Thabiea : Journal of Natural Science Teaching
Vol. 6(1), pp. 90-105, 2023
Available online at
http://journal.iainkudus.ac.id/index.php/Thabiea
p-issn: 2580-8474, e-issn: 2655-898X

Construct validity analysis with messick validity approach and rasch model application on scientific reasoning test items

Yuni Arfiani[1]*, Purwo Susongko[2], Mobinta Kusuma[3]
[1,2,3]Universitas Pancasakti Tegal, Jl. Halmahera KM.1, Tegal, 52122, Indonesia
**Correspondence: yuniarfiani@upstegal.ac.id*

Keywords:
Construct Validity;
Messick Validity;
Rasch Model;
Scientific Reasoning;

Abstract

The Scientific Reasoning Test is basically inseparable from the three aspects of the Scientific Literacy Test. This study aims to test the feasibility of the instrument for measuring scientific reasoning in terms of content validity, psychometrics, and constructs. The form of this research is Research and Development (Research and Development). In the instrument development research design using the ADDIE procedural model (Analysis, Design, Development, Implementation, Evaluation). This test was given to 194 high school students in Math and Natural Sciences class XII in the 2022-2023 academic year. Analysis of the validity of the test using Rasch modeling. The type of validity applied is Messick's validity which includes content validity, psychometric validity and construct validity consisting of content aspects, substantive aspects, structural aspects and external aspects. The results show that of the 50 questions made, there are 25 items that are feasible to use as measument in scientific reasoning test. This research contributes to the improvement of assessment practice in science education.

To cite this article:
Arfiani, Y., Susongko, P., Kusuma, M. (2023). Construct validity analysis with messick validity approach and rasch model application on scientific reasoning test items. *Thabiea : Journal of Natural Science Teaching*, 6(1), 90-105.

Introduction

High social science literacy has a very significant effect on the progress of a nation. This is because community scientific literacy has a positive effect on the quality of economic development, democracy, and societal culture (Hanushek & Woessmann, 2016; Rudolph & Horibe, 2016; Roth & Lee 2016). Therefore, the achievement of students' scientific literacy must be the main goal in science education universally (McFarlane, 2013). Many efforts have been made to increase students' scientific literacy in science learning both through developing science learning models and developing assessments (Fakhriyah *et al.*, 2019; Ratini et al., 2018; Ardianto & Rubini, 2016; Rusilowati et al., 2016).

In Indonesia, the goals of science education in high schools aim to: (1) build and apply information, knowledge, and technology logically, critically, creatively, and innovatively; (2) demonstrate the ability to think logically, critically, creatively, and innovatively independently; (3) demonstrate the ability to analyze and solve complex problems; (4) demonstrate the ability to analyze natural phenomena, utilize the environment productively and responsibly, and master the knowledge needed for higher levels of education (Ministry of National Education, 2006). This is in line with the scientific literacy skills developed by PISA (Program for International Science Student Assessment) 2015, which include (1) explaining phenomena

scientifically, (2) evaluating and designing scientific investigations, and (3) interpreting data and evidence scientifically (Chiang & Tzou, 2018).

According to Hanson (2016), at the senior high school level, global scientific literacy consists of at least three dimensions: scientific reasoning skills, scientific inquiry skills, and knowledge of the nature of science. After we conducting need analysis in the preliminary research, there is currently no study on the development of scientific literacy instruments in Indonesia that combines the three dimensions of scientific literacy in a comprehensive manner. Similarly, few researchers have used the Rasch model approach to validation or digital applications in their presentations (Susongko, 2016; Susongko et al., 2019; Susongko et al., 2021)

The Scientific Reasoning Test is a test that measures scientific reasoning ability for high school students or its equivalent. Scientific reasoning basically consists of abilities in: (1) mathematical deduction; (2) evaluation of experiments; (3) making hypothetical models; (4) categorization and classification; (5) probabilistic reasoning; and (6) history-based evolutionary reasoning (Kind & Osborne, 2017).

The Scientific Reasoning Test is basically inseparable from the three aspects of the Scientific Literacy Test. The concept of scientific reasoning in this test is formulated based on the theoretical framework developed by (Wenning & Vierva ,2015), and adapted according to the context of the curriculum that applies in Indonesia. The difference is that divide the category of scientific reasoning into six levels: rudimentary, basic, intermediate, integrated, culminating, and advanced, whereas in this test it only limits it to the culminating level. This is because competence at the advanced level is generally above the level of ability of high school students. This is supported by the results of (Hanson (2016) research, which measures scientific reasoning abilities at the high school level only up to the culminating level. There are 25 indicators measured in this scientific reasoning test, consisting of 5 competencies: rudimentary (7 competencies), basic (5 competencies), intermediate (5 competencies), integrated (3 competencies), and culminating (5 competencies). This test consists of 25 items that measure the 25 indicators.

