

Characterization of the Test of Scientific Reasoning Ability of Static Fluid Material for Prospective Physics Teachers: the Polytomous Rasch Analysis Authors'

Unang Purwana¹, Dadi Rusdiana², Winny Liliawati³
(purwana.u@upi.edu¹, dadirusdiana@upi.edu², winny@upi.edu³)

Physics Education Program Study Universitas Pendidikan Indonesia, Jl. Setiabudhi No 229 Bandung, Indonesia^{1,3}, Physics Program Study, Universitas Pendidikan Indonesia, Jl. Setiabudhi No 229 Bandung, Indonesia²

Abstract. This study aims to design a scientific reasoning ability test on static fluid material for prospective physics teachers and test the reliability of the test with the analysis of the polytomous Rasch model. The method used is a descriptive explanation. The test is in the form of a description of five questions measuring proportional reasoning ability, probability, correlation, and deductive hypothesis. The test was given to 28 (M=3 dan F=25) prospective teachers in one of the state universities in Bandung. The results obtained are item reliability 0.93, person reliability 0.84, and Cronbach Alpha reliability 0.85 with excellent categories. The result of the raw variance validity test was able to test the diversity of students' scientific reasoning abilities. The difficulty of the questions in the medium category, the accuracy, and the distinguishing power is in the excellent category.

Keywords: scientific reasoning, rasch model, static fluid

1 Introduction

Scientific reasoning ability is essential for science education throughout the world in this 21st century [1]. Scientific reasoning is one of the 21st-century skills and must be given to students and prospective teachers as an effort to prepare them to be able to face global challenges [2-5]. Although content mastery is still emphasized, skills such as formulating hypotheses, designing solutions to problems, and reasoning skills are still considered important [1,6]. Today, the world is more dependent on technology and society, and this makes the understanding of reasoning abilities is becoming increasingly important, especially for developing countries that need to focus on creating a STEM workforce (Science, Technology, Engineering, and Mathematics) [1,7].

Scientific reasoning is the process of applying logical principles to the scientific method, namely looking for problems, formulating hypotheses, making predictions, solutions and problems, creating experiments, controlling variables and analyzing data [8], processing information based on direct observation, and drawing more complex conclusions from the observed object [9]. Scientific reasoning abilities applies the principles of scientific investigation, starting with proposing hypotheses, planning experiments, devising or designing, conducting experiments to conclude [10], combining content knowledge, controlling variables, and using hypothetical-deductive reasoning to find solutions to scientific problems [1,9,11,12]. Scientific reasoning ability affects the level of ease of students in solving a problem [13-16].

Scientific reasoning ability is one of the skills tested in the Programme for International Students Assessment (PISA) test. PISA test results from 2009 showed the student's low scientific reasoning ability. In the PISA test, especially in science, Indonesia ranks 60 out of 65 countries and gets an average score of 383 from the OECD's average score, which is 501. Trends in Mathematics and Science Study (TIMSS) is an international study of students' mathematics and science achievement. The science achievement of Indonesian students in 2011 ranks last, namely 40th out of 40 countries. This is due to the low percentage of correct answers by TIMSS participants. The low ability of Indonesian students in PISA and TIMSS is because students cannot be challenged to think scientifically and have scientific reasoning. So far, education in Indonesia is still focused on the ability to think conceptually alone and have not trained scientific reasoning skills [17]

In physics education, assessment addresses various forms of scientific knowledge. Most of the existing assessment instruments emphasize the evaluation of content knowledge [18]. This assessment instrument fails to measure investigative ability. This study applies Rasch analysis to help validate assessment instruments in static fluid material to measure prospective high school physics teachers' reasoning abilities. The static fluid is one of the physics materials that require students' reasoning abilities to understand the concept of static fluid in everyday life. Fluid phenomena are physical phenomena that occur in everyday life.

2 Method research

The purpose of this study was to test the validity, reliability, level of difficulty, and distinguishing power of the prospective teacher's scientific reasoning ability instruments. The research method used is descriptive quantitative. The device of scientific reasoning ability on static fluid material is in the form of description questions totaling five questions. Problem no.1 (S1) deals with proportional reasoning, question no.2 (S2) probabilistic reasoning, question no.3 (S3) correlational reasoning, and question no.4 & no.5 (S4 & S5) hypothetical-deductive reasoning. The instrument was given to 28 prospective teachers (3 male and 25 female) in the 5th semester of the 2020/2021 academic year who took part in mechanics courses. Processing and test analysis using the Rasch method [19] with polyatomic data on a scale of 0 to 8 for each question. The information obtained from the Rasch model results is item measure, person measure, and wright map. Before the instrument of scientific reasoning ability was distributed, the device was tested for validation to 3 validators and obtained excellent results.

