Journal of English Language and Education
ISSN 2597- 6850 (Online), 2502-4132 (Print)
Journal Homepage: https://jele.or.id/index.php/jele/index
Article

Construct Validation of Science Literacy Instrument with Rasch Modeling on Students of Grade VIII Junior High School

https://doi.org/10.31004/jele.v10i1.652

* Fahriana Alfiokti Tabacholly, Purwo Susongko, Dewi Amaliah Nafiati [1,2,3]
[1,2,3] Universitas Pancasakti Tegal
Corresponding Author: fahrianatabacholly10@guru.smp.belajar.id

ABSTRACT

This study aims to develop and analyze science literacy assessment instruments for grade VIII junior high school students using Rasch modeling. The development and analysis of the instrument was carried out with an approach that includes content, substantive, structural, and consequential aspects. The science literacy assessment instrument was developed based on eight science learning materials for grade VIII semester 1, designed to represent science literacy competencies according to PISA standards, namely: explaining phenomena scientifically, interpreting data and evidence scientifically, and evaluating and designing scientific enquiry. The questions are presented in the form of multiple choice with a total of 48 itemsThis study used the Research and Development (R&D) method with Tessmer's Formative Research design. The research approach applied is mixed methods, which combines qualitative and quantitative approaches. Research data were obtained from 158 students at SMPN 7 Brebes. The results of data analysis using the Rasch model with the R version 4.4.2 program and the eRm package showed that 85% of the items were valid with a p value <0.01. Based on validation covering content, psychometric, and construct aspects (consisting of content, substantive, structural, and external), this instrument meets the requirements as a good quality item.

Keywords: *Instrument Validation, Science Literacy, Rasch Model, Junior High School Students*

Article History:
Received 5th January 2025
Accepted 21st January 2025
Published 02nd February 2025

INTRODUCTION

Science is a scientific discipline that explores phenomena through a process known as the scientific method, which is based on scientific attitudes. This makes learning science more meaningful. With this approach, students can master and apply science knowledge and concepts effectively. Science learning is contextual, because it is closely related to everyday life (Syarif, 2020). The specific purpose of science learning is to form students to have a scientific attitude, curiosity, the ability to think logically, critically, and analytically, as well as an honest and responsible attitude. In addition, students are expected to be able to ask questions related to science phenomena, conduct experiments, record and present the results of investigations in the form of tables and graphs, then draw conclusions and report the results both orally and in writing to answer these questions (Diniya, 2019). The success of science learning is reflected when students can apply the knowledge they have learned to solve problems in everyday life, one of which is through science literacy (Pratiwi et al., 2019).

The level of science literacy in Indonesia is still relatively low, which is reflected in the results of the PISA and TIMSS assessments. Indonesia began participating in TIMSS in 1999 and PISA since 2000. TIMSS (Trends in International Mathematics and Science Study) aims to assess the development of math and science skills of 13-year-old students (grade VIII SMP/MTs) and is conducted every four years. Based on the TIMSS 2015 results, Indonesian students' performance in math and science literacy is not satisfactory, with a score of 397 for science literacy, placing Indonesia in 45th place out of 48 countries, far below the international average of 500. Meanwhile, PISA (Programme for International Student Assessment) evaluates the ability of 15-year-old students in science, reading and math every three years. Based on the 2018 PISA report, Indonesia ranked 70 out of 78 countries in science literacy with a score of 396, which is also below the international average of 489, (Sriyatun, 2024).

The low science literacy skills of Indonesian students are generally caused by learning activities that have not been oriented towards the development of science literacy (Sutrisna, 2021). For this reason, science literacy can be improved, one of which is through learning that focuses on developing science literacy so that students are trained in solving test questions that emphasize science literacy (Sumaryatun & Rusilowati, 2016). However, in reality, many schools have not implemented learning that focuses on developing science literacy. Aryani et al. (2016) suggested that teachers should design assessment tools in schools that not only cover aspects of concepts, but also include elements of science literacy in them. Therefore, teachers as professionals need to continue to innovate and adjust strategies, approaches, and learning technologies, (Nafiati, 2011).