As has been explained, scientific literacy is the goal of science education universally, and its achievement needs to be known so that the effectiveness of learning science can be realized. To achieve this, a test instrument that has good quality and can measure scientific literacy achievements in a comprehensive manner is needed. The test must be built using a strong test theory approach so that all aspects of its validity are guaranteed.

There are two approaches to validating test items, namely, the use of classical test theory and modern test theory. Modern test theory has two methods that are paradigmatically different, namely Rasch and Item Response Theory. In modern measurement theory tests, the Rasch model is seen as the most objective measurement model. The concept of objective measurement in educational assessment has five criteria, namely: (1) producing a linear measure at the same interval; (2) the right estimation process; (3) identifying misfits or outliers; (4) being able to overcome missing data; and (5) producing measurements that are independent of the parameters studied (Bond et al., 2020). So far, only the Rasch model can meet these five requirements.

Validity with Rasch modeling refers to Messick's validity, where construct validity is considered a single concept consisting of several aspects (Ravand & Firoozi, 2016; Chiang & Tzou, 2018). Messick's validity is the demonstration that an instrument is valid in at least five

ways: content validity, substantive validity, structural validity, external validity, and consequential validity. Rasch analysis explains construct validity, which is more comprehensive than classical test theory. There are at least six aspects of construct validity, namely content, substantive, structural, external, generalizable, and consequential aspects (Kuo et al., 2015; Wang et al., 2014; Jong et al., 2015).

In connection with the need for a standardized test to measure one aspect of scientific literacy, namely the aspect of scientific reasoning, and the availability of a fairly comprehensive test validation model based on the Rasch model, it is necessary to develop an adaptive scientific reasoning test for high school students in Indonesia. The purpose of this study was to validate the construct of a scientific reasoning test for high school students in the MIPA program in Indonesia using the Rasch model with the Messick validity approach.

Method

The form of this research is Research and Development. The object of this study was an instrument for assessing the scientific literacy of high school students in the Mathematics and Natural Sciences program on the aspect of scientific reasoning which was compiled, revised, and validated using the Rasch model. In the instrument development research design using the ADDIE procedural model (Analysis, Design, Development, Implementation, Evaluation). This test was given to 194 high school students in Math and Natural Sciences class XII in the 2022-2023 academic year, ranging in age from 17 to 18 years. There were as many as 50 test items that were tested. Analysis of test validity using Rasch modeling. The type of validity applied is Messick Validity (Messick, 1995), which includes content validity, psychometric validity, and construct validity consisting of content aspects, substantive aspects, structural aspects, and external aspects. Construct validity uses the Rasch modeling application based on the concept of Messick validity, which has been formulated by Susongko, (2016), as described in Table 1 below.

Table 1. Construct Validity Criteria with the Messick Approach

Construct Validity Aspects	Indicator	Criteria
Content	Itemfit	P > 0.05
Substantive	Person fit statistic accuracy, sensitivity, dan specificity	P > 0.05 close to 1,0
Structural	Invariansi Test (LRtest)	P< 0,05
External	Separation Person Strata	close to1,0

Results and Discussion

The item suitability test in the scientific reasoning test can be seen in Table 2, while the item difficulty level can be seen in Table 3. Student responses that do not match the model can be seen in Table 4. Table 2 shows that item numbers 4, 7, 11, 15, 35, 38, and 39, or as many as 6 out of 50 test items, do not match the Rasch modeling used. Thus, these items cannot be used as precise items for measuring scientific reasoning abilities.

Table 2. Item Fit Test on Scientific Reasoning Tests

Items	Nilai Chi Square	P Value	Description	Items	Nilai Chi Square	P Value	Description
1	164.018	0.936	Yses	26	163.219	0.941	Yes
2	141.835	0.998	Yes	27	172.724	0.850	Yes
3	217.068	0.113	Yes	28	172.129	0.857	Yes
4	232.333	0.028	No	29	176.355	0.799	Yes
5	221.031	0.081	Yes	30	151.033	0.989	Yes
6	179.542	0.748	Yes	31	195.698	0.432	Yes
7	300.853	0.000	No	32	183.773	0.671	Yes
8	180.609	0.729	Yes	33	167.859	0.904	Yes
9	152.309	0.986	Yes	34	216.094	0.122	Yes
10	188.100	0.586	Yes	35	551.280	0.000	No
11	227.422	0.045	No	36	208.579	0.210	Yes
12	198.718	0.374	Yes	37	169.725	0.885	Yes
13	217.902	0.106	Yes	38	347.975	0.000	No
14	185.034	0.647	Yes	39	231.313	0.031	No
15	165.284	0.926	Yes	40	191.466	0.518	Yes
16	188.534	0.577	Yes	41	150.356	0.990	Yes
17	207.816	0.221	Yes	42	164.845	0.930	Yes
18	200.464	0.341	Yes	43	182.413	0.697	Yes
19	140.317	0.998	Yes	44	203.973	0.280	Yes
20	165.340	0.926	Yes	45	172.025	0.859	Yes
21	196.339	0.420	Yes	46	162.806	0.944	Yes
22	220.492	0.085	Yes	47	176.691	0.794	Yes
23	171.225	0.868	Yes	48	191.053	0.526	Yes
24	200.203	0.346	Yes	49	207.855	0.220	Yes
25	183.770	0.671	Yes	50	179.444	0.749	Yes