3. Result and Discussion

3.1. Reliability

Based on the results of data processing in Figure 1, it can be seen that the reliability of the person is .84 logit and the reliability of the item is .93 logit. Personnel reliability is included in the excellent category (0.80-0.90), and item reliability is included in the outstanding category. Besides, Cronbach's Alpha value is included in a particular type (> 0.83). Based on this reliability data, it can be concluded that this instrument is perfect for measuring the prospective teacher's scientific reasoning ability.

SUMMARY OF 28 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	26.4	5.0	.55	.62	.97	-.1	.99	.0
S.D.	4.7	.0	1.78	.04	.67	1.1	.73	1.1
MAX.	35.0	5.0	4.12	.69	2.69	2.2	2.88	2.3
MIN.	18.0	5.0	-2.54	.57	.09	-2.2	.10	-2.0
REAL RMSE	.70	TRUE SD	1.64	SEPARATION	2.33	Person	RELIABILITY	.84
MODEL RMSE	.62	TRUE SD	1.67	SEPARATION	2.67	Person	RELIABILITY	.88
S.E. OF Person MEAN = .34								

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .85

SUMMARY OF 5 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	147.6	28.0	.00	.26	.96	-.2	.99	-.1
S.D.	15.6	.0	1.06	.00	.31	1.2	.33	1.1
MAX.	168.0	28.0	1.66	.27	1.36	1.3	1.48	1.5
MIN.	123.0	28.0	-1.39	.26	.62	-1.5	.67	-1.2
REAL RMSE	.28	TRUE SD	1.02	SEPARATION	3.67	Item	RELIABILITY	.93
MODEL RMSE	.26	TRUE SD	1.03	SEPARATION	3.92	Item	RELIABILITY	.94
S.E. OF Item MEAN = .53								

Fig. 1. Reliability Test

3.2. Validity

Based on the raw-variance data, the results were 69.6% (> 40%). This shows whether the items used can measure the diversity of the respondent's ability. Data from natural variance indicates that the items used can measure the variety of prospective teachers' scientific reasoning abilities.

3.3. Fit Statistic (item analysis)-How much (item difficulty level).

This analysis shows the grouping of each item's level of difficulty regarding the scientific reasoning abilities of prospective teachers. The difficulty level of the thing can be seen through the measured value in the following Figure. 2.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	123	28	1.66	.26	.75	-.9	.82	-.6	.84	.80	57.1	51.9	S5
4	138	28	.65	.26	1.36	1.3	1.48	1.5	.71	.80	46.4	53.2	S4
3	152	28	-.29	.26	.62	-1.5	.67	-1.2	.85	.79	71.4	53.9	S3
1	157	28	-.63	.26	1.30	1.1	1.29	1.0	.71	.78	32.1	53.6	S1
2	168	28	-1.39	.27	.77	-.8	.71	-1.0	.85	.76	67.9	52.8	S2
MEAN	147.6	28.0	.00	.26	.96	-.2	.99	-.1			55.0	53.1	
S.D.	15.6	.0	1.06	.00	.31	1.2	.33	1.1			14.4	.7	

Fig. 2. Item Measure

This classification uses the logit average value and standard deviation information. Based on the data obtained in Figure 2, it is known that the mean is .00 logit and the standard deviation is 1.06 logit. The sum of the mean and SD values is 1.06 logit (+ 1 SD). The sum of the 1.06 logit value and the SD value is 2.12 logit (+ 2 SD). Next, the < SD group was obtained by reducing the mean value against the SD value, and it was obtained that <1 SD was - 1.06 logit and <2 SD was - 2.12 logit. Based on this information, it is known that there are no items that exceed >2 SD and <2 SD or items that are too difficult and too easy (outliers).

Overall, there are two items below the average with codes S4 and S5, namely hypothetical-deductive ability, while the other three-item problems have a difficulty level above the average of 0.0 logit. On the other hand, based on the data from this data processing, the standard deviation value is 1.06 logit. This information shows that the question of a prospective teacher's scientific reasoning abilities is spreading. The difficulty level of the question items differs from one thing to another. The characteristics of the items that have different levels of difficulty are excellent to be used as an instrument to measure the scientific reasoning ability of prospective teachers in a static fluid material. Through data processing using the Rasch model, information on a difficulty can be obtained more specifically through the mean and standard deviation values.