So it is necessary to develop a PISA standardized science literacy instrument that has 3 science literacy competencies namely explaining phenomena scientifically, interpreting data and evidence scientifically, and evaluating and designing scientific enquiry, (Susongko, et al. 2019). To produce a quality science literacy instrument, a development research is needed. This research is a systematic process that includes the stages of analysis, planning, development, testing, and revision to create better and more effective products. Each development model has different characteristics. Some factors to consider in choosing a development research model are the purpose of the research, the type of research, and the resources available. One of the tools used in the assessment process using instruments is a written test. Written tests aim to measure students' understanding of science literacy. Test instruments need to be carefully designed so that they can accurately measure the aspects to be assessed. As stated by Imania & Bariah (2019) in (Lestari et al., 2022), "Referring to its important role in research, instruments must meet certain qualifications in accordance with scientific standards."

Therefore, in this study, a validity measurement is needed so that the instrument developed is truly in accordance with the purpose of its creation, namely to measure the pedagogical competence of teachers. In the process of making tests, item analysis is very important to evaluate the quality of the questions so that it can provide information related to the characteristics or empirical analysis of the items (Sarea & Hadi, 2015 in Mulyani et al., 2021). There are two types of approaches in testing questions, namely classical theory and modern test theory using Item Response Theory (IRT). IRT is one of the methods used to analyse question items. According to (Ayub et al., 2020), the IRT approach was developed as a solution to the weaknesses of classical measurement theory. Classical measurement theory (classical true score model) states that the characteristics of an item depend on the group of test takers and the ability of the individual being measured as a representation of performance values.

According to (Pudjiati & Madani, 2023), the Rasch Model, which is part of Item Response Theory (IRT), is used to determine the probability that a test taker can answer an item correctly based on their ability level. This model is based on the assumption that the ability of test takers can be measured accurately and objectively, and the items in the test have varying levels of difficulty and produce scores that are independent of the sample items used. The Rasch model also calibrates three main elements: the measurement scale, test takers, and items. This calibration aims to ensure the validity and reliability of the measurement results, so that the test can provide complete and comprehensive information (Ardiyanti, 2016 in Pratama, 2020). Based on this, researchers chose to use the Rasch Model to analyze science literacy test items, especially in measuring instrument validity. The research objectives can be formulated to prove the construct validity of the science literacy assessment instrument in class VIII junior high school students.

METHOD

This research uses a type of development research (Research and Development), namely the development of reading literacy assessment instruments to measure students' abilities. According to Sugiyono (2019:40), the Research and Development (R&D) method is a research method that aims to produce certain products and test the validity and effectiveness of these products. To produce the desired product, this research begins with a needs analysis, and to ensure that the product is effective and acceptable to the community, it is necessary to test the validity and effectiveness of the product. The research design used is Tessmer's Formative Research type development model. There are four stages, namely: preliminary stage, self-evaluation stage, formative evaluation (prototyping) stage which includes expert reviews, one-to-one, and small group; and field test stage, (Anggarsika et al, 2024). This development model is arranged programmatically and systematically to produce a science literacy assessment instrument that can measure the science literacy skills of VIII grade junior high school students. Thus, it is expected to obtain a measuring instrument that can map the true ability of students and minimize the level of error.

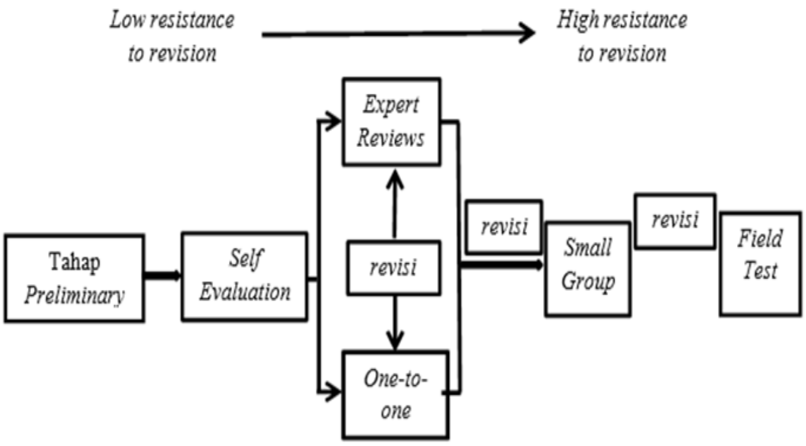


Figure 1: Stages of Tessmer's Formative Research Model

The research subjects were students of class VIII d SMPN 7 Brebes totaling 158 students. Researchers applied purposive sampling techniques to students of SMP N 7 Brebes. The number was chosen because it was considered capable of representing the results of the instrument trials using the Rasch model. In addition, the characteristics of students who become samples have been adjusted to the criteria of academic ability. The three criteria used include: high ability students as many as 53 people, medium ability as many as 53 people, and low ability as many as 52 people. By considering these criteria, it is expected that the science literacy assessment instrument can accommodate all levels of students' abilities.