Table 3. Difficulty of Scientific Reasoning Test Items

Items	Difficulty Level	Error Standar	Items	Difficulty Level	Error Standar
1	-1.402	0.195	26	-0.729	0.165
2	-4.620	-4.620	27	0.030	0.150
3	-0.084	-0.084	28	2.162	0.206
4	0.822	0.822	29	-0.517	0.159
5	-1.364	-1.364	30	0.008	0.151
6	-0.569	-0.569	31	-0.492	0.159
7	1.960	1.960	32	0.121	0.150
8	-1.564	-1.564	33	0.233	0.149
9	-2.485	-2.485	34	-0.543	0.160
10	-0.595	-0.595	35	3.555	0.357
11	0.053	0.053	36	-0.038	0.151
12	2.345	2.345	37	-1.185	0.183
13	-0.130	-0.130	38	2.955	0.276

Items	Difficulty Level	Error Standar	Items	Difficulty Level	Error Standar
14	-1.744	-1.744	39	0.869	0.154
15	-1.793	-1.793	40	1.114	0.159
16	1.325	1.325	41	0.591	0.151
17	1.089	1.089	42	0.277	0.149
18	-1.607	-1.607	43	0.941	0.155
19	1.649	1.649	44	0.798	0.153
20	-1.564	-1.564	45	-1.054	0.177
21	-1.086	-1.086	46	-0.343	0.156
22	0.255	0.255	47	-0.543	0.160
23	0.434	0.434	48	0.917	0.155
24	1.244	1.244	49	-0.153	0.152
25	-0.813	-0.813	50	1.271	0.163

In Table 3, it can be seen that there are several items that have an extreme level of difficulty, which means that they are outside the domain of general student abilities, which is around -2 to +2. These items are item number 2 (-4.620), item number 9 (-2.485), item number 12 (2.345), and item number 35 (3.555). Thus, these items cannot be used as good items for measuring scientific reasoning abilities.

Table 4. List of Test-taker Responses That Don't Fit the Rasch Model

Testee	Nilai Chi Square	P Value	Testee	Nilai Chi Square	P Value
13	75.949	0.008	155	72.825	0.015
34	78.124	0.005	164	343.011	0.000
61	78.264	0.005	165	113.128	0.000
64	73.683	0.013	174	107.134	0.000
78	99.034	0.000	177	84.194	0.001
88	143.357	0.000	178	74.057	0.012
137	88.118	0.001	182	127.631	0.000
149	78.095	0.005	184	69.116	0.031
150	77.326	0.006	186	70.514	0.024
13	75.949	0.008	155	72.825	0.015
34	78.124	0.005	164	343.011	0.000
61	78.264	0.005	165	113.128	0.000
64	73.683	0.013	174	107.134	0.000
78	99.034	0.000	177	84.194	0.001
88	143.357	0.000	178	74.057	0.012
137	88.118	0.001	182	127.631	0.000
149	78.095	0.005	184	69.116	0.031
150	77.326	0.006	186	70.514	0.024

From Table 4, there are 18 students, or 9.27%, whose responses do not match the model used (the Rasch model) at the 49th level of freedom. This difference in student responses is known as an irregular response (abberance response), and the more students who have an irregularity response, the weaker the substantive validity of the test. The magnitude of the response that does not fit in Table 4 is 9.27%, so the magnitude of the response according to the model used is 90.73%. This shows the level of substantive validity of scientific reasoning tests at 90.73%, a fairly high index for a test's validity. Other parameters as a measure of substantive validity are accuracy, sensitivity, and specificity. The results of the analysis showed an accuracy value of 0.744, a sensitivity value of 0.798, and a specificity value of 0.68. These three values are close to 1, so it can be stated that the substantive validity of the scientific reasoning test is met.

The measure of structural validity uses a consistency test for estimating item parameters based on the sub-samples used. This consistency test uses the Likelihood Ratio Test (LR Test) by dividing the sample into two sub-samples based on the median total score of each student. The results of the LR Test show that test item number 2 is inconsistent in the two subsamples. Thus, test item number 2 cannot be used as a good test item. The value of the separation person strata, or reliability in the context of classical test theory, is the measure of external validity. The results of the analysis show a value of 0.787, a value close to 1.0, so it can be concluded that the scientific reasoning test that has been prepared fulfills the construct validity of the structural aspect.

By paying attention to the results of the trial analysis and paying attention to the distribution of difficulty levels of the test items, it was determined that 25 test items were suitable for measuring the ability of scientific literacy in the aspect of scientific reasoning. Table 5 describes the scientific reasoning test items that are not used as appropriate test instruments or appropriate test items. The level of difficulty of the test items that are suitable for use is shown in Table 6.