3.4. How Precise (Level of Measurement Accuracy).

The level of measurement accuracy by items can be known through Standard Error Measurement (SEM). SEM score of > 1.0 logit is included in the question category that lacks research power and distinguishing power. Based on the results of data processing in Figure 2, it is known that the SEM value of all items is <1.0. This data shows that all things about static fluid material's scientific reasoning ability have good accuracy and distinguishing power level. Thus, this item is perfect to be used as a measurement instrument.

3.5. Prospective Teacher's Scientific Reasoning Ability

The scientific reasoning ability of a prospective teacher is shown in measure values with logit as units. There are several classifications of future teacher's scientific reasoning abilities, as seen in Figure 3 below. Based on the data in Figure. 3, prospective teachers' scientific reasoning skills in static fluid material can be seen. With a value of + 4.12 logit, the teacher's highest power comes from two students with codes 10 and 22. At the same time, the lowest ability is - 2.54 logit from prospective teachers with code 16. Future teachers' power is slightly higher than the average difficulty level of the question items 0.0 logit as in the high, very tall, outlier classification. Overall, this item can distinguish the level of diversity in the prospective teacher's scientific reasoning abilities.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBSN	EXACT MATCH EXPN	Person
Outlier													
10	35	5	4.12	.69	1.21	.5	1.22	.5	.23	.59	60.0	54.2	10
22	35	5	4.12	.69	.57	-.7	.56	-.7	.92	.59	60.0	54.2	22
Very High													
14	33	5	3.20	.68	2.04	1.5	1.93	1.3	-.02	.56	20.0	56.0	14
19	32	5	2.74	.67	.34	-1.1	.34	-1.1	.65	.54	60.0	57.1	19
High													
2	30	5	1.85	.66	.47	-.8	.39	-.9	.00	.56	60.0	55.5	02
6	30	5	1.85	.66	1.98	1.4	2.05	1.4	.57	.56	40.0	55.5	06
7	30	5	1.85	.66	2.01	1.4	2.17	1.5	.54	.56	20.0	55.5	07
11	30	5	1.85	.66	.15	-1.8	.16	-1.6	.91	.56	100.0	55.5	11
21	30	5	1.85	.66	.61	-.4	.70	-.2	.92	.56	60.0	55.5	21
26	30	5	1.85	.66	.97	.2	.81	.0	.92	.56	60.0	55.5	26
3	28	5	1.03	.63	.62	-.4	.75	-.1	.85	.63	60.0	53.4	03
1	27	5	.65	.61	.09	-2.2	.10	-2.0	.94	.65	100.0	58.7	01
4	27	5	.65	.61	.56	-.6	.46	-.7	.72	.65	60.0	58.7	04
5	27	5	.65	.61	.67	-.3	.64	-.3	.64	.65	60.0	58.7	05
17	27	5	.65	.61	.67	-.3	.64	-.3	.64	.65	60.0	58.7	17
Low													
24	26	5	.28	.60	.94	.1	.93	.1	.80	.67	60.0	51.1	24
27	25	5	-.06	.58	.22	-1.7	.20	-1.6	.92	.67	60.0	43.6	27
23	24	5	-.40	.57	.58	-.6	.49	-.8	.78	.66	60.0	48.9	23
12	23	5	-.72	.57	.29	-1.6	.28	-1.5	.81	.64	80.0	47.1	12
18	23	5	-.72	.57	.46	-1.0	.41	-1.0	.73	.64	80.0	47.1	18
25	23	5	-.72	.57	.95	.1	1.01	.2	.93	.64	40.0	47.1	25
28	23	5	-.72	.57	.95	.1	1.01	.2	.93	.64	40.0	47.1	28
9	22	5	-1.04	.57	2.69	2.2	2.88	2.3	.35	.62	20.0	44.1	09
Very Low													
13	21	5	-1.38	.59	1.43	.8	1.74	1.2	-.78	.60	60.0	50.7	13
8	20	5	-1.73	.61	1.72	1.2	1.88	1.3	-.38	.58	20.0	52.8	08
15	20	5	-1.73	.61	1.82	1.3	2.17	1.6	.51	.58	20.0	52.8	15
20	19	5	-2.12	.64	.57	-.5	.55	-.6	.74	.57	60.0	52.2	20
16	18	5	-2.54	.67	1.50	.9	1.37	.7	.29	.56	60.0	58.9	16
MEAN	26.4	5.0	.55	.62	.97	-.1	.99	.0			55.0	53.1	
S.D.	4.7	.0	1.78	.04	.67	1.1	.73	1.1			21.1	4.5	