There are three data collection techniques carried out in this study, namely, first, data collection techniques using interviews which are carried out by giving a set of written questions to respondents to answer which has a function so that researchers know exactly the variables to be measured and know what is wanted from respondents. Interviews were conducted with respondents, namely principals, teachers and students who aimed to find out how the respondents' needs for science literacy assessment instruments. The second test, according to Arikunto and Jabar (2004) in a book written by (Esty Aryani Safithry, 2018) suggests that a tool or procedure used to determine or measure something using certain methods or rules. Technical tests are carried out to develop products in the form of science literacy assessment instruments. The test is given to obtain the necessary data, after the test results are tested; the validity is then carried out using the Rasch model. Data collection techniques in the form of documentation support the research process and research results. The third researcher used a documentation study to examine the necessary documents related to the science literacy assessment instrument.

Data analysis was used to validate the science literacy assessment instrument using the Racsh model, in R programming version 4.4.2 with the *eRm* package, namely on the validity of the content aspects, substantive aspects, structural aspects and external aspects. The steps taken to validate are as follows: (a) Data preparation in the form of inputting test takers' answers. The data used are all the answers of the test participants then typed into notepad. (b) Validate the content aspect, (c) validate the substantive aspect. (c) Validate the structural aspects. (d) Validate the external aspects.

FINDINGS AND DISCUSSION

Construct Validity of the Content Aspects of the Science Literacy Instrument

Construct validity is validity that measures the extent to which test items are truly in accordance with established concepts or conceptual definitions (Djaali & Pudji Muljono, n.d.). Measurement of construct validity using the Rasch model on science literacy instrument questions totaling 48 questions that were tested on 158 test participants. Construct validity measured includes content aspect construct validity, substance aspect construct validity. Rasch model-based construct validity on dichotomous items is done by analyzing item fit to the model (Item fit). The content aspect construct validity test also has indicators in the form of item difficulty, respondent ability, item fit test, and respondent fit test. The results of the Item fit analysis of pedagogical literacy assessment items in table 1 below:

Table 1. Non-Fit Items in the Item Fit Analysis of the Science Literacy Instrument for Class VIII Junior High School Students

Item	Chisq	df	p-Value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
5	293.054	157	0.000	1.855	1.204	2.993	1.214
6	287.931	157	0.000	1.822	1.263	3.610	1.904
8	204.038	157	0.007	1.291	1.268	3.368	3.922
9	217.082	157	0.001	1.374	1.184	2.548	1.860
10	300.804	157	0.000	1.904	1.227	3.343	1.427
46	234.074	157	0.000	1.481	1.337	3.265	3.324
47	305.139	157	0.000	1.931	1.520	5.989	5.106

Item fit explains whether an item functions normally or not. Quantitatively, test items that are declared fit or can be said to behave consistently with what is expected by the first model are if the p-value is greater than the significance level (p> 0.01). From table 2, it can be seen that from 7 items out of 48 items of the science literacy instrument, there are items, namely questions 5, 6, 8, 9, 10, 46 and 47, that cannot be accepted as a suitable model with a significance level of 0.01. While the other 41 items can be accepted as a suitable item with a significance level of 0.01. Meanwhile, when viewed from the parameter of the difficulty level of the item if the P value is as follows: P ≤ -2 (items categorized as very easy), -2 ≤ P ≤ -1 (items categorized as easy), -1 ≤ P ≤ 1 (items categorized as medium), 1 ≤ P ≤ 2 (items categorized as difficult), and P ≥ 2 (items categorized as very difficult). The magnitude of the difficulty level value on each item can be seen in table 2.

Table 2. Value of the Level of Difficulty of the Items of the Science Literacy Measurement Instrument in Class VIII Junior High School Students

Question Item	Value	Question Item	Value

1	0.598	25	-0.340
2	-0.508	26	0.630
3	-0.974	27	-0.854
4	0.083	28	-0.593
5	1.810	29	-0.650
6	1.438	30	-0.795
7	0.798	31	0.026
8	0.258	32	-0.172
9	0.941	33	0.630
10	1.695	34	0.348
11	-0.593	35	-1.005
12	0.229	36	-0.200
13	0.318	37	-0.914
14	-0.424	38	-0.312
15	-1.067	39	-0.396
16	0.288	40	-0.593
17	0.141	41	-0.172
18	-0.941	42	0.055
19	1.174	43	0.288
20	0.764	44	0.440
21	-1.067	45	-0.825
22	-0.480	46	0.905
23	-0.424	47	0.833
24	-0.766	48	0.348

