pp. 42–50, Jun. 2018, doi: 10.31539/joeai.v1i1.239.

[5] S. Fajrina, L. Lufri, and Y. Ahda, "Science, Technology, Engineering, and Mathematics (STEM) as A Learning Approach to Improve 21st Century Skills: A Review," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 07, pp. 95–104, Jun. 2020, doi: 10.3991/ijoe.v16i07.14101.

[6] M. P. Simanjuntak, N. Bukit, Y. D. A. Sagala, R. K. Putri, Z. L. Utami, and M. Motlan, "Desain Pembelajaran Berbasis Proyek Terhadap 4C," *Inpafi: Inovasi Pembelajaran Fisika*, vol. 7, no. 3, pp. 38–46, 2019, doi: 10.24114/inpafi.v7i3.14570.

[7] D. Sambada, "Peranan Kreativitas Siswa Terhadap Kemampuan Memecahkan Masalah Fisika dalam Pembelajaran Kontekstual," *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, vol. 2, no. 2, p. 37, Dec. 2012, doi: 10.26740/jpfa.v2n2.p37-47.

[8] M. A. Runco and S. Acar, "Divergent Thinking as an Indicator of Creative Potential," *Creat Res J*, vol. 24, no. 1, pp. 66–75, Jan. 2012, doi: 10.1080/10400419.2012.652929.

[9] I. Tiyaswati, "Pengembangan Perangkat Pembelajaran IPA Dengan Pendekatan Scientific dan Model Search, Solve, Create and Share (SSCS) Untuk Meningkatkan Keterampilan 4CS (Communication, Collaboration, Critical Thinking and Creativity) Siswa," Universitas Sebelas Maret, 2021. [Online]. Available: https://digilib.uns.ac.id/dokumen/detail/89626/Pengembangan-Perangkat-Pembelajaran-IPA-Dengan-Pendekatan-Scientific-dan-Model-Search-Solve-Create-and-Share-SSCS-Untuk-Meningkatkan-Keterampilan-4CS-Communication-Collaboration-Critical-Thinking-and-Creativi

[10] B. I. N. Obi, T. I. Eze, and N. F. Chibuzo, "Experiential learning activities in business education for developing 21st century competencies," *Journal of Education for Business*, vol. 97, no. 1, pp. 36–42, Jan. 2022, doi: 10.1080/08832323.2021.1884521.

[11] U. Munandar, *Pengembangan kreativitas anak berbakat*. Jakarta: Rineka cipta, 2014.