Fig.3. Person Measure

3.6. Wright Map

Based on the data distribution presented in Figure 4, it can be seen that the average line of the reasoning ability level of prospective teachers (.55 logit) is higher than the middle line of problem difficulty level (0.0 logit). This data shows that all future teachers' average ability in answering scientific reasoning questions is higher than the intermediate difficulty level of all questions. There was 16 prospective teacher who had logit scores above the moderate problem difficulty and 12 future teachers who obtained logit scores below the average problem difficulty level. Prospective teachers 10, 22, 14, and 19 were the groups of prospective teachers who had the most success in answering all scientific reasoning questions. Meanwhile, four future teachers, 08, 15, 20, and 16, are prospective teachers who have difficulty answering all scientific reasoning questions. On the other hand, it can be seen that all question items do not have the same problem. All question items have a different difficulty level from S5 as the most challenging item, while S2 is the most comfortable item to answer correctly.

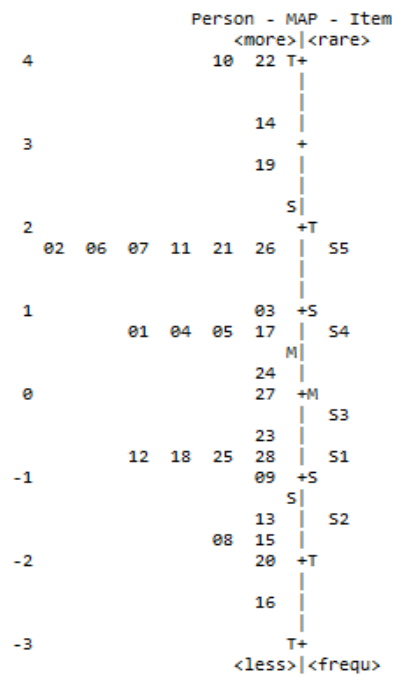


Fig. 4. Wright Map

The most challenging problem is hypothesis-deductive reasoning ability. Hypothesis-deductive reasoning ability is a higher-order thinking ability level used by scientists [20]. Hypothesis-deductive reasoning is related to testing the hypotheses obtained through deductive reasoning. A reasoned argument is an attempt to show that a conclusion must follow a set of premises. A deductive argument is valid if its decision follows precisely from the assumption; if the conclusion is correct, then the hypothesis is true. Reasoned arguments make sense if they are right and the premises are true. Deductive reasoning can be a factor that can help a prospective teacher recognize cognitive conflicts and solve the problems [21]. A person's emotions affect one's deductive reasoning. They tend to damage the reason [22]. This is influenced by the balance between two factors, namely personal and physiological relevance, acting independently. This is consistent with the new utility theory [22-25], which suggests that emotions play a significant role in human rationality [22].

4. Conclusion

The purpose of this study was to test the instrument for the scientific reasoning ability of static fluid material of prospective high school teachers. The device includes description questions totaling five questions testing proportional, probability, correlation, and hypothesis-deductive reasoning ability. The results obtained are item reliability 0.93, person reliability 0.84, and Cronbach's Alpha reliability 0.85 with excellent categories. The result of the raw variance validity test tested the diversity of students' scientific reasoning abilities. The difficulty of the questions in the medium category, the accuracy, and the distinguishing power is in an outstanding variety. The scientific reasoning ability test developed is suitable for use.

Acknowledgments. The author would like to thank the head of the science education study program Dr. Ida Kaniawati for motivating me to complete my doctoral studies. Thank you also to Dr. Muslim, who has helped complete the research and provided suggestions for improvement, as well as Dr. A. Rusli for the motivation and advice given.