The table above shows the difference in the level of difficulty of each item. Based on the data output, it can be said that the level of difficulty of the items is within normal limits because it is still between the range of numbers -4 to 4. Based on table 3, there are 3 items classified as easy, 41 items classified as medium, 4 items classified as difficult. The percentage of categories in the difficulty level of each item is: 6.25% of the questions are categorized as easy, 85.42% are categorized as moderate, 8.33% are categorized as difficult. The difficulty level value ranges from -1,067 to 1,810. The item with the lowest difficulty value is in question number 15 and number 21, which is -1.067 and the highest difficulty value is in number 5, which reaches 1.810. The item difficulty value moves from -1.067 to 1.810. An effective test has an item difficulty level between -3.00 to 3.00.

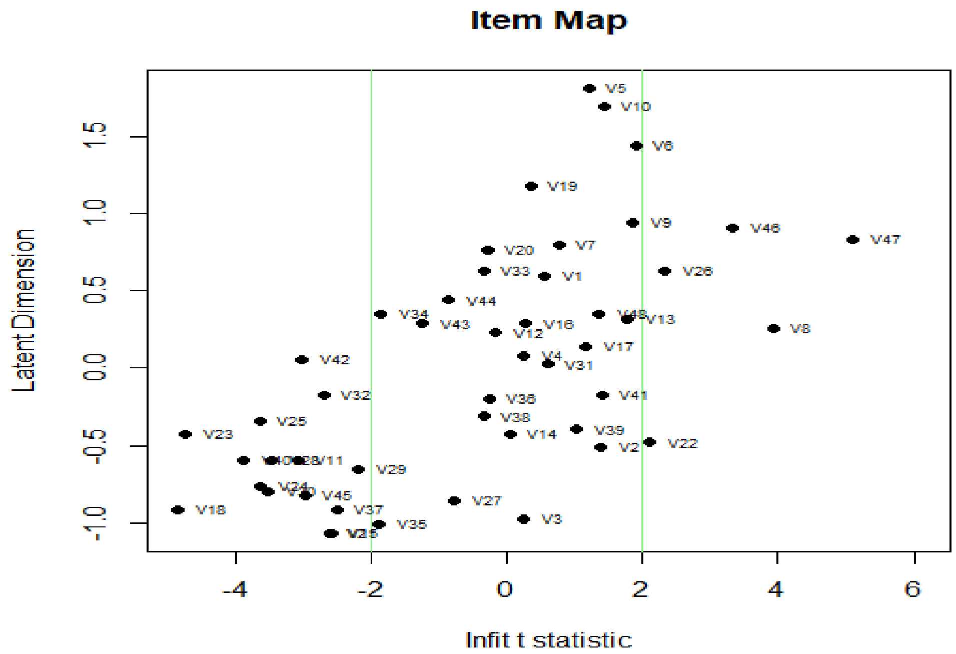


Figure 1. Item map of Science Literacy Assessment Measurement Instrument Items in Class VIII Junior High School Learners

Figure 2 is a picture of the item map of science measurement instruments with the help of eRm software. It can be seen that the item at the highest point is V5 which describes the most difficult item followed by number 10, and the lowest point is item number 21 which is the easiest item category.