[12] W. Hu and P. Adey, "A scientific creativity test for secondary school students," *Int J Sci Educ*, vol. 24, no. 4, pp. 389–403, Apr. 2002, doi: 10.1080/09500690110098912.
[13] M. H. Tumangger, K. Kartono, and S. Ridlo, "Analisis Kebutuhan Pengembangan Instrumen Penilaian Kemampuan Berpikir Kreatif pada Mata Pelajaran Matematika untuk Siswa Sekolah Dasar," *Basicedu Journal*, vol. 6, no. 5, pp. 9170–9179, 2022, doi: 10.31004/basicedu.v6i5.3623.
[14] R. Widyatiningtyas, B. A. Sritumini, and S. Rohartati, "Pembuatan Alat Tes Kemampuan Berpikir Kreatif Bagi Guru SMP Kemala Bhayangkari Bandung," *Jurnal Pengabdian Tri Bhakti*, vol. 3, no. 1, pp. 52–56, Jun. 2021, doi: 10.36555/tribhakti.v3i1.1703.
[15] A. N. Solihat, ""Pengembangan instrumen berpikir kreatif menggunakan item response theory (IRT) dalam pembelajaran ekonomi dengan model konstruktivis," Universitas Pendidikan Indonesia, 2024. [Online]. Available: http://repository.upi.edu/123662/1/D_PEKO_2104932_TITLE.pdf
[16] R. P. Z, R. Sari, J. Jumadi, and A. Ariswan, "Pengembangan dan Validasi Instrumen Tes untuk Mengukur Keterampilan Menyelesaikan Masalah Peserta Didik SMA pada Pelajaran Fisika," *Jurnal Penelitian Pembelajaran Fisika*, vol. 11, no. 1, pp. 17–26, Mar. 2020, doi: 10.26877/jp2f.v11i1.3993.
[17] Aulia Sefti Rahmadhani and Selvia Erita, "Pengembangan Dan Uji Validitas Soal-Soal Kemampuan Pemahaman Konsep Pada Materi Soal Himpunan," *Jurnal Bima : Pusat Publikasi Ilmu Pendidikan bahasa dan Sastra*, vol. 2, no. 1, pp. 139–147, Jan. 2024, doi: 10.61132/bima.v2i1.553.
[18] S. Yusuf, N. Budiman, E. S. Yudha, D. Suryana, and S. M. J. B. Yusof, "Rasch Analysis of the Indonesian Mental Health Screening Tools," *Open Psychol J*, vol. 14, no. 1, pp. 198–203, Sep. 2021, doi: 10.2174/1874350102114010198.
[19] H. Y. Muslihin, D. Suryana, A. Ahman, U. Suherman, and T. H. Dahlan, "Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using Rasch Model," *International Journal of Instruction*, vol. 15, no. 2, pp. 207–222, Apr. 2022, doi: 10.29333/iji.2022.15212a.
[20] D. Indihadi, D. Suryana, and A. B. Ahmad, "The Analysis of Construct Validity of Indonesian Creativity Scale using Rasch Model," *Creativity Studies*, vol. 15, no. 2, pp. 560–576, Aug. 2022, doi: 10.3846/cs.2022.15182.
[21] L. Nur, A. Yulianto, D. Suryana, A. A. Malik, M. A. Al Ardha, and F. Hong, "An Analysis of the Distribution Map of Physical Education Learning Motivation through Rasch Modeling in Elementary School," *International Journal of Instruction*, vol. 15, no. 2, pp. 815–830, Apr. 2022, doi: 10.29333/iji.2022.15244a.
[22] I. H. Yakin, *Metodologi Penelitian (Kuantitatif & Kualitatif)*. Garut: Aksara Global Akademia, 2023.
[23] Agustini *et al.*, *Metode Penelitian Kualitatif: Teori dan Panduan Praktis Analisis Data Kualitatif*. Mifandi Mandiri Digital, 2024.
[24] "Torrance Tests of Creative Thinking," in *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, 2455 Teller Road, Thousand Oaks, California 91320: SAGE Publications, Inc., 2018. doi: 10.4135/9781506326139.n705.
[25] L. R. Aiken, "Three Coefficients for Analyzing the Reliability and Validity of Ratings," *Educ Psychol Meas*, vol. 45, no. 1, pp. 131–142, Mar. 1985, doi: 10.1177/0013164485451012.
[26] B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch: pada Assessment Pendidikan*. Penerbit Trim Komunikata, 2015.
[27] N. R. An Nabil, I. Wulandari, S. Yamtinah, S. R. D. Ariani, and M. Ulfa, "Analisis Indeks Aiken untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia," *PAEDAGOGIA*, vol. 25, no. 2, p. 184, Sep. 2022, doi: 10.20961/paedagogia.v25i2.64566.
[28] H. Retnawati, *Validitas Reliabilitas & Karakteristik Butir*. Yogyakarta: Parama Publishing, 2020.
[29] G. Engelhard, Jr. and J. Wang, *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. New York: Routledge, 2024. doi: 10.4324/9781003458746.
[30] C. Han, "Rasch Analysis in the Human Sciences," *Journal of Research Design and Statistics in Linguistics and Communication Science*, vol. 5, no. 1–2, pp. 208–211, Dec. 2019, doi: 10.1558/jrds.37535.
[31] T. G. Bond and C. M. Fox, *Applying the Rasch Model : Fundamental Measurement in the Human Sciences*. Psychology Press, 2013. doi: 10.4324/9781410614575.
[32] L. Sürücü and A. Maslakçi, "Validity and Reliability in Quantitative Research," *Business & Management Studies: An International Journal*, vol. 8, no. 3, pp. 2694–2726, Sep. 2020, doi: 10.15295/bmij.v8i3.1540.
[33] S. S. Fahmina, M. Masykuri, D. G. Ramadhani, and S. Yamtinah, "Content validity uses Rasch model on computerized testlet instrument to measure chemical literacy capabilities," 2019, p. 020023. doi: 10.1063/1.5139755.
[34] S. Azwar, "Reliabilitas dan Validitas Aitem," *Buetin Psikologi*, vol. 3, no. 1, 1995, [Online]. Available: https://jurnal.ugm.ac.id/buletinpsikologi/article/view/13381/9595
[35] S. Azwar, *Reliabilitas dan validitas*. Pustaka Pelajar, 2012.