References

- [1] Ibrahim B, Ding L, Mollohan KN, Stammen A. Scientific reasoning: theory evidence coordination in physics-based and non-physics-based tasks. *African Journal of Research in Mathematics, Science and Technology Education*. 2016 May 3;20(2):93-105.
- [2] Shofiyah N, Supardi ZA, Jatmiko B. Mengembangkan penalaran ilmiah (scientific reasoning) siswa melalui model pembelajaran 5e pada siswa kelas X SMAN 15 Surabaya. *Jurnal Pendidikan IPA Indonesia*. 2013;2(1).
- [3] Yulianti E, Zhafirah NN. Analisis Komprehensif pada Implementasi Pembelajaran dengan Model Inkuiri Terbimbing: Aspek Penalaran Ilmiah. *Jurnal Penelitian Pendidikan IPA*. 2020 Jan 31;6(1):125-30.
- [4] Utami P, Supeno S, Bektiarso S. Lembar Kerja Siswa (LKS) Berbasis Inkuiri dengan Bantuan Scaffolding Konseptual Untuk Meningkatkan Keterampilan Penalaran Ilmiah Fisika Siswa SMA. *FKIP E-Proceeding*. 2019 Dec 27;4(1):134-40.
- [5] Anjani F, Supeno S, Subiki S. Kemampuan Penalaran Ilmiah Siswa Sma Dalam Pembelajaran Fisika Menggunakan Model Inkuiri Terbimbing Disertai Diagram Berpikir Multidimensi. *Lantanida Journal*. 2020 Aug 6;8(1):13-28.
- [6] Ding L, Wei X, Mollohan K. Does higher education improve student scientific reasoning skills?. *International Journal of Science and Mathematics Education*. 2016 May 1;14(4):619-34.
- [7] Jegede OJ. School science and the development of scientific culture: A review of contemporary science education in Africa. *International Journal of Science Education*. 1997 Jan 1;19(1):1-20.
- [8] Hanson S. The assessment of scientific reasoning skills of high school science students: A standardized assessment instrument.
- [9] Lawson AE. The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*. 2004 Sep 1;2(3):307.
- [10] Zimmerman C, Klahr D. Development of scientific thinking. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. 2018 Mar 21; 4:1-25.
- [11] Kuhn D, Katz JB, Dean, Jr D.: Developing reason. *Thinking & Reasoning*. 2004 May 1;10(2):197-219.
- [12] Zeineddin A, Abd-El-Khalick F. Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of research in science teaching*. 2010 Nov;47(9):1064-93.
- [13] Hejnová E, Eisenmann P, Cihlár J, Pribyl J. Relations between Scientific Reasoning, Culture of Problem Solving and Pupil's School Performance. *Journal on Efficiency and Responsibility in Education and Science*. 2018;11(2):38-44.
- [14] Hong JC, Hwang MY, Liu MC, Ho HY, Chen YL. Using a "prediction–observation–explanation" inquiry model to enhance student interest and intention to continue science learning predicted by their Internet cognitive failure. *Computers & Education*. 2014 Mar 1; 72:110-20.
- [15] Develaki M. Using computer simulations for promoting model-based reasoning. *Science & Education*. 2017 Nov 1;26(7-9):1001-27.
- [16] Fawaiz S, Handayanto SK, Wahyudi HS. Eksplorasi Keterampilan Penalaran Ilmiah Berdasarkan Jenis Kelamin Siswa SMA. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*. 2020 Jul 1;5(7):934-43.
- [17] Novanti SK, Yulianti E, Mustikasari VR. Pengembangan Instrumen Tes Literasi Sains Siswa SMP Materi Tekanan Zat dan Penerapannya dalam Kehidupan Sehari-Hari. *Jurnal Pembelajaran Sains*. 2018 Dec 31;2(2):6-12.
- [18] Tarekegn G, Alemu M, Tadesse M, Kind PM. Development and Validation of a Test of In-service Science Teachers' Epistemic Knowledge in Electricity and Magnetism Using Rasch Analysis.

- African Journal of Research in Mathematics, Science and Technology Education. 2020 May 3;24(2):192-204.
- [19] Sumintono B, Widhiarso W. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi). Trim Komunikata Publishing House; 2014.
 - [20] Hidayah NN, Wiyanto W, Sopyan A. Analisis kemampuan berpikir deduksi hipotesis terhadap pemahaman konsep rangkaian resistor pada listrik arus searah. Physics Communication. 2017;1(1):34-42.
 - [21] Park J, Han S. Using deductive reasoning to promote the change of students' conceptions about force and motion. International Journal of Science Education. 2002 Jun 1;24(6):593-609.
 - [22] Caparos S, Blanchette I. Independent effects of relevance and arousal on deductive reasoning. Cognition and emotion. 2017 Jul 4;31(5):1012-22.
 - [23] Blanchette I, Caparos S. When emotions improve reasoning: The possible roles of relevance and utility. Thinking & Reasoning. 2013 Sep 1;19(3-4):399-413.
 - [24] Bonnefon JF. A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. Psychological Review. 2009 Oct;116(4):888.
 - [25] Mercier H, Sperber D. Why do humans reason? Arguments for an argumentative theory. Behavioral and brain sciences. 2011 Apr;34(2):57-74.