Table 5. Feasibility of Scientific Reasoning Test Items according to Validity Trial Results

Scientific Reasoning Test Items unfit for use	Scientific Reasoning Test Items proper to use
2,4,5,7,9,11,14,15,18,20,21,23,25,27,29,31, 34,35,38,39,41,44,45,47,49	1,3,6,8,10,12,13,16,17,19,22,24,26,28,30,32, 33,36,37,40,42,43,46,48,50

Table 6. Difficulty Level of Items in Scientific Reasoning Tests that are Eligible to use

Items	Difficulty Level	Items	Difficulty Level	Items	Difficulty Level
1	-1.402	10	1.649	19	-1.185
2	-0.084	11	0.255	20	1.114
3	-0.569	12	1.244	21	0.591
4	-1.564	13	-0.729	22	0.941
5	-0.595	14	2.162	23	-0.343
6	2.345	15	0.008	24	0.917
7	-0.130	16	0.121	25	1.271
8	1.325	17	0.233		
9	1.089	18	-0.038		

By paying attention to the four aspects of the validity of McNamara (2006), the results of the scientific reasoning tests carried out can be concluded, as explained in Table 7.

Table 7. Results of Construct Validity Analysis Using the Messick Validity Approach and Rasch Model Application on Scientific Reasoning Test Items

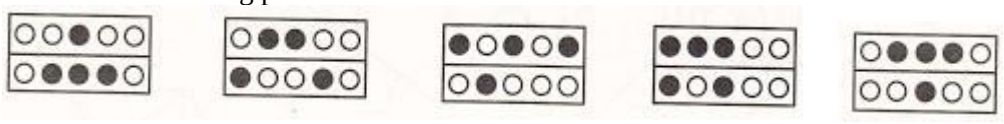
Construct Validity Aspect	Indicator	Criteria	Result
Content	Itemfit	P > 0.05	a total of 44 (88%) test items fit the model
Substantive	Person fit statistic	P > 0.05	a total of 176 (90.7%) student responses matched the model
	accuracy, sensitivity, dan specificity	close to 1,0	Accuracy: 0.744 Sensitivity: 0.798 Specificity: 0.68
Structural	Invariansi Test (LRtest)	P< 0,05	inconsistent item is item no 2
External	separation Person strata	Close to 1,0	0.787

The results of the feasibility analysis of this instrument are in line with the The concept of objective measurement in educational assessment has five criteria, namely: (1) producing a linear measure at the same interval; (2) the right estimation process; (3) identifying misfits or outliers; (4) being able to overcome missing data; and (5) producing measurements that are independent of the parameters studied (Bond et al., 2020). By paying attention to the results of the trial analysis and paying attention to the distribution of difficulty levels of the test items, it was determined that 25 test items were suitable for measuring the ability of scientific literacy in the aspect of scientific reasoning that's shown in Table 6.

The Scientific Reasoning Test that has been tested and meets the validity aspects of Messick (1995) with the application of the Rasch model can be shown by the following discussion.

The Scientific Reasoning Test Items

1. Look at the following picture:

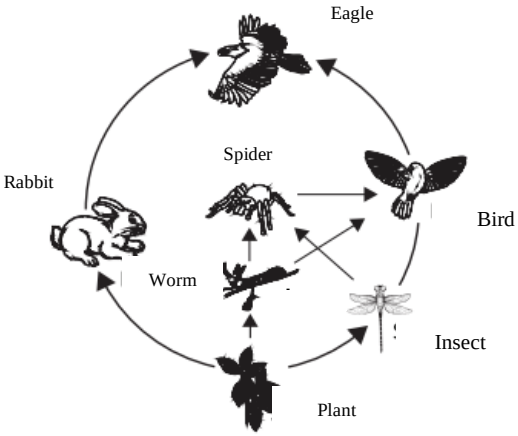


(1) (2) (3) (4) (5)

Different pictures are …

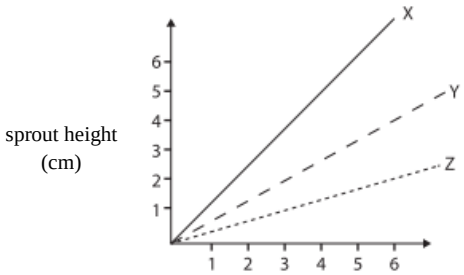
a. 1
b. 2
c. 3
d. 4
e. 5

2. Look at the pictures of the following food webs:



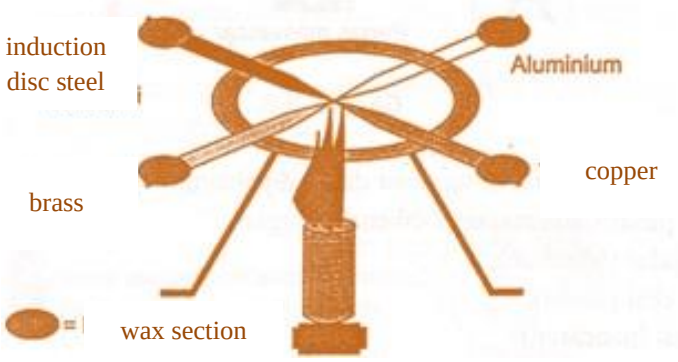
Organisms that are at the second trophic level …
a. rabbit and worm
b. eagle and spider
c. eagle dan bird
d. spider and bird
e. insect and bird