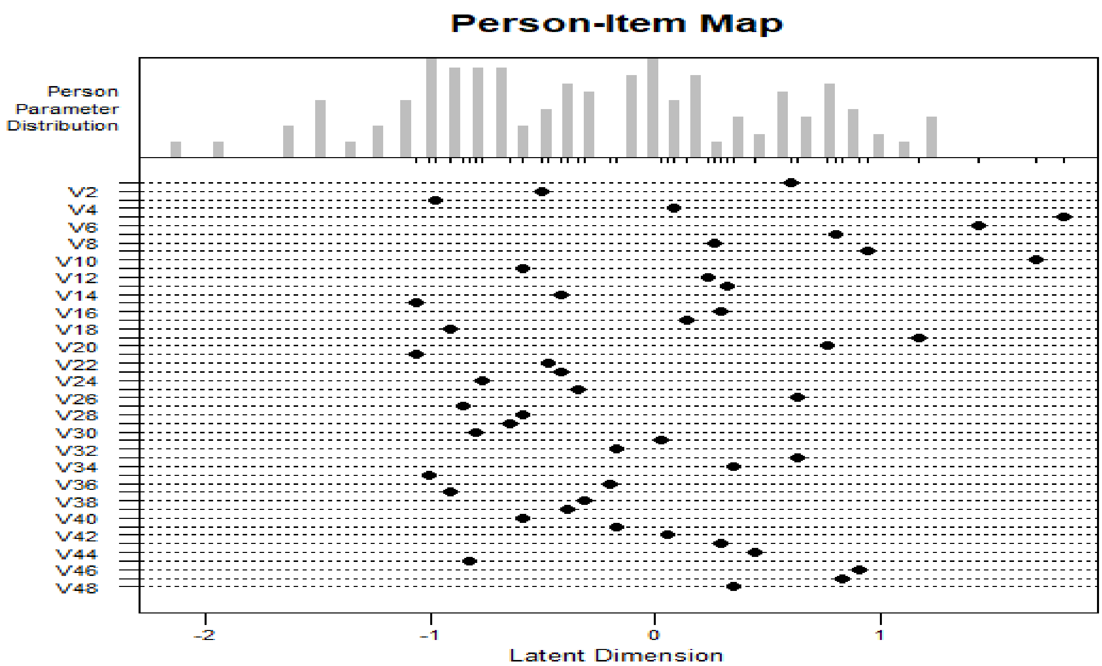


Figure 2. Person-Item Map of Science Literacy Measurement Instrument Items of Participants in Class VIII Junior High School

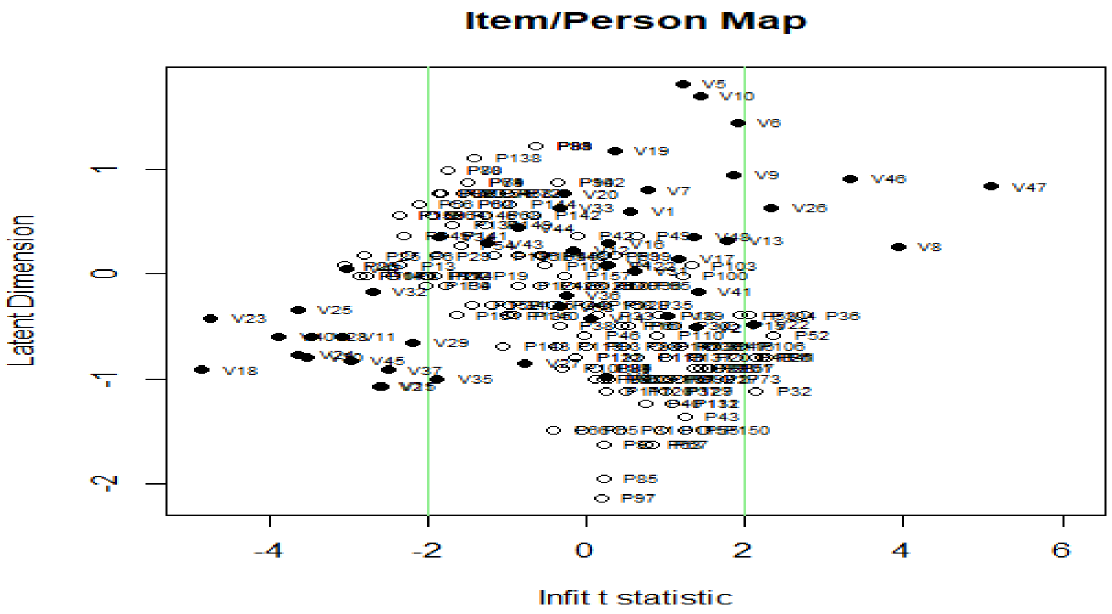


Figure 3. Item/Person Map of Science Literacy Measurement Instrument Items in Class VIII Junior High School Learners

The evidence that the science literacy assessment instrument items are effective for learner abilities between -2.00 and 2.00 is explained by the item and test information functions in the figure. The figure explains that the information function will be maximized at the interval of learner ability between 0 to 1.0 and effective between -2.00 to 2.00.

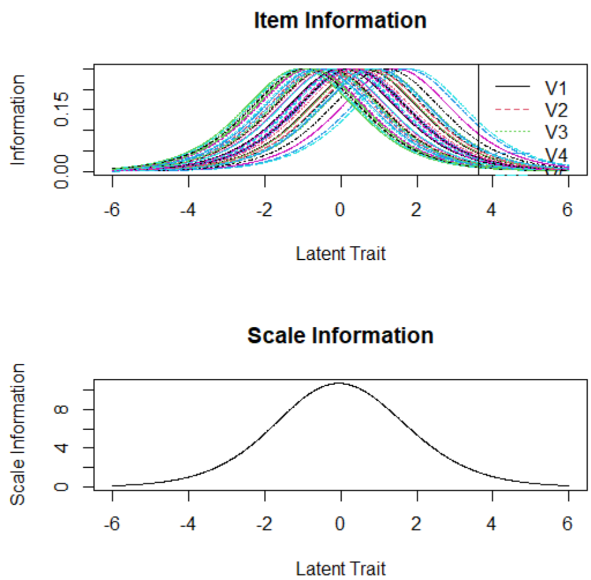


Figure 4: Information Function of the Items of the Science Literacy Assessment Instrument for Class VIII Junior High School Learners

Construct Validity of Substantive Aspects of the Pedagogical Literacy Assessment of Teachers in the Junior High School Environment

The quality of construct validity from the substantive aspect is used to test the test taker's ability to match the instrument. The substantive aspect is related to the content aspect. Achieved by empirically finding to ensure that the test taker actually involves the ability of the field being measured in answering the test items. Different response patterns are inconsistent test takers who have an ability (ϴ) of 1.5 should be able to answer all items that have a difficulty level below 1.5, but in the field there are certainly some students who are inconsistent or cause abberant responses. How many test takers experience this abberant response becomes substantive type construct validity.

This deviant response can be caused by inattention, cheating activities (cheating) or even misconceptions. Testing whether a person's response is deviant or not is called person fit. Quantitatively, the Outfit MSQ value response is between 0.5 to 1.5 while the outfit t value is between -2.0 to 2.0 and the chance of accepting Ho (model fit) is greater than 0.01 (p> 0.01). The table contains the results of the person fit test of 158 responses from test participants of the science literacy assessment instrument.

Table 3. Test takers with deviant responses *(Aberrant Response)*

Peserta	Chisq	df	p-Value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P2	73.482	47	0.008	1.531	1.214	2.23	1.48
P17	73.482	47	0.008	1.531	1.214	2.23	1.48
P31	76.196	47	0.004	1.587	1.073	1.77	0.43
P32	86.621	47	0.000	1.805	1.350	2.91	2.14
P37	75.823	47	0.005	1.580	1.149	2.22	0.99
P55	82.004	47	0.001	1.708	1.250	2.07	1.23
P63	73.227	47	0.008	1.526	1.158	1.50	0.76
P106	72.900	47	0.009	1.519	1.230	2.75	1.93
P111	73.244	47	0.008	1.526	1.178	1.90	1.08
P115	78.041	47	0.003	1.626	1.182	1.87	0.93
P132	73.244	47	0.008	1.526	1.178	1.90	1.08
P150	107.074	47	0.000	2.231	1.304	3.18	1.47

The results of the Person fit test for science literacy test participants on 158 test participants were 12 (twelve) test participants who experienced responses that deviated from the model. This is because twelve participants did not meet the specified p-value of 0.01. From the explanation in the table it can be concluded that there are 92.41% of test takers' responses that are reasonable according to the model or do not experience deviation (aberrant response), while 7.59% of test takers' responses experience deviation. The large percentage of test takers who have reasonable responses according to the model can be used as a basis that the test sufficiently meets substantive validity.

Construct Validity of Structural Aspects of Pedagogical Literacy Assessment of Teachers in Junior High School Environment

Indicators on the science literacy assessment instrument show construct validity in structural aspects, namely the test is unidimensional. A science literacy test designed based on a one-dimensional framework must be valid and have relevant meaning. The instrument consisting of 48 items from 8 learning materials was tested on 158 participants, then analyzed using the Rasch model. The following are the results of the unidimensionality test of the science literacy measurement instrument items in class VIII junior high school students

Test the unidimensionality of the science literacy assessment instrument using Kaiser-Meyer-Olkin factor adequacy which is used to assess the adequacy of the sample in factor analysis. KMO measures how well the existing data is suitable for factor analysis by assessing the proportion of variance among variables that may be common variance. The result was an MSA (Measure of Sampling Adequacy) value of 0.66, which is considered adequate for factor analysis. However, this value is at the minimum threshold, so data quality needs to be considered to get optimal results in factor analysis. For the KMO value per item, most variables have values above the 0.50 threshold, which indicates that most variables support factor analysis.