3. Look at the graph of the experimental results of sprout growth in the following dark, shady, and bright places.:



Based on the graph, it shows that X and Z, respectively, are sprouts placed in place.....
a. light and dark
b. dark and light
c. light and shady
d. dark and shady
e. shady and light

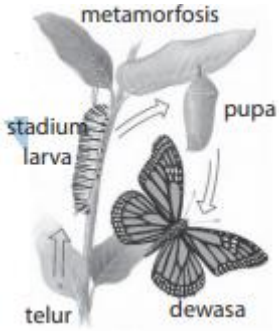
4. Conduction events can be practiced using a conduction disc with four different types of metal, as shown in the following figure:



Each metal tip is dripped with wax and then heated over the burner. Then we can observe what happens to the wax by seeing which metal melts it first. Conduction events can be observed in nature. …

a. when the tip of an aluminum spoon gets hot when you stir hot tea
b. when burning satay si over hot charcoal.
c. Get warm when near a campfire
d. When boiling water to a boil
e. When drying clothes in the washing machine

5. Look at the following picture:



Animals that have the same developmental phase as the animals in the picture above are ….

a. Grasshopper
b. Cricket
c. flies
d. snake
e. cockroach

6. The following table shows the experimental results of five types of conducting wires that have the same resistance. …

wire to	Long	wide
1	X	Y
2	2X	3Y
3	0,5X	Y
4	3X	2Y
5	2X	0,5Y

Based on the above table, the wire with the greatest resistance is the to wire …

a. 1
b. 2
c. 3
d. 4
e. 5

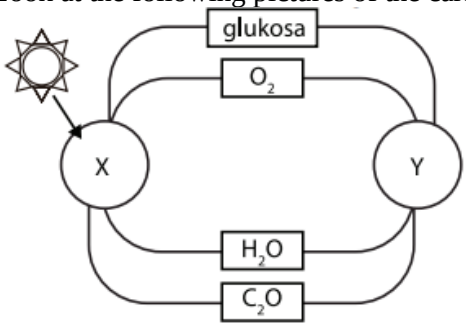
7. In the experiment at the level of organization of life, by making observation plots using the quadratic method (square size of 1 meter) and counting and then recording all organisms (animals and plants) found in the square, the following data is obtained:

No	Spesies	Plot 1	Plot 2	Plot 3	Amount	Average	K	F
1	Plant A	-	10	30	40	13.33	40	0.67
2	Plant B	2	5	8	15	5.00	15	1
3	Plant C	30	30	-	60	20.00	60	0.67
4	Plant D	11	23	-	44	14.67	44	0.67
Amount					159	53.00	159	3
Average					39.75			

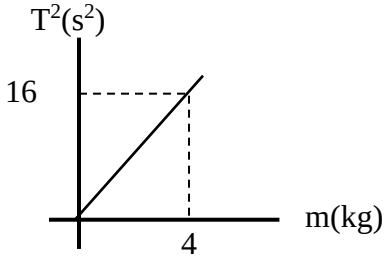
The best research questions to ask based on the data above are …

a. "Does the type of species affect density?"
b. "How to calculate density?"
c. "Which species has the highest density and frequency?"
d. "How to calculate relative frequency?"
e. "how to calculate the frequency?"

8. A force **F = (5i + 2j) N** does work with the capture point moving according to the equation **r = (2i + aj)** m, with i and j respectively denoting the unit vector in the x and y directions. If the work done by the force is 30 joules, then the value of a is equal to …
   a. 2
   b. 4
   c. 5
   d. 8
   e. 10
9. Look at the following pictures of the carbon and oxygen cycles:

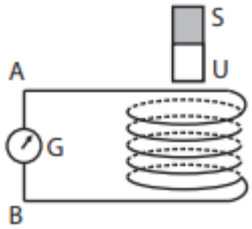


   The process that occurs in section X is …
   a. catabolism, the breakdown of glucose to produce CO_2 and O_2
   b photosynthesis, the formation of glucose to secrete CO_2 and O_2
   c. photosynthesis, the formation of carbohydrates
   d catabolism, the breakdown of glucose into O_2 and energy
   e. respiration, breakdown of glucose into H_2O and CO_2
10. An object connected to a spring has a period T and a force constant k. The relationship between the graph of T^2 and the mass of object m is in accordance with the following graph:



    The magnitude of the force constant k based on the graph is …
    a. 0,025 N/m
    b. 1 N/m
    c. 2 N/m
    d. 4 N/m
    e. 40 N/m
11. If Reva buys a skirt, she also buys a scarf.
    If Reva buys a scarf, then Sasa buys a jacket.
    Sasa buys pants. The conclusion that can be drawn is…
    a. Sasa and Reva bought skirt
    b. Reza didn't buy pants
    c. Reva bought scarf and jacket
    d. Reva and Sasa bought pants
    e. Reva bought skirt and scarf

12. The coil, like in the picture, is connected to a sensitive galvanometer. If current flows from A to B, the galvanometer needle moves to the right. If the magnetic north pole is inserted again, then the needle moves. …



a. turn right and freeze
b. turn left and freeze
c. to the right, to the left and then silence
d. to the left, to the right then silence
e. to the right then to the left

13. If two species with the same ecological niche and isolated habitat compete for food, then….
a. both will look for additional food sources
b. one of the species will become extinct
c. both will survive better
d. one of the species will seek a new habitat
e. both will create new ecological niches

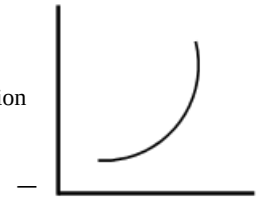
14. Consider the following experimental table:

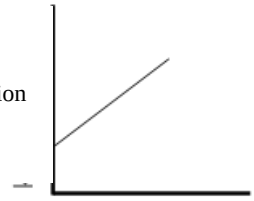
	Substance	Mass(grams)	Melting calories (J/kg)	received calories (J)
A	Water (ice)	500	$3{,}36 \times 10^5$	10000
B	Ammonia	500	$0{,}33 \times 10^5$	10000
C	Lead	500	$0{,}25 \times 10^5$	10000
D	Silver	500	$0{,}88 \times 10^5$	10000
E	Iron	500	$2{,}89 \times 10^5$	10000

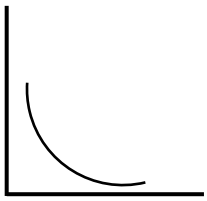
The substance that melts the most is …
a. water (ice)
b. Ammonia
c. Lead
d. Silver
e. Iron

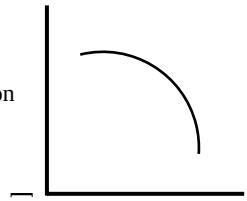
15. The equation for a transverse wave traveling along a very long string is $y = 6 \sin(0.02\pi x + 4\pi t)$. (y and x are in cm and t is in seconds), then:
(1) wave amplitude of 6 cm
(2) wave amplitude 100 cm
(3) wave frequency of 2 Hz
(4) positive x-th wave travels
The correct statement is …
a. 1, 2 and 3 are correct
b. 4 alone is correct
c. 1 and 3 are correct
d. everything is correct
e. 2 and 4 are correct

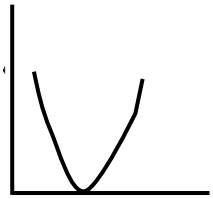
16. The graph showing the relationship between humidity and the rate of transpiration in plants is ….

A. transpiration rate

humidity

B. transpiration rate

Humidity

C. transpiration rate

humidity

D. transpiration rate

humidity

E. transpiration rate

humidity

17. A pair of identical springs is used to pull a train up a hill at a constant speed in two states. In the first state, the spring is connected to the carriage side by side. In the second situation, the springs connected to the train are connected to each other. In the first situation, the spring extends to provide a certain distance. In the second case, the spring extends twice as much as in the first situation. What problems might be being studied in the above situation?
    a. How is the steepness of a hill determined by a stretched spring?
    b. How does the material of the spring determine the amount of strain?
    c. How does the arrangement of the spring determine the amount of strain?
    d. How does the mass of the train determine the amount of strain?
    e. How does the speed of the train determine the amount of strain?
18. Lately, rainfall in various regions of Indonesia has increased. Not a few areas were damaged due to the rain that kept falling. Jakarta, which is prone to flooding every year, must brace itself for more flooding this rainy season. Poor drainage systems and changes in the function of agricultural land could be the causes of this disaster. Which of the following statements is true?
    a. The causes of flooding on earth are solely due to human activities
    b. Continuous rainfall is the main cause of flooding
    c. Good urban planning can prevent flooding
    d. Some of the causes of flooding are natural and some are caused by humans
    e. Nothing can be concluded from these readings
19. Ani conducted a growth experiment with the formulation of the problem, "Does humidity affect the growth of green beans?" Plant 1 is planted in a cardboard box covered with gauze with a 2 cm hole in the gauze and placed on the terrace. Plant 1 is sprayed with 1 bottle of water every morning. Plant 2 was planted in a box with gauze on top and a 2 cm hole in the gauze, on the same terrace, and was sprayed with 3 bottles of water every morning.