The MSA values for each item varied, ranging from 0.48 to 0.91. Some items had MSA values below 0.50, such as V6 (0.51), V36 (0.48), and V39 (0.52), indicating that these items were less suitable for inclusion in factor analysis due to their low contribution to the data structure. However, some items had high MSA values, such as V18 (0.91), V22 (0.84) and V32 (0.82), indicating that these items fit well into the factor structure and contributed well. The following are the results of the scree plot of the science literacy assessment instrument

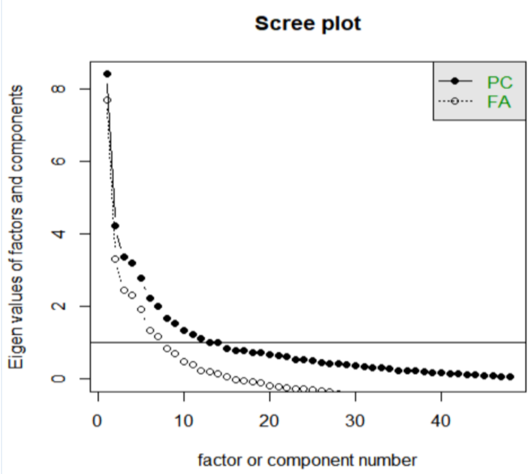
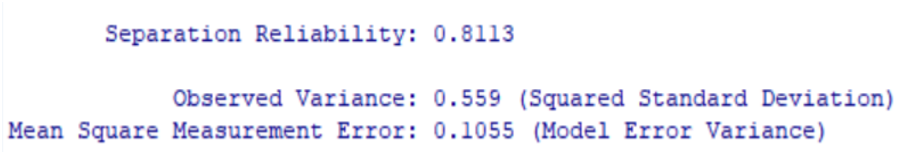


Figure 5. Scree Plot of Science Literacy Assessment Items in Class VIII Junior High School Students

The scree plot shows that the items on the science literacy instrument show one dimension. It can be seen from the graph that there is one dominant dimension. So based on the unidimensional feasibility criteria, the science literacy assessment in class VIII junior high school students is fulfilled. Because there are two eligibility criteria that are met, the first has a KMO value of 0.66 greater than the threshold limit of 0.5 and the second there is a dominant factor that has a higher eigenvalue than other factors.

Construct Validity of External Aspects of Pedagogical Literacy Assessment of Teachers in the Junior High School Environment

The science literacy assessment instrument from the external aspect construct validity is carried out to ensure that the research results can be generalized and applied to a wider context outside the sample studied. To determine the construct validity of external aspects, one of the approaches used is Person Separation Reliability or person separation. This person separation is used to enable researchers to understand the effectiveness of measurement instruments. By conducting this analysis, areas of improvement can be identified to increase the accuracy and reliability of the measuring instrument, as well as ensuring that the instrument can properly distinguish between individuals with different levels of ability. The results of the person separation analysis using eRm can be seen in the following figure:



```
Separation Reliability: 0.8113

        Observed Variance: 0.559 (Squared Standard Deviation)
Mean Square Measurement Error: 0.1055 (Model Error Variance)
```

Figure 6. ***Person separation Reliabilty on the Items of the Science Literacy Assessment Instrument in Class VIII Junior High School Learners***

From the data above, it can be seen that the Separation Reliability of 0.8113 indicates the level of reliability in distinguishing individuals based on their ability or level in the Rasch model. This value is in the high category (above 0.8), which means that the instrument is reliable enough to group participants consistently based on their ability. The higher this value, the better the instrument's ability to separate participants into different ability levels.

Furthermore, the Observed Variance of 0.559 reflects the total variance observed in the data, which is calculated as the square of the standard deviation. This value indicates the extent to which participants' scores are dispersed based on the measured data. A large variance indicates a wider spread, while a small variance indicates a more concentrated distribution. The Mean Square Measurement Error (Model Error Variance) result of 0.1055 indicates the average amount of error generated by the model in measuring participants' abilities. This value is relatively small compared to the observed variance, which indicates that the model does a good job of reducing measurement error.