Based on the treatment described above, the purpose of spraying water with varying amounts of water is to weld. …

a. give different light intensities
b. different humidity
c. regulates water uptake by plants
d. maintain water content in the soil
e. retain moisture due to the gauze closure

20. A curved bridge has a radius of curvature R. The center of curvature is under the bridge. The force caused on the bridge by a car weighing W moving with speed n while at the top of the bridge, if g is the acceleration due to gravity, is equal to..…

a. $\frac{W\left(1+\frac{v^2}{R}\right)}{g}$
b. $W\left(1+\frac{v^2}{gR}\right)$
c. $\frac{Wv^2}{W+gR}$
d. $\frac{W\left(1-\frac{v^2}{R}\right)}{g}$
e. $W\left(1-\frac{v^2}{gR}\right)$

21. An ideal monatomic gas in a confined space has a pressure of 1.2 x 105 Pa and a volume of 40 liters. If the gas expands isobarically so that its volume becomes 50 liters, then the gas will absorb heat from the environment at a rate of 2 x 103 J. At the end of the process:
(1) gas temperature will decrease
(2) the gas does 1.2 x 10^3 J of work
(3) The gas undergoes an energy change in 80 J.
Is the statement above true or false, including the size and/or unit?
a. true gas temperature will decrease
b. the size of the effort is too large, the unit is correct.
c. the size of the business is too small, the units are correct.
d. the change in energy is too small, the units are correct
e. the amount of work and energy is wrong, the unit is wrong

22. Rahmat observed the diversity of animals in his school garden and grouped the animals as follows:

Group A	Group B	Group C
Earthworms	Snail	Spider
Termite	caterpillar	Scorpion
keluwing	Sparrows	Owl
	Bee	

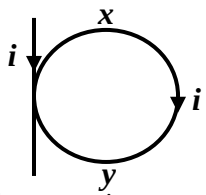
Suddenly, Rahmat found grasshoppers and butterflies. Based on the groupings made by Rahmat, the animals were put into groups. …

a. grasshoppers in group A and butterflies in group B
b. grasshoppers and butterflies in group B
c. grasshoppers in group B, butterflies in group C
d. grasshoppers and butterflies made a separate group
e. grasshoppers are made into a separate group, butterflies in group B

23. Ahmad heard the sound of crows at night. Then, the next day, he heard the sad news that his neighbour had died. What can be stated on the basis of this?

a. The sound of crows is a sign that someone has died
b. There is a connection between the sound of crows and sad news
c. It cannot be stated that there is a connection between the sound of crows and the news of someone dying
d. The sound of crows brings bad luck
e. The sound of crows is a sign that you have to be vigilant

24. Vehicles that use diesel fuel and power plants that use coal fuel will release SO^2 and NO_2 into the air, which will react with rainwater to form acid rain, which can cause…
    a. depletion of the ozone layer
    b. increase in soil and surface water pH
    c. damage to agricultural crops
    d. increase in ambient temperature
    e. increase in the amount of carbon
25. A long, straight, vertical wire is tangent to a coiled wire as shown in the figure.



The two wires carry the same electric current. Magnetic induction at the center of the coiled wire is …
    a. zero
    b. pointing outside the image area
    c. point into the image area
    d. point to X
    e. point to Y

Regarding the instructions for using the test, they can be explained as follows: Scientific reasoning tests have the potential to be high stakes or risk tests. As a result, students are not advised to study prior to attending, and their implementation must be closely monitored. Test administration can be done digitally or conventionally. In digital implementation, it can be done by randomly determining the number of test items that must be completed by students. The test is in the form of multiple-choice questions with five answer choices and has a working duration of 90 minutes. The scoring method is dichotomous: if it is successful, it is given a score of 1, and if it fails to answer, it is given a score of 0.

This test can be used as a graduation supplement for high school students in the MIPA program or MIPA specialization. This test is arranged based on an assessment with a benchmark reference. The graduation limit can be determined by each school according to the average ability of its students. The standard passing limit, taking the average test item difficulty level (0.345) into account, is around 55 to 57.5 in the standard 100. This test score will explain the competence of scientific literacy competencies in the aspect of scientific reasoning as an integral part of aspects of scientific investigation and aspects of knowledge of the nature of science. This research contributes to the improvement of assessment practice in science education especially in providing scientific literacy research instruments.

Conclusion

The results of the construct validity show that of the 50 items of the Scientific Reasoning Test Instrument that have been tried out, there are 25 questions that have met the aspects of Messick's validity. It has five criteria, namely: (1) producing a linear measure at the same interval; (2) the right estimation process; (3) identifying misfits or outliers; (4) being able to overcome missing data; and (5) producing measurements that are independent of the parameters studied.

Acknowledgment

Ministry of Education and Culture of the Republic of Indonesia for the 2022 Higher Education Excellence Applied Research Grant

Credit Authorship Contribution Statement

Yuni Arfiani: Visualization, Writing – original draft, Writing – review & editing. **Purwo Susongko**: Conceptualization, Methodology, Formal analysis, Resources, Writing – review & editing. **Mobinta Kusuma**: Writing – review & editing.