CONCLUSIONS

The test construction of the Science Literacy Assessment Instrument for Class VIII Junior High School Students in the form of multiple choice 1) At the 99% confidence level there are 41 items that match the modeling, (2) The level of item difficulty is in the range of -3 to 3, meaning that the test items are suitable for the test takers' abilities. (3) At the 99% confidence level, 92.41% of test takers' responses matched the modeling. (4) The science literacy assessment instrument for grade VIII junior high school students has a unidimensional concept eligibility criterion that each item only measures one ability. (5) The science literacy assessment instrument for grade VIII junior high school students distinguishes test takers into two categories, namely test takers who have minimal adequacy and have not experienced minimal adequacy.

REFERENCES

Anggarsika, P., Mulyono, T., & Muljani, S. (2024). Pengembangan Instrumen Penilaian Literasi Membaca dalam Pembelajaran Bahasa Jawa pada Peserta Didik SMP. Journal of Education Research, 5(3), 3238–3244. https://doi.org/10.37985/jer.v5i3.1441

Aryani, A. K., Suwono, H., & Parno. (2016). Profil kemampuan literasi sains siswa SMPN 3 Batu. In Prosiding Seminar nasional Pendidikan IPA Pascasarjana UM

Diniya, D. (2019). Pembelajaran IPA Terpadu Tipe Integrated Melalui Model Inkuiri Terbimbing Tingkat Sekolah Menengah Pertama. Journal of Natural Science and Integration, 2(2), 22. https://doi.org/10.24014/jns.v2i2.7580

Djaali dan Muljono, Pudji. (2020). Pengukuran Dalam Bidang Pendidikan. Jakarta: PT Grasindo

D. Pratama, "Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch," Tarbawy J. Pendidik. Islam, vol. 7, no. 1, pp. 61–70, 2020, doi: 10.32923/tarbawy.v7i1.1187.

Esty Aryani Safithry, asesmen teknik tes dan non tes. Malang: CV IRDH, 2018. [Online]. Available: https://www.google.co.id/books/edition/ASESMEN_TEKNIK_TES_DAN_NON_TES/EcbODwAAQBAJ?hl=id&gbpv=1&dq=asesmen+teknik+tes+dan+non+tes&printsec=frontcover

Indri Pudjiati and Faisal Madani, "Asesmen Autentik Analisis Butir Soal Dengan Rasch Model Di Sekolah Dasar: Literature Review," Mutiara J. Penelit. dan Karya Ilm., vol. 1, no. 4, pp. 01–12, 2023, doi: 10.59059/mutiara.v1i4.325.

Nafiati, D. (2011). Meningkatkan Efektivitas Pembelajaran Melalui Penggunaan Multimedia. Https://Cakrawala.Upstegal.Ac.Id/Index.Php/Cakrawala/Article/View/338, 1–7.

W. Lestari, I. Wigati, M. I. Sholeh, D. Desi Pramita, U. Islam Negeri Raden Fatah Palembang, and S. Selatan, "Instrumen Literasi Digital Guru Menggunakan Model Rasch," Orbital J. Pendidik. Kim., vol. 6, no. 2, 2022.

Pratiwi, S. N., Cari, C., & Aminah, N. S. (2019). Pembelajaran IPA abad 21 dengan literasi sains siswa. Jurnal Materi Dan Pembelajaran …, 9, 34–42.

https://jurnal.uns.ac.id/jmpf/article/view/31612%0Ahttps://jurnal.uns.ac.id/jmpf/article/download/31612/21184

Purwo Susongko, Mobinta Kusuma, Y. A. (2019). Model Asesmen Literasi Sains Berbasis Ipa Terpadu. 1–118.S. Mulyani, R. Efendi, and T. R. Ramalis, "Karakterisasi Tes Keterampilan Pemecahan Masalah Fisika Berdasarkan Teori Respon Butir," J. Pendidik. dan Ilmu Fis., vol. 1, no. 1, p. 1, 2021, doi: 10.52434/jpif.v1i1.1006.

M. R. S. S. N. Ayub, E. Istiyono, S. Munadi, C. Permadi, A. Pattiserlihun, and D. N. Sudjito, "Analisa Penilaian Soal Fisika Menggunakan Model Rasch Dengan Program R," J. Sains dan Edukasi Sains, vol. 3, no. 2, pp. 46–52, 2020, doi: 10.24246/juses.v3i2p46-52.

Sriyatun, Siti. (2024). PISA dan TIMSS sebagai Acuan AKM. https://www.gurusiana.id/read/sitisriyatun/article/pisa-dan-timss-sebagai-acuan-akm-3711194

Sugiyono. (2017). Metode Penelitian Kuantitatif Kualitatif Dan R & D. Bandung: Alfabeta.