References

Ardianto, D., & Rubini, B. (2016). Comparison of students' scientific literacy in integrated science learning through model of guided discovery and problem based learning. *Jurnal Pendidikan IPA Indonesia*, *5*(1), 31–37. https://doi.org/10.15294/jpii.v5i1.5786

Bond, T., Yan, Z., & Heene, M. (2020). Applying the Rasch Model: Fundamental measurement in the human sciences. Routledge.

Chiang, P. M., & Tzou, H. I. (2018). The application of differential person functioning on the science literacy of Taiwan PISA 2015. *Humanities and Social Sciences Reviews*, *6*(1), 8–13. https://doi.org/10.18510/hssr.2018.612

Fakhriyah, F., Masfuah, S., & Mardapi, D. (2019). Developing scientific literacy-based teaching materials to improve students' computational thinking skills. *Jurnal Pendidikan IPA Indonesia*, *8*(4), 482–491. https://doi.org/10.15294/jpii.v8i4.19259

Hanushek, E. A., & Woessmann, L. (2016). Knowledge capital, growth, and the East Asian miracle: Access to schools achieves only so much if quality is poor. In *Science* (Vol. 351, Issue 6271, pp. 344–345). American Association for the Advancement of Science. https://doi.org/10.1126/science.aad7796

Hanson, S. (2016). The assessment of scientific reasoning skills of high school science students: A standardized assessment instrument.

Jong, C., Hodges, T. E., Royal, K. D., & Welder, R. M. (2015). Instruments to Measure Elementary Preservice Teachers' Conceptions: An Application of the Rasch Rating Scale Model. *Educational Research Quarterly*, *39*(1), 21–48.

Kind, P. E. R., & Osborne. (2017). Styles of Scientific Reasoning-a Cultural Rationale for Science Education? *Science Education*, *101*(1), 8–31.

Kuo, C. Y., Wu, H. K., Jen, T. H., & Hsu, Y. S. (2015). Development and Validation of a Multimedia-based Assessment of Scientific Inquiry Abilities. *International Journal of Science Education*, *37*(14), 2326–2357. https://doi.org/10.1080/09500693.2015.1078521

McFarlane, D. A. (2013). Understanding the Challenges of Science Education in the 21st Century: New Opportunities for Scientific Literacy. *International Letters of Social and Humanistic Sciences*, *4*, 35–44. https://doi.org/10.18052/www.scipress.com/ilshs.4.35

McNamara, T. (2006). Validity in Language Testing: The Challenge of Sam Messick's Legacy. *Language Assessment Quarterly*, *3*(1), 31–51. https://doi.org/10.1207/s15434311laq0301_3

Ratini, Muchtar, H., Suparman, M. A., Tamuri, A. H., & Susanto, E. (2018). The influence of learning models and learning reliance on students' scientific literacy. *Jurnal Pendidikan IPA Indonesia*, *7*(4), 458–466. https://doi.org/10.15294/jpii.v7i4.12489

Ravand, H., & Firoozi, T. (2016). Examining Construct Validity of the Master's UEEUsing the Rasch Model and the Six Aspects of the Messick's Framework. *International Journal of Language Testing*, *6*(1), 1–24. https://www.researchgate.net/publication/312093938

Roth, W. M., & Lee, S. (2016). Scientific literacy as collective praxis. Public understanding of Science.

Rudolph, J. L., & Horibe, S. (2016). What do we mean by science education for civic engagement? *Journal of Research in Science Teaching*, *53*(6), 805–820. https://doi.org/10.1002/tea.21303

Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an Instrument of Scientific Literacy Asessment on the Cycle Theme open access. In *international journal of environmental & science education* (Vol. 11, Issue 12).

Susongko, P. (2016). Validation of science achievement test with the Rasch model. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 268–277. https://doi.org/10.15294/jpii.v5i2.7690

Wang, C. C., Ho, H. C., Cheng, C. L., & Cheng, Y. Y. (2014). Application of the Rasch Model to the Measurement of Creativity: The Creative Achievement Questionnaire. *Creativity Research Journal*, *26*(1), 62–71. https://doi.org/10.1080/10400419.2013.843347

Wenning, C. J., & Vieyra, R. E. (2015). Teaching High School Physics Volume III. Rebecca Vieyra.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, *50*(9), 741.

Susongko, P. U. R. W. O., Arfiani, Y., & Kusuma, M. (2021). Determination of Gender Differential Butir Functioning in Tegal Students' Scientific Literacy Skills with Integrated Science (SLiSIS) Test Using Rasch Model. *Jurnal Pendidikan IPA Indonesia*, *10*(2), 270-281.

Susongko, P., Kusuma, M., & Widiatmo, H. (2019). Using Rasch Model to Detect Differential Person Functioning and Cheating Behavior in Natural Sciences Learning Achievement Test. *Jurnal Penelitian dan Pembelajaran IPA*, *5*(2), 94-111.